2013 Federated Conference on Computer Science and Information Systems

September 8–11, 2013. Kraków Poland

ISBN 978-1-4673-4471-5

2013 Federated Conference on Computer Science and Information Systems

September 8–11, 2013. Kraków Poland

ISBN 978-1-4673-4471-5

2013

2013 Federated Conference on Computer Science and Information Systems

2013

M. Ganzha, L. Maciaszek, M. Paprzycki (editors)

ISBN 978-1-4673-4471-5
IEEE Catalog Number CFP1385N-ART

Al. Solidarności 82A m. 5,
01-003 Warsaw
Poland

10662 Los Vaqueros Circle
Los Alamitos, CA 90720
USA

TEXnical editor: Aleksander Denisiuk

Dear Reader, it is our pleasure to present to you Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), which took place in Kraków, Poland, on September 8–11, 2013. Each of the papers, found in this volume, was refereed by at least two referees and the acceptance rate of full papers was approximately 43%.

FedCSIS was organized by the Polish Information Processing Society (Mazowsze Chapter), AGH University of Mining and Metalurgy, Wrocław University of Economics and Systems Research Institute Polish Academy of Sciences. FedCSIS was organized in technical cooperation with: IEEE Region 8, Computer Society Chapter Poland, Gdańsk Computer Society Chapter, Poland, Polish Chapter of the IEEE Computational Intelligence Society (CIS), ACM Special Interest Group on Applied Computing, Łódź ACM Chapter, SERSC: Science & Engineering Research Support soCiety, Informatics Europe, Asociación de Técnicos de Informática, Committee of the Computer Science of the Polish Academy of Sciences, The Polish Association for Information Systems, Polish Society for Business Informatics and Polish Chamber of Commerce for High Technology. Furthermore, the 8th International Symposium Advances in Artificial Intelligence and Applications (AAIA'13) was organized in technical cooperation with: IEEE SMC Technical Committee on Computational Collective Intelligence, International Rough Set Society, International Fuzzy Systems Association, Romanian Association for Artificial Intelligence, and Polish Neural Networks Society.

FedCSIS consisted of the following events:

- **AAIA'13—8th International Symposium Advances in Artificial Intelligence and Applications**
  - AIMA'13—3rd International Workshop on Artificial Intelligence in Medical Applications
  - ASIR'13—3rd International Workshop on Advances in Semantic Information Retrieval
  - WCO'13—6th Workshop on Computational Optimization
- **CSNS—Computer Science & Network Systems**
  - CANA'13—6th Computer Aspects of Numerical Algorithms
  - MMAP'13—6th International Symposium on Multimedia Applications and Processing
- **ECRM—Education, Curricula & Research Methods**
  - DS-RAIT'13—Doctoral Symposium on Recent Advances in Information Technology
  - ISEC'13—Information Systems Education & Curricula Workshop
- **iNetSApp—Innovative Network Systems and Applications**
  - WSN'13—2nd International Conference on Wireless Sensor Networks
  - SoFAST-WS'13—2nd International Symposium on Frontiers in Network Applications, Network Systems and Web Services
- **IT4MBS—Information Technology for Management, Business & Society**
  - ABICT'13—4th International Workshop on Advances in Business ICT
  - Agent Day'13
  - AITM'13—11th Conference on Advanced Information Technologies for Management
  - IT4L'13—2nd Workshop on Information Technologies for Logistics
  - KAM'13—19th Conference on Knowledge Acquisition and Management
  - TAMoCo'13—Techniques and Applications for Mobile Commerce
- **SSD&A—Software Systems Development & Applications**
  - ATSE'13—4th International Workshop Automating Test Case Design, Selection and Evaluation
  - IWCPS'13—International Workshop on Cyber-Physical Systems
  - PBDA'13—Performance of Business Database Applications
  - WAPL'13—4th Workshop on Advances in Programming Languages

Each of these events had its own Organizing and Program Committee. We would like to express our warmest gratitude to members of all of them for their hard work attracting and later refereeing 420 submissions.

FedCSIS was organized under the auspices of Prof. Barbara Kudrycka, Minister of Science and Higher Education, dr Michał Boni, Minister of Administration and Digitization, Prof. Michał Kleiber, President of Polish Academy of Sciences, Marek Sowa, Marshal of Małopolska and Prof. Jacek Majchrowski, Mayor of Kraków. It was sponsored by Ministry of Science and Higher Eduction and Intel.

Maria Ganzha, *Conference Co-Chair, Systems Research Institute Polish Academy of Sciences, Warsaw, Poland, and Gdańsk University, Gdańsk, Poland*

Leszek Maciaszek, *Conference Co-Chair, Wrocław University of Economics, Wrocław, Poland and Macquarie University, Sydney, Australia*

Marcin Paprzycki, *Conference Co-Chair, Systems Research Institute Polish Academy of Sciences, Warsaw and Management Academy, Warsaw, Poland*

Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS)

September 8–11, 2013. Kraków, Poland

TABLE OF CONTENTS

3RD International Workshop on Artificial Intelligence in Medical Applications

3RD International Workshop on Advances in Semantic Information Retrieval

6TH Workshop on Computational Optimization

Computer Science & Network Systems

Computer Aspects of Numerical Algorithms

International Symposium on Multimedia Applications and Processing

Information Systems Education & Curricula Workshop

Innovative Network Systems and Applications

2ND International Symposium on Frontiers in Network Applications, Network Systems and Web Services

2ND International Conference on Wireless Sensor Networks

Information Technology for Management, Business & Society

4TH International Workshop on Advances in Business ICT

Agent Day

11TH Conference on Advanced Information Technologies for Management

2ND Workshop on Information Technologies for Logistics

19TH Conference on Knowledge Acquisition and Management

Techniques and Applications for Mobile Commerce

Software Systems Development & Applications

4TH International Workshop Automating Test Case Design, Selection and Evaluation

International Workshop on Cyber-Physical Systems

Performance of Business Database Applications

4TH Workshop on Advances in Programming Languages

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1–10

A General Divide and Conquer Approach for Process Mining

Wil M.P. van der Aalst
Architecture of Information Systems, Eindhoven University of Technology,
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
International Laboratory of Process-Aware Information Systems,
National Research University Higher School of Economics (HSE),
33 Kirpichnaya Str., Moscow, Russia.
Email: w.m.p.v.d.aalst@tue.nl

***Abstract*—Operational processes leave trails in the information systems supporting them. Such event data are the starting point for process mining – an emerging scientific discipline relating modeled and observed behavior. The relevance of process mining is increasing as more and more event data become available. The increasing volume of such data ("Big Data") provides both opportunities and challenges for process mining. In this paper we focus on two particular types of process mining: *process discovery* (learning a process model from example behavior recorded in an event log) and *conformance checking* (diagnosing and quantifying discrepancies between observed behavior and modeled behavior). These tasks become challenging when there are hundreds or even thousands of different activities and millions of cases. Typically, process mining algorithms are linear in the number of cases and exponential in the number of different activities. This paper proposes a very general divide-and-conquer approach that decomposes the event log based on a partitioning of activities. Unlike existing approaches, this paper does not assume a particular process representation (e.g., Petri nets or BPMN) and allows for various decomposition strategies (e.g., SESE- or passage-based decomposition). Moreover, the generic divide-and-conquer approach reveals the core requirements for decomposing process discovery and conformance checking problems.**

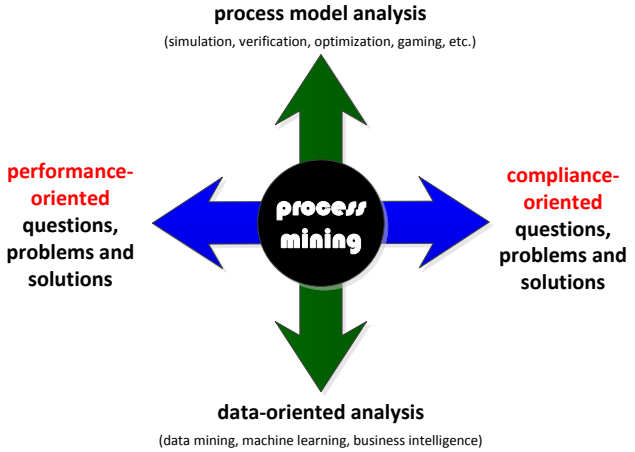


Fig. 1. Process mining is on the interface between process model analysis and data-oriented analysis and can be used to answer a variety of performance and compliance-related questions.

I. Introduction

RECENTLY, *process mining* emerged as a new scientific discipline on the interface between process models and event data [1]. Conventional *Business Process Management* (BPM) [2] and *Workflow Management* (WfM) [3] approaches and tools are mostly model-driven with little consideration for event data. *Data Mining* (DM) [4], *Business Intelligence* (BI), and *Machine Learning* (ML) [5] focus on data without considering end-to-end process models. Process mining aims to bridge the gap between BPM and WfM on the one hand and DM, BI, and ML on the other hand (cf. Figure 1).

The practical relevance of process mining is increasing as more and more event data become available (cf. the recent attention for "Big Data"). Process mining techniques aim to *discover, monitor and improve real processes by extracting knowledge from event logs*. The two most prominent process mining tasks are: (i) *process discovery*: learning a process model from example behavior recorded in an event log, and (ii) *conformance checking*: diagnosing and quantifying discrepancies between observed behavior and modeled behavior.

Starting point for any process mining task is an *event log*. Each *event* in such a log refers to an *activity* (i.e., a well-defined step in some process) and is related to a particular *case* (i.e., a *process instance*). The events belonging to a case are ordered, and can be seen as one "run" of the process. Such a run is often referred to as a *trace*. It is important to note that an event log contains only example behavior, i.e., we cannot assume that all possible runs have been observed.

Lion's share of process mining research has been devoted to process discovery [1]. Here the challenge is to turn a multiset of example traces (observed cases) into a process model. Process representations allowing for concurrency and choice, e.g., Petri nets, BPMN models, UML activity diagrams, or EPCs, are preferred over low-level notations such as finite state machines of hidden Markov models [1].

Given a process model (discovered or made by hand) and an event log one can try to *align modeled and observed behavior*. An alignment relates a trace in a event log to its corresponding path in the model. If there is not a direct match, the trace is aligned with the closest or most likely path. Such alignments can be used to answer *performance-oriented* and *compliance-oriented* questions (cf. Figure 1). Alignments can be used to show how often paths are taken and activities are being executed. Moreover, events often bear a timestamp which

978-83-60810-53-8/$25.00

can be used to compute flow times, waiting times, service times, etc. For example, alignments can be used to highlight bottlenecks in the process model. Similarly, alignments can be used to show where model and event log disagree. This is commonly referred to as *conformance checking*.

The incredible growth of event data is also posing new challenges [6]. As event logs grow, process mining techniques need to become more efficient and highly scalable. Moreover, torrents of event data need to be distributed over multiple databases and large process mining problems need to be distributed over a network of computers. Several approaches have been described in literature [7], [8], [9], [10], [11] (also see the related work described in Section VII). In this paper, we describe a *generic divide-and-conquer approach* based on a (valid) *partitioning* of the activities in sets. The activity sets should overlap if there is a direct dependency. We will illustrate the divide-and-conquer approach for the two main process mining tasks:

- For conformance checking, we decompose the process model into smaller partly overlapping submodels using projection. The event log is decomposed into sublogs, also using projection. Any trace that fits into the overall model also fits all submodels. The reverse only holds if the partitioning is *valid*. Metrics such as the fraction of fitting cases can be computed by checking the conformance of the submodels.
- To decompose process discovery, we first create an activity partitioning, i.e., we split the set of activities into a collection of partly overlapping activity sets. For each activity set, we project the log onto a sublog and discover a submodel for it. The different submodels can be merged to create an overall process model. Again it is guaranteed that all traces in the event log that fit into the overall model also fit into the submodels and vice versa.

Unlike existing papers [7], [8], [9], [10], [11], we abstract from a concrete process representation and do not select a particular decomposition strategy. Instead we focus on the core requirements to enable decomposition.

The remainder is organized as follows. Section II introduces preliminaries ranging from multisets to event logs. Section III provides abstract high-level definitions for process discovery and conformance checking. Section IV shows that any process mining problem can be decomposed trivially, but with a possible loss of precision. Section V shows that exact results can be obtained (not just bounds) if the activity partitioning is valid. Section VI discusses possible strategies to obtain valid (or otherwise suitable) activity partitionings. Related work is described in Section VII. Section VIII concludes the paper.

II. PRELIMINARIES

Before describing the two main process mining tasks and the ways in which these tasks can be distributed, we introduce some basic notations to reason about event logs and process models.

A. Multisets, Sequences and Projection

Multisets are used to represent the state of a Petri net and to describe event logs where the same trace may appear multiple times.

$\mathcal{B}(A)$ is the set of all multisets over some set A. For some multiset $b \in \mathcal{B}(A)$, $b(a)$ denotes the number of times element $a \in A$ appears in b. Some examples: $b_1 = [\,]$, $b_2 = [x, x, y]$, $b_3 = [x, y, z]$, $b_4 = [x, x, y, x, y, z]$, $b_5 = [x^3, y^2, z]$ are multisets over $A = \{x, y, z\}$. b_1 is the empty multiset, b_2 and b_3 both consist of three elements, and $b_4 = b_5$, i.e., the ordering of elements is irrelevant and a more compact notation may be used for repeating elements.

The standard set operators can be extended to multisets, e.g., $z \in b_3$, $b_2 \uplus b_3 = b_4$, $b_5 \setminus b_2 = b_3$, $|b_5| = 6$, etc. Bags are compared in the usual manner, i.e., $b_2 \leq b_4$ and $b_2 \not\leq b_3$. $\{a \in b\}$ denotes the set with all elements a for which $b(a) \geq 1$. $[f(a) \mid a \in b]$ denotes the multiset where element $f(a)$ appears $\sum_{x \in b | f(x) = f(a)} b(x)$ times.

$\mathcal{P}(X)$ is the powerset of X, i.e., $Y \in \mathcal{P}(X)$ if $Y \subseteq X$.

$\sigma = \langle a_1, a_2, \ldots, a_n \rangle \in X^*$ denotes a sequence over X. $|\sigma| = n$ is its length. $a \in \sigma$ if and only if $a \in \{a_1, a_2, \ldots, a_n\}$. $\langle\,\rangle$ is the empty sequence.

Projection is defined for sequences and sets or bags of sequences.

Definition 1 (Projection): Let $\sigma \in X^*$ and $Y \subseteq X$. $\sigma{\restriction_Y}$ is the projection of σ on Y, i.e., all elements in $X \setminus Y$ are removed (e.g., $\langle x, y, z, x, y, z \rangle{\restriction_{\{x,y\}}} = \langle x, y, x, y \rangle$). Projection is generalized to sets and bags. If $s \in \mathcal{P}(X^*)$, then $s{\restriction_Y} = \{\sigma{\restriction_Y} \mid \sigma \in s\}$ If $b \in \mathcal{B}(X^*)$, then $b{\restriction_Y} = [\sigma{\restriction_Y} \mid \sigma \in b]$. In the latter case frequencies are respected, e.g., $[\langle x, y, z, y \rangle^{10}, \langle z, y, z, y \rangle^5]{\restriction_{\{x,y\}}} = [\langle x, y, y \rangle^{10}, \langle y, y \rangle^5]$.

Without proof we mention some basic properties for sequences and projections.

Lemma 1 (Projection Properties): Let $\sigma \in X^*$, $Y \subseteq X$, $s \in \mathcal{P}(X^*)$, and $b \in \mathcal{B}(X^*)$.

- $\sigma \in s \;\Rightarrow\; \sigma{\restriction_Y} \in s{\restriction_Y}$,
- $\sigma \in b \;\Rightarrow\; \sigma{\restriction_Y} \in b{\restriction_Y}$,
- $\sigma \in s{\restriction_Y} \;\Leftrightarrow\; \exists_{\sigma' \in s}\ \sigma = \sigma'{\restriction_Y}$,
- $\sigma \in b{\restriction_Y} \;\Leftrightarrow\; \exists_{\sigma' \in b}\ \sigma = \sigma'{\restriction_Y}$, and
- for any $\sigma_1 \in {X_1}^*$ and $\sigma_2 \in {X_2}^*$: ${\sigma_1}{\restriction_{X_2}} = {\sigma_2}{\restriction_{X_1}} \;\Leftrightarrow\; \exists_{\sigma_3 \in (X_1 \cup X_2)^*}\ {\sigma_3}{\restriction_{X_1}} = \sigma_1 \;\wedge {\sigma_3}{\restriction_{X_2}} = \sigma_2$.

B. Activities, Traces, Event Logs, and Models

Event logs serve as the starting point for process mining. An event log is a multiset of traces. Each trace describes the life-cycle of a particular case (i.e., a process instance) in terms of the activities executed. Process models are represented as sets of traces. As indicated earlier, we avoid restricting ourselves to a specific process notation. However, we will show some Petri nets and a BPMN model for illustration purposes.

Definition 2 (Universe of Activities, Universe of Traces): $\mathcal{A}$ is the universe of *activities*, i.e., the set of all possible and relevant activities. Other activities cannot be observed (or are

abstracted from). Elements of $\mathcal{A}$ may have *attributes*, e.g., costs, resource information, duration information, etc. A *trace* $\sigma \in \mathcal{A}^*$ is a sequence of activities found in an event log or corresponding to a run of some process model. $\mathcal{U} = \mathcal{A}^*$ is the universe of all possible traces over $\mathcal{A}$.

We assume that an activity is identified by attributes relevant for learning, i.e., irrelevant attributes are removed and attribute values may be coarsened. $|\mathcal{A}|$ is the number of unique activities. Process models with hundreds of activities (or more) tend to be unreadable. In the remainder we will refer to activities using a single letter (e.g. a), however, an activity could also be $decide(gold, manager, reject)$ to represent a decision to reject a gold customer's request by a manager.

In a process model a specific trace $\sigma \in \mathcal{U}$ is possible or not. Hence, a model can be characterized by its set of allowed traces.

Definition 3 (Process Model): A *process model* M is a non-empty collection of traces, i.e., $M \in \mathcal{P}(\mathcal{U})$ and $M \neq \emptyset$.

$A_M = \bigcup_{\sigma \in M} \{a \in \sigma\}$ is the set of activities possible in M. Figure 2 shows a process model M using the Business Process Model and Notation (BPMN) [12]. For this paper the representation itself is irrelevant. Trace $\langle a, b, d, e, f, c, d, g \rangle$ is one of the infinitely many possible traces of M.

An event log is a multiset of *sample* traces from a known or unknown process. The same trace can appear multiple times in the log. Moreover, the event log contains only example behavior. Often only few of the possible traces are observed [1].

Definition 4 (Event Log): An *event log* $L \in \mathcal{B}(\mathcal{U})$ is a multiset of observed traces.

$A_L = \bigcup_{\sigma \in L} \{a \in \sigma\}$ is the set of activities occurring in L. Note that projection (see Definition 1) is defined for both models and event logs.

$L = [\langle a, b, d, e, g \rangle^5, \langle a, c, d, e, h \rangle^4, \langle a, b, d, e, f, c, d, e, g \rangle]$ is an event log containing 10 traces that could have been generated by the BPMN model in Figure 2, e.g., five cases followed the path $\langle a, b, d, e, g \rangle$.

III. Process Discovery and Conformance Checking

In the introduction we already informally introduced the two main process mining tasks: *process discovery* (learning a model from a collection of example behaviors) and *conformance checking* (identifying mismatches between observed and modeled behavior). Using definitions 3 and 4 we can now formalize these notions at a high abstraction level.

Definition 5 (Process Discovery Technique): A process discovery technique $disc \in \mathcal{B}(\mathcal{U}) \rightarrow \mathcal{P}(\mathcal{U})$ is a function that produces a process model $disc(L) \in \mathcal{P}(\mathcal{U})$ for any event log $L \in \mathcal{B}(\mathcal{U})$.

Given an event log $L = [\langle a, c \rangle^5, \langle a, b, c \rangle^4, \langle a, b, b, b, b, c \rangle]$, the discovery technique may discover the process model that always starts with activity a, followed by zero or more b activities, and always ends with a c activity: $disc(L) = \{\langle a, c \rangle, \langle a, b, c \rangle, \langle a, b, b, c \rangle, \ldots\}$.

An example of a discovery algorithm is the α algorithm [13] that produces a Petri net based on the patterns identified in the event log. Many discovery techniques have been proposed in literature [14], [13], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] and are supported by open source tools such as ProM and commercial tools such as Disco (Fluxicon), Perceptive Process Mining (also known as Futura Reflect), ARIS Process Performance Manager (Software AG), QPR ProcessAnalyzer, Interstage Process Discovery (Fujitsu), Discovery Analyst (StereoLOGIC), and XMAnalyzer (XMPro). It is impossible to provide an complete overview of all techniques here. Very different approaches can be followed, e.g., using heuristics [19], [23], inductive logic programming [20], state-based regions [14], [18], [22], language-based regions [16], [24], and genetic algorithms [21].

There are four quality dimensions for comparing model and log: (1) *fitness*, (2) *simplicity*, (3) *precision*, and (4) *generalization* [1]. A model with good *fitness* allows for most of the behavior seen in the event log. A model has a perfect fitness if all traces in the log can be replayed by the model from beginning to end. The *simplest* model that can explain the behavior seen in the log is the best model. This principle is known as Occam's Razor. Fitness and simplicity alone are not sufficient to judge the quality of a discovered process model. For example, it is very easy to construct an extremely simple Petri net ("flower model") that is able to replay all traces in an event log (but also any other event log referring to the same set of activities). Similarly, it is undesirable to have a model that only allows for the exact behavior seen in the event log. Remember that the log contains only example behavior and that many traces that are possible may not have been observed yet. A model is *precise* if it does not allow for "too much" behavior. Clearly, the "flower model" lacks precision. A model that is not precise is "underfitting". Underfitting is the problem that the model over-generalizes the example behavior in the log (i.e., the model allows for behaviors very different from what was seen in the log). At the same time, the model should generalize and not restrict behavior to just the examples seen in the log. A model that does not *generalize* is "overfitting". Overfitting is the problem that a very specific model is generated whereas it is obvious that the log only holds example behavior (i.e., the model explains the particular sample log, but there is a high probability that the model is unable to explain the next batch of cases).

We often focus on fitness, e.g., event log $L = [\langle a, b, d, e, g \rangle^5, \langle a, c, d, e, h \rangle^4, \langle a, b, d, e, f, c, d, e, g \rangle]$ is perfectly fitting model M described in Figure 2.

Definition 6 (Conformance Checking Technique): A conformance checking technique $check \in (\mathcal{B}(\mathcal{U}) \times \mathcal{P}(\mathcal{U})) \rightarrow \mathcal{D}$ is a function that computes conformance diagnostics $check(L, M) \in \mathcal{D}$ (e.g., fitness or precision metrics) given an event log $L \in \mathcal{B}(\mathcal{U})$ and process model $M \in \mathcal{P}(\mathcal{U})$. $\mathcal{D}$ is the set of all possible diagnoses (e.g., a fitness value between 0 and 1) and depends on the metric chosen.

As indicated, we will often focus on fitness. Hence, we introduce some functions characterizing fitness.

Definition 7 (Conformance Checking Functions): Given an event log $L \in \mathcal{B}(\mathcal{U})$ and process model $M \in \mathcal{P}(\mathcal{U})$, we define the following functions:

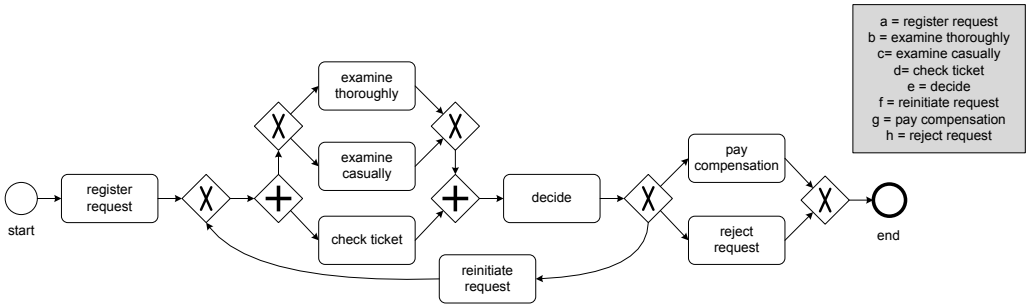


Fig. 2. A process model $M = \{\langle a,b,d,e,g\rangle, \langle a,c,d,e,g\rangle, \langle a,d,b,e,g\rangle, \langle a,d,c,e,g\rangle, \langle a,b,d,e,h\rangle, \langle a,c,d,e,h\rangle, \langle a,d,b,e,h\rangle, \langle a,d,c,e,h\rangle, \langle a,b,d,e,f,c,d,e,g\rangle, \langle a,c,d,e,f,b,d,e,h\rangle, \ldots\}$ expressed in terms of BPMN.

- $\mathit{fit}(L,M) = [\sigma \in L \mid \sigma \in M]$ is the multiset of fitting traces,
- $\mathit{nofit}(L,M) = [\sigma \in L \mid \sigma \notin M]$ is the multiset of non-fitting traces,
- $check_{pft}(L,M) = \frac{|\mathit{fit}(L,M)|}{|L|}$ is the fraction of traces in the event log perfectly fitting the model,
- $check_{bpf}(L,M) = (\mathit{fit}(L,M) = L)$ is a boolean function returning true if all traces in the event log fit.

Note that $\mathit{fit}(L,M) \uplus \mathit{nofit}(L,M) = L$. Clearly, $check_{pft}(L,M)$ and $check_{bpf}(L,M)$ are conformance checking functions in the spirit of Definition 6.

Simply counting the fraction of fitting cases is often too simplistic. Typically, one is interested in conformance metrics at the event level. For example, Petri-net based conformance checking approaches [25] may count the number of missing and remaining tokens during replay. Many conformance checking techniques have been proposed [26], [27], [28], [29], [30], [31], [20], [32], [33], [25], [34]. The alignment-based approach described in [26], [27] is much more sophisticated and provides much better diagnostics than simple metrics such as $check_{pft}(L,M)$. To illustrate the notion of alignments, let us consider the non-fitting trace $\langle a,b,c,e,g\rangle$ and process model M depicted in Figure 2. An optimal alignment is:

$$\gamma_1 = \left|\begin{array}{c|c|c|c|c|c} a & b & c & \gg & e & g \\ a & b & \gg & d & e & g \end{array}\right|$$

The $\gg$ symbols indicate the problems. The third column shows a "move on log only" $(c, \gg)$ indicating that the observed c event cannot be matched with a move of the model. The fourth column shows a "move on model only" $(\gg, d)$ indicating that the necessary execution of d in the model cannot be matched with an event in the log. All other columns refer to "synchronous moves", i.e., model and log agree. Alignment γ_1 has lowest costs assuming equal costs to all non-synchronous moves, i.e., simply count the number of $\gg$ symbols. A non-optimal alignment is:

$$\gamma_2 = \left|\begin{array}{c|c|c|c|c|c|c|c|c|c} a & b & \gg & \gg & \gg & c & \gg & e & g \\ a & b & d & e & f & c & d & e & g \end{array}\right|$$

Alignment γ_2 has four non-synchronous moves, i.e., double the number of non-synchronous moves compared to γ_1. Therefore, it is not optimal and not considered during analysis.

The alignment-based approach is very flexible because it can deal with arbitrary cost functions and any model representation. For example, one can associate costs to activities that are executed too late or by the wrong person. Alignments can also be used for computing precision and generalization [26], [33]. However, the approach can be rather time consuming. Therefore, the efficiency gains obtained through decomposition can be considerable for larger processes.

For simplicity we will focus in the remainder on the fraction of perfectly fitting traces $check_{pft}(L,M)$. However, as illustrated by the results in [7], we can use our decomposition approach also for more sophisticated alignment-based approaches.

IV. Decomposing Models and Logs

As events logs and process models grow in size, process mining may become a time consuming activity. Conformance checking may become intractable when many different traces need to be aligned with a model that allows for an exponential (or even infinite) number of traces. Event logs may contain millions of events. Finding the best alignment may require solving many optimization problems or repeated state-space explorations. In worst case, a state-space exploration of the model is needed per event. When using genetic process mining, one needs to check the fitness of every individual model in every generation. As a result, millions of conformance checks may be required. For each conformance check, the whole event log needs to be traversed.

For process discovery there are similar problems. Now the model is not given and the challenge is to find a model that scores good with respect to different objectives, e.g., fitness, simplicity, precision, and generalization. Depending on the representational bias, there may be infinitely many candidate models.

Given these challenges, we are interested in reducing the time needed for process mining tasks by decomposing the associated event log. In this section, we show that it is possible to decompose any process mining problem by *partitioning the set of activities*.

Definition 8 (Activity Partitioning): $P = \{A_1, A_2, \ldots, A_n\}$ is an *activity partitioning* of a set A if $A = \bigcup_{1 \leq i \leq n} A_i$. The activity sets may overlap. $\tilde{A}_i = A_i \cap \bigcup_{j \neq i} A_j$ are all activities that A_i shares with other activity sets. $A_i^I = A_i \setminus \tilde{A}_i$

are the internal activities of A_i. $\overline{A}_i = A \setminus (A_i^I) = \bigcup_{j \neq i} A_j$ are the non-internal activities of A_i.

Note that $\overline{A}_i \cap A_i = \tilde{A}_i$. A possible activity partitioning for the activities used in Figure 2 is $P = \{\{a, b, c, d, e, f\}, \{e, f, g, h\}\}$. Note that both activity sets in P share activities e and f.

Given an activity partitioning P, activity set $A \in P$, trace $\sigma \in \mathcal{U}$, model $M \in \mathcal{P}(\mathcal{U})$, and event log $L \in \mathcal{B}(\mathcal{U})$, we define the following terms:

- $\sigma{\restriction_A}$ is a *subtrace* of σ,
- $M{\restriction_A} \in \mathcal{P}(\mathcal{U})$ is a *submodel* of M, and
- $L{\restriction_A} \in \mathcal{B}(\mathcal{U})$ is a *sublog* of L.

Given an activity partitioning P consisting of n activity sets, we can partition the overall model into n submodels and the overall event log into n sublogs.[1]

It is easy to see that any trace that fits the overall model also fits any submodel (use the first property of Lemma 1). The reverse does not need to hold in case the behavior of one submodel may depend on internal behavior of another submodel. Nevertheless, we can compute bounds for conformance and use this insight for decomposed process discovery.

Definition 9 (Alternative Conformance Functions): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model, $P = \{A_1, A_2, \ldots, A_n\}$ an activity partitioning of A_M, and $L \in \mathcal{B}(\mathcal{U})$ an event log. We define variants of the functions in Definition 7 that only use submodels and sublogs.

- $\mathit{fit}^P(L, M) = [\sigma \in L \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}]$,
- $\mathit{check}^P_{pft}(L, M) = \frac{|\mathit{fit}^P(L,M)|}{|L|}$, and
- $\mathit{check}^P_{bpf}(L, M) = (\mathit{fit}^P(L, M) = L)$.

Theorem 1 (Conformance Bounds): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model and $P = \{A_1, A_2, \ldots, A_n\}$ an activity partitioning of A_M. For any event log $L \in \mathcal{B}(\mathcal{U})$:

- $\mathit{fit}(L, M) \leq \mathit{fit}^P(L, M)$,
- $\mathit{check}_{pft}(L, M) \leq \mathit{check}^P_{pft}(L, M)$, and
- $\mathit{check}_{bpf}(L, M) \Rightarrow \mathit{check}^P_{bpf}(L, M)$.

Proof: Let $\sigma \in \mathit{fit}(L, M)$, i.e., $\sigma \in L$ and $\sigma \in M$. Using Lemma 1 we can deduce that $\sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}$ for any i. Hence, $\sigma \in \mathit{fit}^P(L, M)$. This proves that $\mathit{fit}(L, M) \leq \mathit{fit}^P(L, M)$. The two other statements follow directly from this. ■

Consider activity partitioning $P = \{A_1, A_2\}$ with $A_1 = \{a, b, c, d, e, f\}$, and $A_2 = \{e, f, g, h\}$ for model M in Figure 2 and $L = [\langle a, b, d, e, g\rangle^5, \langle a, c, d, e, h\rangle^4, \langle a, b, d, e, f, c, d, e, g\rangle]$. $M{\restriction_{A_1}} = \{\langle a, b, d, e\rangle, \langle a, c, d, e\rangle, \langle a, d, b, e\rangle, \langle a, d, c, e\rangle, \langle a, b, d, e, f, c, d, e\rangle, \langle a, c, d, e, f, b, d, e\rangle, \ldots\}$ and $M{\restriction_{A_2}} = \{\langle e, g\rangle, \langle e, h\rangle, \langle e, f, e, g\rangle, \langle e, f, e, h\rangle, \ldots\}$. $L{\restriction_{A_1}} = [\langle a, b, d, e\rangle^5, \langle a, c, d, e\rangle^4, \langle a, b, d, e, f, c, d, e\rangle]$ and $L{\restriction_{A_2}} = [\langle e, g\rangle^5, \langle e, h\rangle^4, \langle e, f, e, g\rangle]$. $\mathit{fit}(L, M) = \mathit{fit}^P(L, M) = L$, i.e., all traces fit into the overall model and all submodels. Hence, $\mathit{check}_{pft}(L, M) = \mathit{check}^P_{pft}(L, M) = 1$, and $\mathit{check}_{bpf}(L, M) = \mathit{check}^P_{bpf}(L, M) = \mathit{true}$.

[1]Note that we use the term "partition" in a loose manner. As indicated before, activity sets may overlap and hence submodels and sublogs can also overlap in terms of events/activities.

V. Valid Activity Partitioning

Consider the following four event logs each containing 100 traces: $L_1 = [\langle a, c, d\rangle^{50}, \langle b, c, e\rangle^{50}]$, $L_2 = [\langle a, c, d\rangle^{25}, \langle a, c, e\rangle^{25}, \langle b, c, d\rangle^{25}, \langle b, c, e\rangle^{25}]$, $L_3 = [\langle a, c, d\rangle^{49}, \langle a, c, e\rangle^{1}, \langle b, c, d\rangle^{1}, \langle b, c, e\rangle^{49}]$, and $L_4 = [\langle a, c, d\rangle^{25}, \langle d, c, a\rangle^{25}, \langle b, c, e\rangle^{25}, \langle e, c, b\rangle^{25}]$. Also consider the process models $M_1 = \{\langle a, c, d\rangle, \langle b, c, e\rangle\}$ and $M_2 = \{\langle a, c, d\rangle, \langle a, c, e\rangle, \langle b, c, d\rangle, \langle b, c, e\rangle\}$. Figure 3 shows both models in terms of a Petri net. M_1 is the model *with* places p_1 and p_2 and M_2 is the model *without* these places. Note that the Petri net is just shown for illustration purposes. None of the definitions or results in this paper depends on a particular representation (see Definition 3).

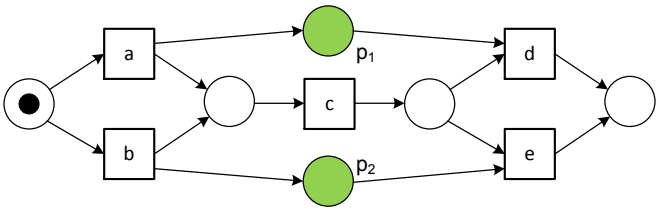


Fig. 3. Two Petri nets used to illustrate valid activity partitioning. $M_1 = \{\langle a, c, d\rangle, \langle b, c, e\rangle\}$ is the model with places p_1 and p_2 and $M_2 = \{\langle a, c, d\rangle, \langle a, c, e\rangle, \langle b, c, d\rangle, \langle b, c, e\rangle\}$ is the model without these places.

Given these three event logs and two process models, we can compute the following fractions of fitting traces:

$\mathit{check}_{pft}(L_1, M_1) = 1$	$\mathit{check}_{pft}(L_1, M_2) = 1$
$\mathit{check}_{pft}(L_2, M_1) = 0.50$	$\mathit{check}_{pft}(L_2, M_2) = 1$
$\mathit{check}_{pft}(L_3, M_1) = 0.98$	$\mathit{check}_{pft}(L_3, M_2) = 1$
$\mathit{check}_{pft}(L_4, M_1) = 0.50$	$\mathit{check}_{pft}(L_4, M_2) = 0.50$

Consider now the activity partitioning $P = \{A_1, A_2\}$ with $A_1 = \{a, b, c\}$ and $A_2 = \{c, d, e\}$. Using this partitioning more events logs will perfectly fit the individual submodels:

$\mathit{check}^P_{pft}(L_1, M_1) = 1$	$\mathit{check}^P_{pft}(L_1, M_2) = 1$
$\mathit{check}^P_{pft}(L_2, M_1) = 1$	$\mathit{check}^P_{pft}(L_2, M_2) = 1$
$\mathit{check}^P_{pft}(L_3, M_1) = 1$	$\mathit{check}^P_{pft}(L_3, M_2) = 1$
$\mathit{check}^P_{pft}(L_4, M_1) = 0.50$	$\mathit{check}^P_{pft}(L_4, M_2) = 0.50$

This illustrates that in general $\mathit{check}_{pft}(L, M) \leq \mathit{check}^P_{pft}(L, M)$, but both values do not need to be the same, e.g., $\mathit{check}_{pft}(L_2, M_1) \neq \mathit{check}^P_{pft}(L_2, M_1)$. In this case the difference is caused by the dependency between the two choices in M_1 which is not "communicated" via the activities in $A_1 \cap A_2 = \{c\}$. This example triggers the question when $\mathit{check}_{pft}(L, M) = \mathit{check}^P_{pft}(L, M)$, or more precisely, what properties must activity partitioning P have such that $\mathit{fit}(L, M) = \mathit{fit}^P(L, M)$?

An activity partitioning is *valid* if the "internal behavior" of one set of activities A_i does not depend on the internals of

another activity set A_j. It other words: the activity sets need to synchronize on non-local phenomena.

Definition 10 (Valid Activity Partitioning): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model with activities A_M and $P = \{A_1, A_2, \ldots, A_n\}$ an activity partitioning of the set A_M. P is *valid* for M if and only if $M = \{\sigma \in {A_M}^* \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\}$.

Activity partitioning $P = \{A_1, A_2\}$ with $A_1 = \{a, b, c\}$ and $A_2 = \{c, d, e\}$ is valid for M_2 but not for M_1. Note that $\langle a, c, e\rangle{\restriction_{A_1}} = \langle a, c\rangle \in M_1{\restriction_{A_1}}$ and $\langle a, c, e\rangle{\restriction_{A_2}} = \langle c, e\rangle \in M_1{\restriction_{A_2}}$, but $\langle a, c, e\rangle \notin M_1$.

The following theorem shows that a valid activity partitioning allows us to compute conformance per submodel without loosing precision, i.e., we get exact values rather than bounds.

Theorem 2 (Conformance Checking Can Be Decomposed): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model and $P = \{A_1, A_2, \ldots, A_n\}$ a valid activity partitioning of A_M. For any event log $L \in \mathcal{B}(\mathcal{U})$:

- $\mathit{fit}(L, M) = \mathit{fit}^P(L, M)$,
- $\mathit{check}_{pft}(L, M) = \mathit{check}^P_{pft}(L, M)$, and
- $\mathit{check}_{bpf}(L, M) = \mathit{check}^P_{bpf}(L, M)$.

Proof: $\mathit{fit}(L, M) = [\sigma \in L \mid \sigma \in M] = [\sigma \in L \mid \sigma \in \{\sigma' \in {A_M}^* \mid \forall_{1 \leq i \leq n}\ \sigma'{\restriction_{A_i}} \in M{\restriction_{A_i}}\}] = [\sigma \in L \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}] = \mathit{fit}^P(L, M)$. This proves the first statement. The two other statements follow directly from this. ■

Theorem 2 shows that conformance checking can be decomposed. Next we show that also discovery can be decomposed. Here we only have an event log to start with. However, while constructing the overall model we can simply assume independence and thus ensure a valid activity partitioning.

Corollary 1 (Process Discovery Can Be Decomposed): Let $L \in \mathcal{B}(\mathcal{U})$ be an event log and $P = \{A_1, A_2, \ldots, A_n\}$ an activity partitioning of A_L. Let $\mathit{disc} \in \mathcal{B}(\mathcal{U}) \rightarrow \mathcal{P}(\mathcal{U})$ be a discovery algorithm used to obtain the submodels $M_i = \mathit{disc}(L{\restriction_{A_i}})$ with $i \in \{1, \ldots, n\}$. $M = \{\sigma \in U \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M_i\}$ is the overall model constructed by merging the discovered submodels.

- P is a valid activity partitioning for M,
- $\mathit{fit}(L, M) = \mathit{fit}^P(L, M)$,
- $\mathit{check}_{pft}(L, M) = \mathit{check}^P_{pft}(L, M)$, and
- $\mathit{check}_{bpf}(L, M) = \mathit{check}^P_{bpf}(L, M)$.

By applying the construction of Corollary 1 we can decompose process discovery. If all the sublogs fit perfectly, then the overall event log will also fit the overall model perfectly. However, if activity sets are not overlapping sufficiently, the model may be underfitting (too general).

One may try to relax the validity requirement. For example, by considering one activity set A_i and its complete environment $\overline{A_i}$.

Definition 11 (Weakly Valid Activity Partitioning): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model with activities A_M and $P = \{A_1, A_2, \ldots, A_n\}$ an activity partitioning of the set A_M. P is *weakly valid* if and only if $M = \{\sigma \in {A_M}^* \mid \exists_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}} \wedge \sigma{\restriction_{\overline{A_i}}} \in M{\restriction_{\overline{A_i}}}\}$.

Clearly, any valid activity partitioning is also weakly valid. If $\sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}$ and $\sigma{\restriction_{\overline{A_i}}} \in M{\restriction_{\overline{A_i}}}$, then we can apply Lemma 1 to show that $\sigma{\restriction_{A_j}} \in M{\restriction_{A_j}}$ for $j \neq i$. The reverse does not hold. Consider for example $M = \{\langle a, b, d, e, g\rangle, \langle a, c, d, f, g\rangle$ and $P = \{A_1, A_2, A_3, A_4\}$ with $A_1 = \{a, b, c\}$, $A_2 = \{b, c, d\}$, $A_3 = \{d, e, f\}$, and $A_4 = \{e, f, g\}$. $P = \{A_1, A_2, A_3, A_4\}$ is not a valid activity partitioning because the traces $\langle a, c, d, e, g\rangle$ and $\langle a, b, d, f, g\rangle$ are not in M. However, P is weakly valid. This example also shows that Theorem 2 in general does not hold for weakly valid activity partitionings.

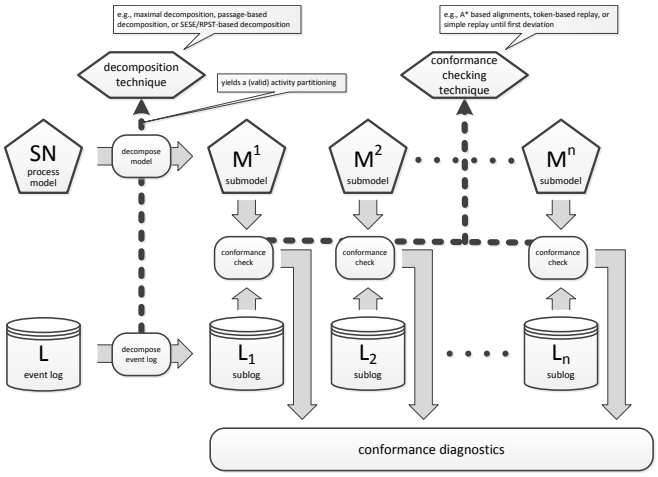


Fig. 4. Overview of decomposed conformance checking showing the configurable elements of the approach: (a) the *decomposition technique* yielding the activity partitioning that is used to split the overall model and event log, and (b) the *conformance checking technique* used to analyze the individual models and sublogs.

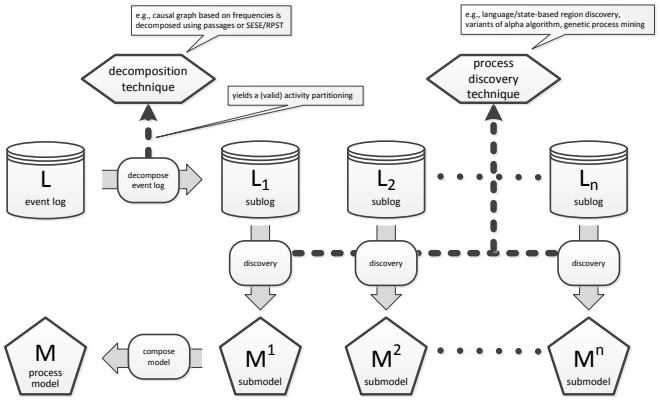


Fig. 5. Overview of decomposed process discovery showing the configurable elements of the approach: (a) the *decomposition technique* yielding the activity partitioning that is the basis for creating sublogs and (b) the *process discovery technique* used to infer submodels from sublogs.

Figures 4 and 5 summarize the overall approach proposed in this paper. Moreover, these figures point out the configurable elements. For conformance checking we need to find an activity partitioning P and a conformance checking technique

$check \in (\mathcal{B}(\mathcal{U}) \times \mathcal{P}(\mathcal{U})) \rightarrow \mathcal{D}$ (see Definition 6). For process discovery we need to find an activity partitioning P (based on only the event log since there is no initial model) and a process discovery technique $disc \in \mathcal{B}(\mathcal{U}) \rightarrow \mathcal{P}(\mathcal{U})$ (see Definition 5).

VI. Finding a (Valid) Activity Partitioning

Theorems 1 and 2 are very general, but stand or fall with a suitable activity partitioning. Consider the following two extreme activity partitionings for an activity set A: $P_{one} = \{A\}$ (just one activity set containing all activities) and $P_{all} = \{\{a\} \mid a \in A\}$ (one activity set per activity). Both are not very useful. P_{one} does not decompose the problem, i.e., there is still one big task. P_{all} decomposes the set of activities into singleton activity sets. P_{all} considers all activities in isolation, hence conformance checking and discovery are only considering frequencies of activities and not their order. For conformance checking P_{all} is typically not valid and decomposed discovery using P_{all} will most likely result in a severely underfitting model.

For conformance checking we can exploit the structure of the process model when searching for a (valid) activity partitioning. For example, the process model (e.g., a Petri net) can be decomposed using the so-called Refined Process Structure Tree (RPST) [35], [36] as shown in [10], [9]. The RPST allows for the construction of a hierarchy of SESE (Single-Exit-Single-Entry) components. Slicing the SESE at the desired level of granularity corresponds to a decomposition of the graph [9] that can be used for process mining.

In [7] an algorithm providing the so-called "maximal decomposition" of a Petri net is given. The construction of the maximal decomposition is based on partitioning the edges. Each edge will end up in precisely one submodel. Edges are taken together if they are connected through a place, through an internal transition (invisible action), or through multiple transitions having the same label. The algorithm iterates until no edges need to joined. Any labeled Petri net has a unique maximal decomposition and this decomposition defines a valid activity partitioning.

The notion of "passages" defined in [8], [37] provides an alternative approach to decompose a Petri net. A passage is a pair of two non-empty sets of activities (X, Y) such that the set of direct successors of X is Y and the set of direct predecessors of Y is X. As shown in [8], any Petri net can be partitioned using passages such that all edges sharing a source vertex or sink vertex are in the same set. This is done to ensure that splits and joins are not decomposed. Note that passages do not necessarily aim at high cohesion and low coupling. Nevertheless, they define a valid activity partitioning.

For discovery we cannot exploit the structure of the model to ensure the validity or suitability of an activity partitioning. Therefore, often an intermediate step is used [7]. For example, one can mine for frequent item sets to find activities that often happen together. Another, probably better performing, approach is to first create a *causal graph* (A, R) where A is the set of activities and $R \subseteq A \times A$ is a relation on A. The interpretation of $(a_1, a_2) \in R$ is that there is a "causal relation" between a_1 and a_2. Most process mining algorithms already build such a graph in a preprocessing step. For example, the α algorithm [13], the heuristic miner [23], and the fuzzy miner [38] scan the event log to see how many times a_1 is followed by a_2. If this occurs above a certain threshold, then it is assumed that $(a_1, a_2) \in R$. Even for large logs it is relatively easy to construct a causal graph (linear in the size of the event log). Moreover, counting the frequencies used to determine a causal graph can be distributed easily by partitioning the cases in the log. Also sampling (determining the graph based on representative examples) may be used to further reduce computation time.

Given a causal graph, one can view decomposition as a graph partitioning problem [39], [40], [41], [42]. There are various approaches to partition the graph such that certain constraints are satisfied while optimizing particular metrics. For example, in [42] a vertex-cut based graph partitioning algorithm is proposed ensuring the balance of the resulting partitions while simultaneously minimizing the number of vertices that are cut (and thus replicated).

Some of the notions in graph partitioning are related to "cohesion and coupling" in software development [43]. Cohesion is the degree to which the elements of a module belong together. Coupling is the degree to which each module relies on the other modules. Typically, one aims at "high cohesion" and "low coupling". In terms of our problem this means that we would like to have activity sets that consist of closely related activities whereas the overlap between the different activity sets is as small as possible while still respecting the causalities.

Definition 10 also suggests to investigate correlations between activity sets. An activity set $A_i \in P$ may be influenced through $\tilde{A}_i = A_i \cap \bigcup_{j \neq i} A_j$ but not through $\overline{A}_i \setminus A_i$. The goal is to find suitable "milestone activities", i.e., shared activities "decoupling" two activity sets.

The ideas mentioned above have only been explored superficially, but nicely illustrate that there are many promising directions for future research.

Interestingly, we can merge activity sets without jeopardizing validity. This allows us to decompose process mining problems at different levels of granularity or to provide a "tree view" on the process and its conformance.

Theorem 3 (Hierarchy Preserves Validity): Let $M \in \mathcal{P}(\mathcal{U})$ be a process model and $P = \{A_1, A_2, \ldots, A_n\}$ a valid activity partitioning of A_M with $n \geq 2$. Activity partitioning $P' = \{A_1 \cup A_2, A_3 \ldots, A_n\}$ is also valid.

Proof: Since P is valid $M = \{\sigma \in {A_M}^* \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\}$. We need to prove: $\{\sigma \in {A_M}^* \mid \sigma{\restriction_{A_1 \cup A_2}} \in M{\restriction_{A_1 \cup A_2}} \wedge \forall_{3 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\} = M$. $\sigma{\restriction_{A_1 \cup A_2}} \in M{\restriction_{A_1 \cup A_2}}$ implies that $\sigma{\restriction_{A_1}} \in M{\restriction_{A_1}}$ and $\sigma{\restriction_{A_2}} \in M{\restriction_{A_2}}$ (Lemma 1). Hence, $\{\sigma \in {A_M}^* \mid \sigma{\restriction_{A_1 \cup A_2}} \in M{\restriction_{A_1 \cup A_2}} \wedge \forall_{3 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\} \subseteq \{\sigma \in {A_M}^* \mid \forall_{1 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\} = M$. Moreover, for any $\sigma \in M$, $\sigma{\restriction_{A_1 \cup A_2}} \in M{\restriction_{A_1 \cup A_2}} \wedge \forall_{3 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}$ trivially holds. The observation that $M \subseteq \{\sigma \in {A_M}^* \mid \sigma{\restriction_{A_1 \cup A_2}} \in M{\restriction_{A_1 \cup A_2}} \wedge \forall_{3 \leq i \leq n}\ \sigma{\restriction_{A_i}} \in M{\restriction_{A_i}}\} \subseteq M$ completes the proof. ■

Theorem 3 can be applied iteratively. Hence, any combination of activity sets originating from a valid activity partitioning yields another valid activity partitioning. This allows us to coarsen any valid activity partitioning. Note that if $P = \{A_1, A_2, \ldots, A_n\}$ is not valid, then $P' = \{A_1 \cup A_2, A_3 \ldots, A_n\}$

may still be valid. Therefore, we can try to merge problematic activity sets in order to get a valid activity partitioning and better (i.e., more precise) process mining results. For example, when we are using alignments as described in [26], [27] we can diagnose the activity sets that disagree. We can also give preference to alignments that to not disagree on the interface of different activity sets. Last but no least, we can create a hierarchy of conformance/discovery results, similar to a dendrogram in hierarchical clustering.

VII. Related Work

For an introduction to process mining we refer to [1]. For an overview of best practices and challenges, we refer to the Process Mining Manifesto [44]. Also note the availability of open source tools such as ProM and commercial tools such as Disco (Fluxicon), Perceptive Process Mining (also known as Futura Reflect), ARIS Process Performance Manager (Software AG), QPR ProcessAnalyzer, Interstage Process Discovery (Fujitsu), Discovery Analyst (StereoLOGIC), and XMAnalyzer (XMPro).

The goal of this paper is to decompose challenging process discovery and conformance checking problems into smaller problems [6]. Therefore, we first review some of the techniques available for process discovery and conformance checking.

Process discovery, i.e., discovering a process model from a multiset of example traces, is a very challenging problem and various discovery techniques have been proposed [14], [13], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Many of these techniques use Petri nets during the discovery process and/or to represent the discovered model. It is impossible to provide an complete overview of all techniques here. Very different approaches are used, e.g., heuristics [19], [23], inductive logic programming [20], state-based regions [14], [18], [22], language-based regions [16], [24], and genetic algorithms [21]. Classical synthesis techniques based on regions [45] cannot be applied directly because the event log contains only example behavior. For state-based regions one first needs to create an automaton as described in [14]. Moreover, when constructing the regions, one should avoid overfitting. Language-based regions seem good candidates for discovering transition-bordered Petri nets that can serve as submodels [16], [24]. Unfortunately, these techniques still have problems dealing with infrequent/incomplete behavior.

There are four competing quality criteria when comparing modeled behavior and recorded behavior: fitness, simplicity, precision, and generalization [1]. In this paper, we focused on fitness, but also precision and generalization can also be investigated per submodel. Various conformance checking techniques have been proposed in recent years [26], [27], [28], [29], [30], [31], [20], [32], [33], [25], [34]. Conformance checking can be used to evaluate the quality of discovered processes but can also be used for auditing purposes [46]. Most of the techniques mentioned can be combined with our decomposition approach. The most challenging part is to aggregate the metrics per model fragment and sublog into metrics for the overall model and log. We consider the approach described in [27] to be most promising as it constructs an optimal alignment given an arbitrary cost function. This alignment can be used for computing precision and generalization [26], [33]. However, the approach can be rather time consuming. Therefore, the efficiency gains obtained through decomposition can be considerable for larger processes with many activities and possible subnets.

Little work has been done on the decomposition and distribution of process mining problems [6], [7]. In [47] MapReduce is used to scale event correlation as a preprocessing step for process mining. In [48] an approach is described to distribute genetic process mining over multiple computers. In this approach candidate models are distributed and in a similar fashion also the log can be distributed. However, individual models are not partitioned over multiple nodes. Therefore, the approach in this paper is complementary. Moreover, unlike [48], the decomposition approach in this paper is not restricted to genetic process mining.

More related are the divide-and-conquer techniques presented in [49]. In [49] it is shown that region-based synthesis can be done at the level of synchronized State Machine Components (SMCs). Also a heuristic is given to partition the causal dependency graph into overlapping sets of events that are used to construct sets of SMCs. In this paper we provide a different (more local) partitioning of the problem and, unlike [49] which focuses specifically on state-based region mining, we decouple the decomposition approach from the actual conformance checking and process discovery approaches.

Also related is the work on conformance checking of proclets [50]. Proclets can be used to define so-called artifact centric processes, i.e., processes that are not monolithic but that are composed of smaller interacting processes (called proclets). In [50] it is shown that conformance checking can be done per proclet by projecting the event log onto a single proclet while considering interface transitions in the surrounding proclets.

Several approaches have been proposed to distribute the verification of Petri net properties, e.g., by partitioning the state space using a hash function [51] or by modularizing the state space using localized strongly connected components [52]. These techniques do not consider event logs and cannot be applied to process mining.

Most data mining techniques can be distributed [53], e.g., distributed classification, distributed clustering, and distributed association rule mining [54]. These techniques often partition the input data and cannot be used for the discovery of Petri nets.

This paper generalizes the results presented in [7], [8], [9], [10], [11] to arbitrary decompositions (Petri-net based or not). In [8], [55], [11] it is shown that so-called “passages” [37] can be used to decompose both process discovery and conformance checking problems. In [10], [9] it is shown that so-called SESE (Single-Exit-Single-Entry) components obtained through the Refined Process Structure Tree (RPST) [35], [36] can be used to decompose conformance checking problems. These papers use a particular particular decomposition strategy. However, as shown in [7], there are many ways to decompose process mining problems.

The results in [7] are general but only apply to Petri nets. This paper further generalizes the divide and conquer approach beyond Petri nets. This allows us to simplify the presentation

and clearly show the key requirements for decomposing both process discovery and conformance checking problems.

VIII. Conclusion

In this paper we provided a high-level view on the decomposition of process mining tasks. Both conformance checking and process discovery problems can be divided into smaller problems that can be distributed over multiple computers. Moreover, due to the exponential nature of most process mining techniques, the time needed to solve "many smaller problems" is less than the time needed to solve "one big problem". Therefore, decomposition is useful even if the smaller tasks are done on a single computer. Moreover, decomposing process mining problems is not just interesting from a performance point of view. Decompositions can also be used to pinpoint the most problematic parts of the process (also in terms of performance) and provide localized diagnostics. This also helps us to better understand the limitations of existing conformance checking and process discovery techniques.

In this paper we discussed a very general divide-and-conquer approach without focusing on a particular representation or decomposition strategy. Nevertheless, it provided new and interesting insights with respect to the essential requirements of more concrete approaches. The paper also provides pointers to approaches using Petri nets as a representational bias and SESEs [10], [9], passages [8], [11], or maximal decompositions [7] as a decomposition strategy. It is clear that these are merely examples of the broad spectrum of possible techniques to decompose process mining problems. Given the incredible growth of event data, there is an urgent need to explore and investigate the entire spectrum in more detail.

Acknowledgements

This work was supported by the Basic Research Program of the National Research University Higher School of Economics (HSE). The author would like to thank Eric Verbeek, Jorge Munoz-Gama and Joseph Carmona for their joint work on decomposing Petri nets for process mining (e.g., using passages and SESEs).

References

[1] W. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011.

[2] ——, "Business Process Management: A Comprehensive Survey," *ISRN Software Engineering*, pp. 1–37, 2013, doi:10.1155/2013/507984.

[3] W. van der Aalst and K. van Hee, *Workflow Management: Models, Methods, and Systems*. MIT press, Cambridge, MA, 2004.

[4] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. MIT press, Cambridge, MA, 2001.

[5] T. Mitchell, *Machine Learning*. McGraw-Hill, New York, 1997.

[6] W. van der Aalst, "Distributed Process Discovery and Conformance Checking," in *International Conference on Fundamental Approaches to Software Engineering (FASE 2012)*, ser. Lecture Notes in Computer Science, J. Lara and A. Zisman, Eds., vol. 7212. Springer-Verlag, Berlin, 2012, pp. 1–25.

[7] ——, "Decomposing Petri Nets for Process Mining: A Generic Approach," BPM Center Report BPM-12-20 (accepted for Distributed and Parallel Databases), BPMcenter.org, 2012.

[8] ——, "Decomposing Process Mining Problems Using Passages," in *Applications and Theory of Petri Nets 2012*, ser. Lecture Notes in Computer Science, S. Haddad and L. Pomello, Eds., vol. 7347. Springer-Verlag, Berlin, 2012, pp. 72–91.

[9] J. Munoz-Gama, J. Carmona, and W. van der Aalst, "Conformance Checking in the Large: Partitioning and Topology," in *International Conference on Business Process Management (BPM 2013)*, ser. Lecture Notes in Computer Science, F. Daniel, J. Wang, and B. Weber, Eds., vol. 8094. Springer-Verlag, Berlin, 2013, pp. 130–145.

[10] ——, "Hierarchical Conformance Checking of Process Models Based on Event Logs," in *Applications and Theory of Petri Nets 2013*, ser. Lecture Notes in Computer Science, J. Colom and J. Desel, Eds., vol. 7927. Springer-Verlag, Berlin, 2013, pp. 291–310.

[11] H. Verbeek and W. van der Aalst, "Decomposing Replay Problems: A Case Study," BPM Center Report BPM-13-09, BPMcenter.org, 2013.

[12] OMG, "Business Process Model and Notation (BPMN)," Object Management Group, formal/2011-01-03, 2011.

[13] W. van der Aalst, A. Weijters, and L. Maruster, "Workflow Mining: Discovering Process Models from Event Logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 9, pp. 1128–1142, 2004.

[14] W. van der Aalst, V. Rubin, H. Verbeek, B. van Dongen, E. Kindler, and C. Günther, "Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting," *Software and Systems Modeling*, vol. 9, no. 1, pp. 87–111, 2010.

[15] R. Agrawal, D. Gunopulos, and F. Leymann, "Mining Process Models from Workflow Logs," in *Sixth International Conference on Extending Database Technology*, ser. Lecture Notes in Computer Science, vol. 1377. Springer-Verlag, Berlin, 1998, pp. 469–483.

[16] R. Bergenthum, J. Desel, R. Lorenz, and S. Mauser, "Process Mining Based on Regions of Languages," in *International Conference on Business Process Management (BPM 2007)*, ser. Lecture Notes in Computer Science, G. Alonso, P. Dadam, and M. Rosemann, Eds., vol. 4714. Springer-Verlag, Berlin, 2007, pp. 375–383.

[17] J. Carmona and J. Cortadella, "Process Mining Meets Abstract Interpretation," in *ECML/PKDD 210*, ser. Lecture Notes in Artificial Intelligence, J. Balcazar, Ed., vol. 6321. Springer-Verlag, Berlin, 2010, pp. 184–199.

[18] J. Carmona, J. Cortadella, and M. Kishinevsky, "A Region-Based Algorithm for Discovering Petri Nets from Event Logs," in *Business Process Management (BPM2008)*, 2008, pp. 358–373.

[19] J. Cook and A. Wolf, "Discovering Models of Software Processes from Event-Based Data," *ACM Transactions on Software Engineering and Methodology*, vol. 7, no. 3, pp. 215–249, 1998.

[20] S. Goedertier, D. Martens, J. Vanthienen, and B. Baesens, "Robust Process Discovery with Artificial Negative Events," *Journal of Machine Learning Research*, vol. 10, pp. 1305–1340, 2009.

[21] A. Medeiros, A. Weijters, and W. van der Aalst, "Genetic Process Mining: An Experimental Evaluation," *Data Mining and Knowledge Discovery*, vol. 14, no. 2, pp. 245–304, 2007.

[22] M. Sole and J. Carmona, "Process Mining from a Basis of Regions," in *Applications and Theory of Petri Nets 2010*, ser. Lecture Notes in Computer Science, J. Lilius and W. Penczek, Eds., vol. 6128. Springer-Verlag, Berlin, 2010, pp. 226–245.

[23] A. Weijters and W. van der Aalst, "Rediscovering Workflow Models from Event-Based Data using Little Thumb," *Integrated Computer-Aided Engineering*, vol. 10, no. 2, pp. 151–162, 2003.

[24] J. van der Werf, B. van Dongen, C. Hurkens, and A. Serebrenik, "Process Discovery using Integer Linear Programming," *Fundamenta Informaticae*, vol. 94, pp. 387–412, 2010.

[25] A. Rozinat and W. van der Aalst, "Conformance Checking of Processes Based on Monitoring Real Behavior," *Information Systems*, vol. 33, no. 1, pp. 64–95, 2008.

[26] W. van der Aalst, A. Adriansyah, and B. van Dongen, "Replaying History on Process Models for Conformance Checking and Performance Analysis," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 182–192, 2012.

[27] A. Adriansyah, B. van Dongen, and W. van der Aalst, "Conformance Checking using Cost-Based Fitness Analysis," in *IEEE International Enterprise Computing Conference (EDOC 2011)*, C. Chi and P. Johnson, Eds. IEEE Computer Society, 2011, pp. 55–64.

[28] ——, "Towards Robust Conformance Checking," in *BPM 2010 Workshops, Proceedings of the Sixth Workshop on Business Process Intelligence (BPI2010)*, ser. Lecture Notes in Business Information Process-

ing, M. Muehlen and J. Su, Eds., vol. 66. Springer-Verlag, Berlin, 2011, pp. 122–133.

[29] A. Adriansyah, N. Sidorova, and B. van Dongen, "Cost-based Fitness in Conformance Checking," in *International Conference on Application of Concurrency to System Design (ACSD 2011)*. IEEE Computer Society, 2011, pp. 57–66.

[30] T. Calders, C. Guenther, M. Pechenizkiy, and A. Rozinat, "Using Minimum Description Length for Process Mining," in *ACM Symposium on Applied Computing (SAC 2009)*. ACM Press, 2009, pp. 1451–1455.

[31] J. Cook and A. Wolf, "Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model," *ACM Transactions on Software Engineering and Methodology*, vol. 8, no. 2, pp. 147–176, 1999.

[32] J. Munoz-Gama and J. Carmona, "A Fresh Look at Precision in Process Conformance," in *Business Process Management (BPM 2010)*, ser. Lecture Notes in Computer Science, R. Hull, J. Mendling, and S. Tai, Eds., vol. 6336. Springer-Verlag, Berlin, 2010, pp. 211–226.

[33] ——, "Enhancing Precision in Process Conformance: Stability, Confidence and Severity," in *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011)*, N. Chawla, I. King, and A. Sperduti, Eds. Paris, France: IEEE, April 2011, pp. 184–191.

[34] J. Weerdt, M. De Backer, J. Vanthienen, and B. Baesens, "A Robust F-measure for Evaluating Discovered Process Models," in *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011)*, N. Chawla, I. King, and A. Sperduti, Eds. Paris, France: IEEE, April 2011, pp. 148–155.

[35] A. Polyvyanyy, J. Vanhatalo, and H. Völzer, "Simplified Computation and Generalization of the Refined Process Structure Tree," in *WS-FM 2010*, ser. Lecture Notes in Computer Science, M. Bravetti and T. Bultan, Eds., vol. 6551. Springer-Verlag, Berlin, 2011, pp. 25–41.

[36] J. Vanhatalo, H. Völzer, and J. Koehler, "The Refined Process Structure Tree," *Data and Knowledge Engineering*, vol. 68, no. 9, pp. 793–818, 2009.

[37] W. van der Aalst, "Passages in Graphs," BPM Center Report BPM-12-19, BPMcenter.org, 2012.

[38] C. Günther and W. van der Aalst, "Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics," in *International Conference on Business Process Management (BPM 2007)*, ser. Lecture Notes in Computer Science, G. Alonso, P. Dadam, and M. Rosemann, Eds., vol. 4714. Springer-Verlag, Berlin, 2007, pp. 328–343.

[39] U. Feige, M. Hajiaghayi, and J. Lee, "Improved Approximation Algorithms for Minimum-Weight Vertex Separators," in *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*. ACM, New York, 2005, pp. 563–572.

[40] G. Karpis and V. Kumar, "A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.

[41] B. Kernighan and S. Lin, "An Efficient Heuristic Procedure for Partitioning Graphs," *The Bell Systems Technical Journal*, vol. 49, no. 2, 1970.

[42] M. Kim and K. Candan, "SBV-Cut: Vertex-Cut Based Graph Partitioning Using Structural Balance Vertices," *Data and Knowledge Engineering*, vol. 72, pp. 285–303, 2012.

[43] H. Dhama, "Quantitative Models of Cohesion and Coupling in Software," *Journal of Systems and Software*, vol. 29, no. 1, pp. 65–74, 1995.

[44] IEEE Task Force on Process Mining, "Process Mining Manifesto," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, F. Daniel, K. Barkaoui, and S. Dustdar, Eds., vol. 99. Springer-Verlag, Berlin, 2012, pp. 169–194.

[45] P. Darondeau, "Unbounded Petri Net Synthesis," in *Lectures on Concurrency and Petri Nets*, ser. Lecture Notes in Computer Science, J. Desel, W. Reisig, and G. Rozenberg, Eds., vol. 3098. Springer-Verlag, Berlin, 2004, pp. 413–438.

[46] W. van der Aalst, K. van Hee, J. van der Werf, and M. Verdonk, "Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor," *IEEE Computer*, vol. 43, no. 3, pp. 90–93, 2010.

[47] H. Reguieg, F. Toumani, H. M. Nezhad, and B. Benatallah, "Using MapReduce to Scale Events Correlation Discovery for Business Processes Mining," in *International Conference on Business Process Management (BPM 2012)*, ser. Lecture Notes in Computer Science, A. Barros, A. Gal, and E. Kindler, Eds., vol. 7481. Springer-Verlag, Berlin, 2012, pp. 279–284.

[48] C. Bratosin, N. Sidorova, and W. van der Aalst, "Distributed Genetic Process Mining," in *IEEE World Congress on Computational Intelligence (WCCI 2010)*, H. Ishibuchi, Ed. Barcelona, Spain: IEEE, July 2010, pp. 1951–1958.

[49] J. Carmona, J. Cortadella, and M. Kishinevsky, "Divide-and-Conquer Strategies for Process Mining," in *Business Process Management (BPM 2009)*, ser. Lecture Notes in Computer Science, U. Dayal, J. Eder, J. Koehler, and H. Reijers, Eds., vol. 5701. Springer-Verlag, Berlin, 2009, pp. 327–343.

[50] D. Fahland, M. de Leoni, B. van Dongen, and W. van der Aalst, "Conformance Checking of Interacting Processes with Overlapping Instances," in *Business Process Management (BPM 2011)*, ser. Lecture Notes in Computer Science, S. Rinderle, F. Toumani, and K. Wolf, Eds., vol. 6896. Springer-Verlag, Berlin, 2011, pp. 345–361.

[51] M. Boukala and L. Petrucci, "Towards Distributed Verification of Petri Nets properties," in *Proceedings of the International Workshop on Verification and Evaluation of Computer and Communication Systems (VECOS'07)*. British Computer Society, 2007, pp. 15–26.

[52] C. Lakos and L. Petrucci, "Modular Analysis of Systems Composed of Semiautonomous Subsystems," in *Application of Concurrency to System Design (ACSD2004)*. IEEE Computer Society, 2004, pp. 185–194.

[53] M. Cannataro, A. Congiusta, A. Pugliese, D. Talia, and P. Trunfio, "Distributed Data Mining on Grids: Services, Tools, and Applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 34, no. 6, pp. 2451–2465, 2004.

[54] R. Agrawal and J. Shafer, "Parallel Mining of Association Rules," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, pp. 962–969, 1996.

[55] H. Verbeek and W. van der Aalst, "An Experimental Evaluation of Passage-Based Process Discovery," in *Business Process Management Workshops, International Workshop on Business Process Intelligence (BPI 2012)*, ser. Lecture Notes in Business Information Processing, M. Rosa and P. Soffer, Eds., vol. 132. Springer-Verlag, Berlin, 2013, pp. 205–210.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 11–16

Nonnegative Matrix Factorization and Its Application to Pattern Analysis and Text Mining

Jacek M. Zurada
Electrical and Computer Engineering
University of Louisville
Louisville, USA
Spoleczna Akademia Nauk,
90-011 Lodz, Poland
jacek.zurada@louisville.edu

Tolga Ensari
Computer Engineering
Istanbul University
Istanbul, Turkey
ensari@istanbul.edu.tr

Ehsan Hosseini Asl
Electrical and Computer Engineering
University of Louisville
Louisville, USA
ehsan.hosseiniasl@gmail.com

Jan Chorowski
TTA Techtra
ul. Muchoborska 18
54-424 Wrocław
Poland
jan.chorowski@techtra.pl

***Abstract*—Nonnegative Matrix Factorization (NMF) is one of the most promising techniques to reduce the dimensionality of the data. This presentation compares the method with other popular matrix decomposition approaches for various pattern analysis tasks. Among others, NMF has been also widely applied for clustering and latent feature extraction. Several types of the objective functions have been used for NMF in the literature. Instead of minimizing the common Euclidean Distance (EucD) error, we review an alternative method that maximizes the correntropy similarity measure to produce the factorization. Correntropy is an entropy-based criterion defined as a nonlinear similarity measure. Following the discussion of maximization of the correntropy function, we use it to cluster document data set and compare the clustering performance with the EucD-based NMF. Our approach was applied and illustrated for the clustering of documents in the 20-Newsgroups data set. The comparison is illustrated with 20-Newsgroups data set. The results show that our approach produces per average better clustering compared with other methods which use EucD as an objective function.**

***Keywords*—Nonnegative Matrix Factorization; Correntropy; Principal Component Analyis; Face recognition**

I. Introduction

The ever-increasing amount of data recorded, stored and processed worldwide necessitates the development of new representations and is becoming a major task for data analysis research [1, 2, 3, 4, 17, and 18]. Dimensionality reduction of the data is a technique that describes each multidimensional data sample with a small number of coefficients that are the sample's coordinates in a new, particular to this dataset, feature space. Often dimensionality reduction is accomplished by finding factorizations of a matrix representing the dataset. Most widely-known methods are Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Singular Value Decomposition (SVD). Recently defined Nonnegative Matrix Factorization (NMF) approach also has been successfully applied in pattern recognition. It is an unsupervised learning method that also reduces the dimensionality of the data. It has also been used for several applications [5-8, 14-16, 19, 20].

Matrix factorization methods treat the data as an $m \times n$ matrix in which every column represents a data sample.

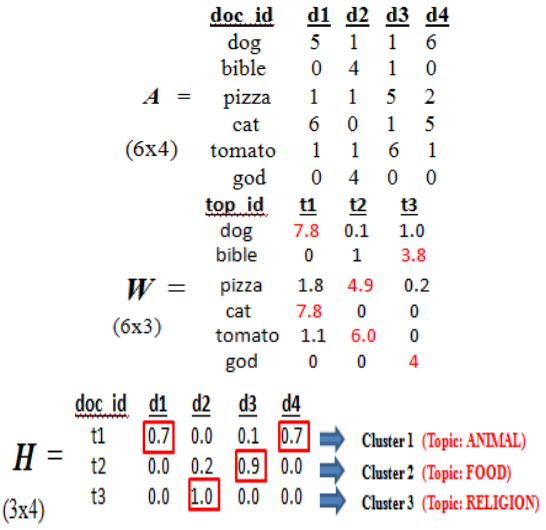


Figure 1. A small document-term matrix and its nonnegative factorization. Indicated are terms that are most important for each topic and assignments of documents into topics (clusters).

This matrix is approximated by a product of two rank k matrices, as follows:

$$A \approx WH,$$

where A is the data matrix, W is the $m \times k$ matrix of basis vectors and H is a $k \times n$ matrix that gives the coordinates of samples in the feature space. We can think of the factorization as of a decomposition of the j-th sample (j-th column of A, $A_{:j}$) into a linear combination of features given by columns of W:

$$A_{:j} = \sum_i W_{:i} H_{ij}.$$

For instance, consider a small artificial data set of documents shown in Figure 1. The documents are encoded in the bag-of-words format. Here, A is a 6×4 data matrix formed by 4 documents, 6 words of interest, and 3 topics. The topics are ANIMAL (d1 and d4), RELIGION (d2) an d FOOD (d4). Once the factorization is computed we can cluster the documents [14]. It can be accomplished by assigning each document to the topic that contributes the

most to its nonnegative representation (has the largest entry in the matrix H). The matrices W and H that form a nonnegative factorization are shown in Figure 1. We have indicated the most important words for each topic and demonstrated the cluster memberships.

II. NONNEGATIVE MATRIX FACTORIZATION (NMF)

The intuitive definition of matrix factorization asks to find two matrices, W and H, whose product approximates a given matrix. Specific matrix factorization schemes are differentiated by the error function used to describe the quality of the approximation and by the constraints imposed on the elements of W and H. The family of nonnegative factorizations imposes that elements of W and H be nonnegative. A necessary prerequisite is that the data matrix A must also contain nonnegative elements only. Fortunately, this is often the case. In example documents in a bag-of-words format or images are non-negative. Application areas for NMF include face recognition, bioinformatics, text mining and audio (speech) processing [2, 3, 6-8, 14, 19]. Clustering task is also one of the main topics for NMF and it has been extensively applied and discussed in the literature [6, 8, 14-16].

Nonnegative factorizations are motivated by their enhanced interpretability. When subtraction is forbidden no cancelations occur in the topic (cluster) definitions and meanings can be deduced, as in the example shown in Figure 1. Indeed, for many applications subtraction is not meaningful. However, except for NMF, other matrix factorization methods generally allow the subtraction of values. These values can be faces, audio or gene expression levels according to application areas. But in these cases, basis values (for instance images for face recognition) are not physically intuitive.

Typically, in NMF the factorization objective is the Euclidean distance between the elements of A and the elements of WH (i.e. the Frobenius norm of the difference $A - WH$). This measure is well-known and often used in the literature. However, other distance (or similarity) measures can be used, and they will often produce different factorizations. In example, calculations derived from the Kullback-Leibler divergence have often been studied in the literature [1-4, 9, and 14]. Often the loss function is chosen to match a specific application domain. In [7], authors used the Itakura-Saito (IS) divergence as an objective function and in [5] authors used the β-divergence.

In [1-4], authors used a distance measure based on the Kullback-Leibler divergence. The measure $D_s(u_j, w_i)$ is a symmetric divergence of u_j with respect w_i given by [2]:

$$D_s(u_j, w_i) = D(u_j||w_i) + D(u_j||w_i),$$

where:

$$D(x||z) = \sum_l \frac{x(l)}{\|x\|_1} \log\left(\frac{x(l)\|z\|_1}{z(l)\|x\|_1}\right).$$

Another distance measure suitable for the NMF is the correntropy function, described in details in Section III.

Extensions of the NMF methodology involve imposing other constraints on the matrices W and H, such as sparseness or orthogonality. Bayesian approaches and other conditions for factorization have also been considered [3, 8].

The typical algorithm used to compute the NMF factorization with the Euclidean distance measure begins with W and H randomly initialized. It then uses the multiplicative update rules to minimize the error function [1,4]:

$$\mathrm{H_{ij}} \leftarrow \mathrm{H_{ij}} \frac{(\mathrm{W^T A})_{\mathrm{ij}}}{(\mathrm{W^T WH})_{\mathrm{ij}}}$$

$$\mathrm{W_{ij}} \leftarrow \mathrm{W_{ij}} \frac{(\mathrm{AH^T})_{\mathrm{ij}}}{(\mathrm{WHH^T})_{\mathrm{ij}}}$$

The rules ensure that at each iteration the error function does not increase, while the matrices W and H stay non-negative. The rules are applied iteratively until convergence.

Faster converging alternatives to the multiplicative updates, that have been proposed for the NMF include the projected gradient descent (PGD) and the alternating least squares (ALS) algorithm [2, 16].

III. CORRENTROPY SIMILARITY MEASURE

We have recently proposed to use the correntropy similarity measure as an objective function for nonnegative matrix factorization [26, 27]. The correntropy is a localized similarity measure between two random variables that was proposed in [9-12, 14]. It can be used as a cost function for NMF. We use it to calculate the element-wise similarity between the matrix A and its factorization:

$$Corr(A, WH) = \sum_{\mathrm{i,j}} \exp\left(\frac{-\left(A_{ij} - (WH)_{ij}\right)^2}{2\sigma^2}\right) \quad (1)$$

where σ is a parameter of the correntropy similarity measure. We note that for NMF we need to minimize the negative of correntropy since it is a similarity and not a distance measure [14].

It can easily be seen from eq. 1 that $Corr(A, WH)$ is always bounded and nonnegative. Moreover, the correntropy saturates when the disagreement between elements of A and its factorization WH is large. This property is important. It makes correntropy insensitive to outliers, because errors for badly approximated elements have less influence on the factorization. We illustrate correntropy as the error surface in Figure 2. It shows the errors for a single element of $1 + Loss(A, WH)$. We can change the shape of the function and control the level of saturation by adjusting the parameter σ. When σ is large little saturation occurs. Lowering σ causes that more and more elements of the difference $A - WH$ saturate and are treated as outliers.

IV. EXEMPLARY APPLICATIONS OF NMF

A. Document Clustering with NMF

For the first real life example we report the result of a comparison between quality of NMF factorizations based on the Euclidean distance and based on correntropy [14, 26]. The evaluation analyses the quality of clusters computed from factorizations. We have used the 20-newsgroups data set, which is one of the popular benchmarks used for clustering and classification of the text data. It has approximately 11,000 documents taken from 20 different newsgroups pertaining to various subjects.

After the factorization process, we obtain W and H. H can be used to group the data (A) into r clusters by choosing the largest value of each column in H.

The 20 newsgroups data contains ground-truth document labels which can be used to evaluate the quality of the clustering. We evaluate the clustering performance with the entropy measure. Total entropy for a set of clusters is calculated as the weighted mean of the entropies of each cluster weighted by the size of each cluster. Firstly, we calculate the distribution of the data for each cluster. For class j we compute p_{ij}, the probability that a member of cluster i belongs to class j as $p_{ij} = m_{ij}/m_i$, where m_i is the number of objects in cluster i and m_{ij} is the number of objects of class j in cluster i. Entropy of each cluster i is defined as:

$$e_i = -\sum_{j=1}^{L} p_{ij} log_2(p_{ij}),$$

where L is the number of classes. Entropy of the full data set as the sum of the entropies of each cluster i weighted by the size of each cluster:

$$e = \sum_{i=1}^{K} \frac{m_i}{m} e_i$$

where K is the number of clusters and m is the total number of data points [24].

Table 1 shows the entropy values of *NMF-PGD (EucD)* and *NMF-Corr* approaches for 20-Newsgroups data set. We graph these values (*NMF-PGD (EucD)* and *NMF-Corr (for* $\sigma = 1$, $\sigma = 0.5$ and $\sigma = 0.01$)) in Figure 3. Here, "k" denotes the assumed number of clusters and equals to the ranks of W, H. We change it from 2 to 20 to track the clustering performance. We show all entropy values in Figure 3, but for brevity we only illustrate 10 data points in Table 1. Since lower entropy values indicate better clustering performance, it can be seen from Table 1 and Figure 3, that *NMF-Corr* ($\sigma = 0.5$) demonstrates superior clustering performance than *NMF-PGD (EucD)* for every evaluated number of clusters.

Experiments and comparative results between NMF-PGD (EucD) and NMF-Corr show that NMF-Corr ($\sigma = 0.5$) has better clustering performance than NMF-PGD (EucD). Therefore, we can conclude that correntropy-based

Table 1. Entropy of 20-Newsgroups data set with NMF-PGD (EucD) and NMF-Corr.

Number of Clusters (k)	NMF-PGD (EucD)	NMF-Corr *(σ = 1)*	**NMF-Corr *(σ = 0.5)***	NMF-Corr *(σ = 0.01)*
r = 2	3.84	3.86	**3.85**	4.30
r = 3	3.86	3.79	**3.58**	4.27
r = 4	3.78	3.49	**3.50**	4.27
r = 5	3.74	3.60	**3.38**	4.24
r = 6	3.49	3.36	**3.30**	4.23
r = 7	3.44	3.28	**3.26**	4.20
r = 8	3.30	3.26	**2.94**	4.19
r = 9	3.30	3.34	**3.13**	4.18
r = 10	3.16	3.23	**2.93**	4.20

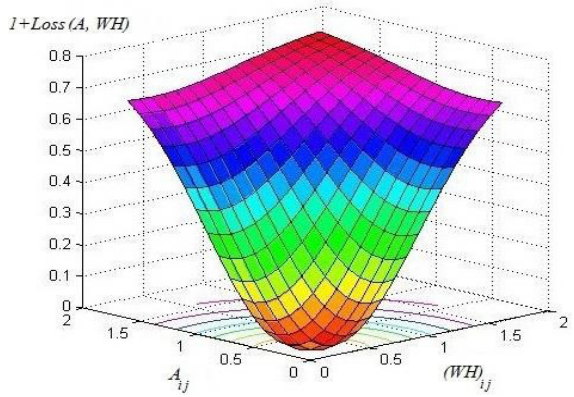


Figure 2. Correntropy objective function 1+Loss(A,WH) with σ=0.5,m=n=1.

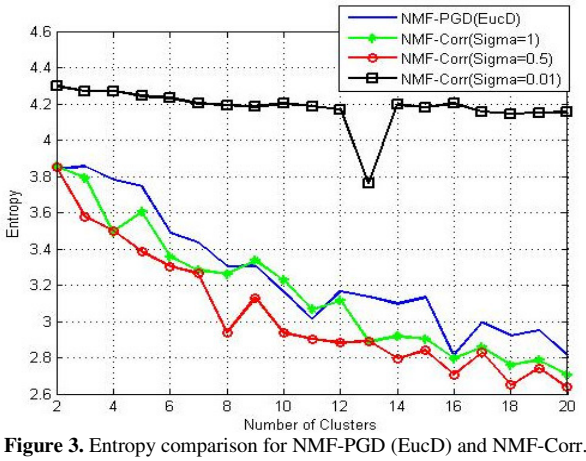


Figure 3. Entropy comparison for NMF-PGD (EucD) and NMF-Corr.

NMF ($for\ \sigma = 0.5$) has comparatively better clustering performance vs. EucD-based NMF for the evaluated data set. However, NMF-Corr does not show improved performance for $\sigma = 1$ and specifically worst performance for $\sigma = 0.01$. This can be seen from Figure 3 and Table 1. Also, the deterioration of clustering results for σ values below 0.5 requires further studies. One additional question

is whether this dependence on σ value is a property of the method or else whether it lies in the properties of the data for which experiments have been conducted. This will warrant further studies.

B. Occluded Face Recognition Using NMF

In the second example we report the results of an application of NMF to the problem of occluded face recognition [26].

Face recognition is one of the well-studied real life problems. Several methods have been defined and applied for this task. Above mentioned methods and Neural Networks (NN) have been studied to recognize face images [1, 19-26]. In fact, faces are not clear for daily life, because some obstacles can be in front of the face. These obstacles can be scarf, glasses, hats or some occlusion on the face. Therefore, occluded face recognition is important area in pattern analysis. There are many studies in the literature for occluded face recognition task, especially using PCA and NMF [19-26].

In this section, we evaluate the recognition performance of occluded face images on ORL face data set. We have compared PCA, NMF and correntropy based NMF (NMF-Corr) formulations by evaluating quality of recognition rates computed from factorizations. The ORL data consists of 40 persons, each photographed in 10 different poses. The data set was partitioned into two equal parts for training and testing. We have resized face images from original 112x92 pixels to 56x46 pixels for efficient computation.

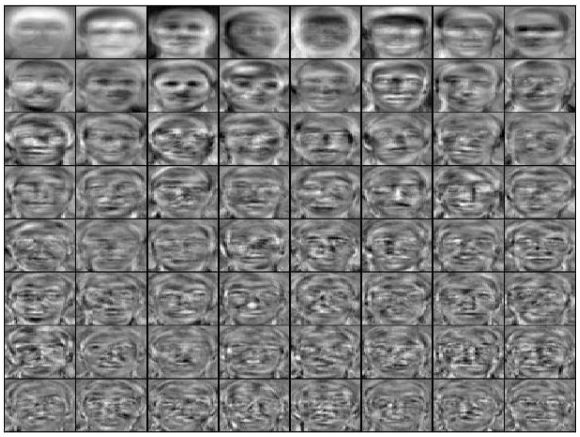
Face recognition in the NMF and NMF-Corr linear subspace is performed by first computing the pseudo-inverse of the W matrix as $W^+ = W^T(WW^T)^{-1}$. Then, all samples were encoded using this pseudo inverse. Finally, we have used 1 nearest-neighbor (1-NN) classifier for the recognition process.

In order to generate occluded faces, we have used randomly located black patches for both training and testing face images. In this way we test the robustness of the compared dimensionality reduction methods to noise on both training and testing data. Each patch covers from 10% to %50 of the face image at a random location. Sample patched face images can be seen from Figure 4.

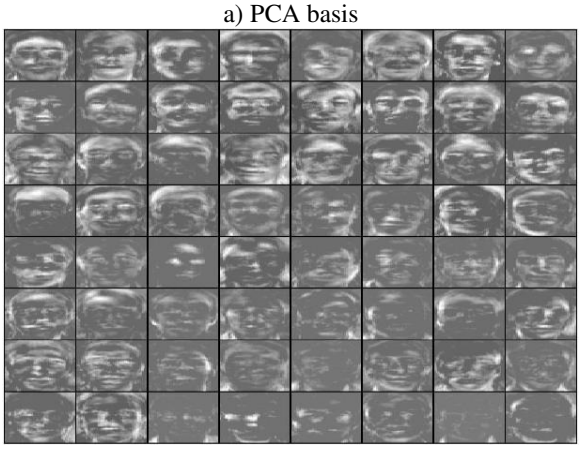
Recognition results have been obtained by running each method (PCA, NMF and NMF-Corr) 10 times, and then average recognition rate has been calculated.

NMF-Corr and NMF algorithms were run with the random initial matrices W and H. For NMF-Corr, we set stopping criteria at most 1000 iterations and relative tolerance 10^{-4}. PCA, NMF and NMF-Corr basis images has been shown in Figure 5, respectively.

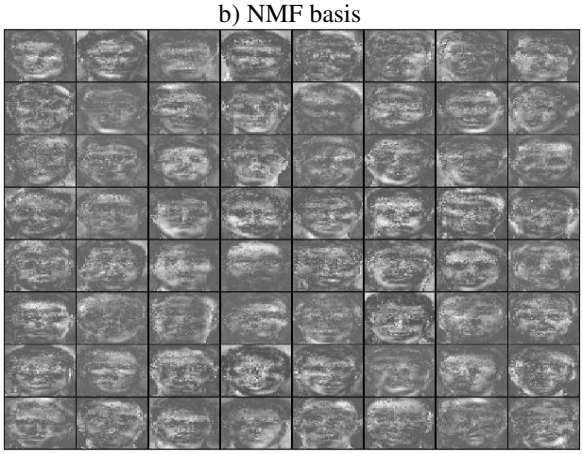
Figure 4 Randomly located occluded face samples from ORL face dataset with 10%, 20%, 30%, 40% and 50% patch sizes (From left to right).



a) PCA basis



b) NMF basis



c) NMF-Corr basis

Figure 5. Basis images of PCA, NMF and NMF-Corr for 64 grids.

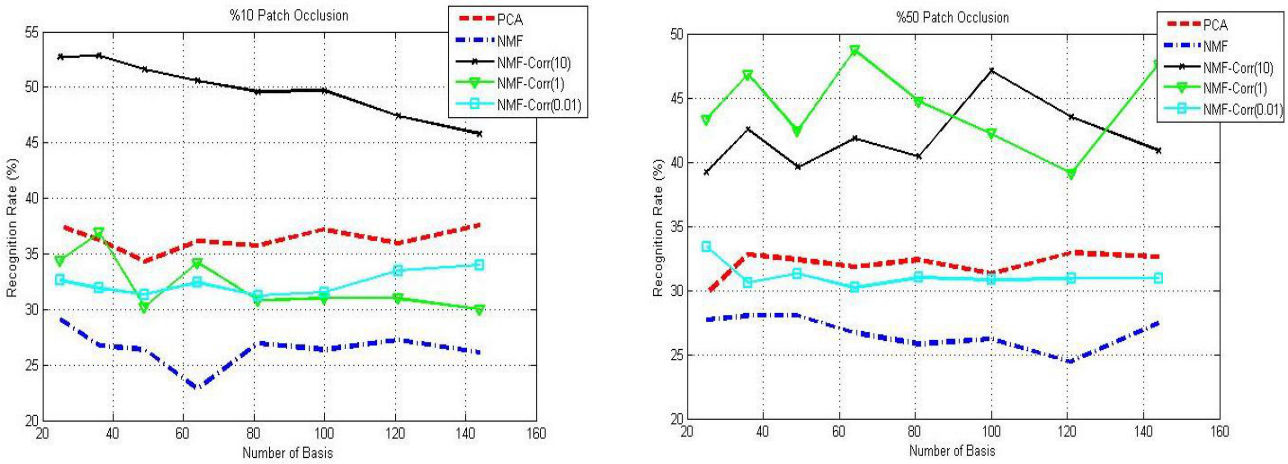


Figure 6. Recognition rates (%) versus number of basis images for 10% and 50% patch occlusions (On the legend, values in paranthesis indicate the corresponding σ parameter for NMF-Corr).

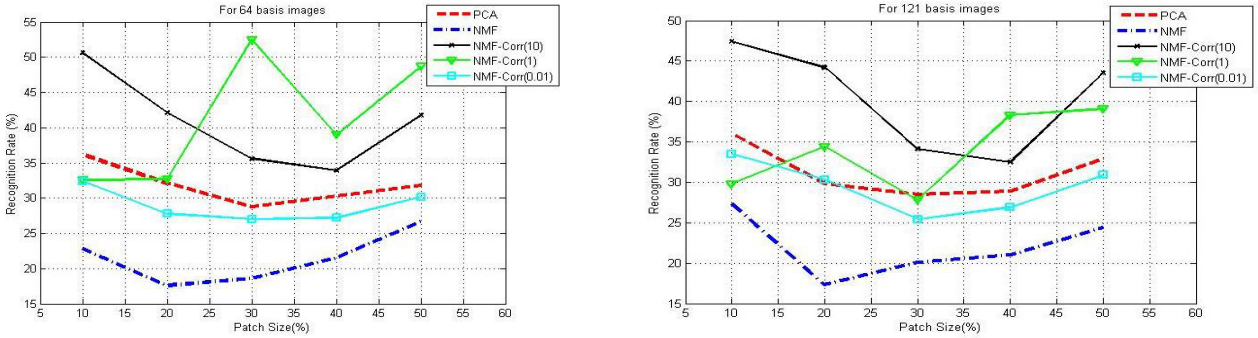


Figure 7. Recognition rates (%) versus patch sizes (%) of face images for 64 and 121 basis images.

For brevity we only illustrate for 10% and 50% occlusions in Figure 6, for different number of basis images. It can be easily seen that NMF-Corr with $\sigma = 10$ demonstrates superior recognition performance than NMF and PCA. Therefore, NMF-Corr with $\sigma = 10$ has the best accuracy. (In the case of 50% patch occlusion, generally $\sigma = 1$ has better accuracy than $\sigma = 10$). Recognition rate plots versus patch occlusion sizes have been also calculated for 25, 36, 49, 64, 81, 100, 121 and 144. Again, we only demonstrate for 64 and 121 basis in Figure 7 for brevity. Here, NMF-Corr has the best recognition rate for all patch sizes. Additionally, it can be seen from Figure 7, the graphic lines are u-shaped, because training and testing parts have been done with occluded face images.

V. CONCLUSION

In this contribution we have first introduced the topic of nonnegative matrix factorization and reviewed its major applications and implementations. The NMF factorizes a given data matrix into a product of two matrices that contain nonnegative elements only. Subtraction is forbidden which enhances sparsity of the patterns that are found in the data. This leads to a better interpretability of the factorization.

The usefulness of nonnegative factorizations was demonstrated using two real-life tasks: document clustering and occluded face recognition. Moreover the demonstrations used correntropy, a novel similarity measure that enhances the robustness to outliers. Experiments on both datasets have shown that using the correntropy criterion has led to better cluster purity and recognition rates than NMF and PCA.

ACKNOWLEDGMENTS

This paper is an extended version of [14] and [26].

REFERENCES

[1] Lee D., Seung H. S. “Learning the Parts of Objects with Nonnegative Matrix Factorization”, Nature, Vol. 401, pp. 788-791, 1999.

[2] Berry, M. W., Browne M., Langville A. N., Pauca V. P., Plemmons R. J. , “Algorithms and Applications for Approximate Nonnegative Matrix Factorization”, Computational Statistics and Data Analysis, Vol. 52, No. 1, pp. 155–173, 2007.

[3] Hoyer P. O., “Non-negative Matrix Factorization with Sparseness Constraints”, Journal of Machine Learning Research 5, pp. 1457-1469, 2004.

[4] Lee D., Seung H. S., “Algorithms for Non-negative Matrix Factorization”, Advances in Neural Information Processing, Vol. 13, pp. 556-562, 2001.

[5] Fevotte C. and Idier J. “Algorithms for Nonnegative Matrix Factorization with the β- Divergence”, Neural Computation, Vol. 13, Issue 3, pp. 1-24, 2010.

[6] Zhao W., Ma H., Li N., “A Nonnegative Matrix Factorization Algorithm with Sparseness Constraints”, Int. Conf. on Machine Learning and Cybernetics, Guilin, China, July 10-13, 2011.

[7] Fevotte C., Bertin N., Durrieu J. L., “Nonnegative Matrix Factorization with the Itakura-Saito Divergence, Neural Computation, Vol. 21, pp. 793-830, 2009.

[8] Shahnaz F., Berry M. W., Pauca V. P., Plemmons R. J. “Document Clustering Using Nonnegative Matrix Factorization”, Int. Journal of Information Processing and Management, Vol. 42, Issue 2, pp. 373-386, 2006.

[9] He R., Zheng W. S., Hu B. G., “Maximum Correntropy Criterion for Robust Face Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 8, pp. 1561-1576, 2011.

[10] Liu W., Pokharel P. P., Principe J. C.: Correntropy, “Properties and Applications in Non-Gaussian Signal Processing”, IEEE Transactions on Signal Processing, Vol. 55, No. 11, pp. 5286-5298, 2007.

[11] He R., Hu B. G., Zheng W. S., Kong X. W., “Robust Principal Component Analysis Based on Maximum Correntropy Criterion”, IEEE Transactions on Image Processing, Vol. 20, No. 6, 2011.

[12] Chalasani R., Principe J. H., “Self Organizing Maps with Correntropy Induced Metric”, Int. Joint Conf. on Neural Networks, Spain, pp. 1-6, 2010.

[13] Matlab Software, www.di.ens.fr/~mschmidt/Software/minConf.html

[14] Ensari T, Chorowski J, Zurada J. M., “Correntropy-based Document Clustering via Nonnegative Matrix Factorization", Int. Conf. on Artificial Neural Networks (ICANN), Lausanne, Switzerland, September 11-14, 2012.

[15] Zhao W., Ma H., Li N., “A New Non-negative Matrix Factorization Algorithm with Sparseness Constraints, Proc. of the 2011 Int. Conf. on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011.

[16] Lin C. J., “Projected Gradient methods for Non-Negative Matrix Factorization”, Neural Computation, 19:2756-2779, 2007.

[17] Tan P., Steinbach M., Kumar V., “Introduction to Data Mining”, Pearson Addison Wesley, 2006.

[18] P. Paatero, “Least Squares Formulation of Robust Non-negative Factor Analysis”, Chemometrics and Intelligent Laboratory Systems 37, 23-35, 1997.

[19] Wang Y., Jia Y., “Non-Negative Matrix Factorization Frame for Face Recognition”, Int. Journal of Pattern Recognition and Artificial Intelligence, Vol. 19. No.4, pp. 495-511, 2005.

[20] Byeon W., Jeon M., “Face Recognition Using Region-based Nonnegative Matrix Factorization”, Communications in Computer and Information Science, Vol. 56, pp. 621-628, 2009.

[21] Feng T., Li S. Z., Shum H. Y., Zhang H. J., “Local Non-negative Matrix Factorization as a Visual Perpeption”, Int. conf. on Development and Learning, June 12-15, 2002.

[22] Shastri B. J. and Levine M. D., “Face Recognition Using Localized Features Based on Non-Negative Sparse Coding”, Machine Vision and Applications, Vol. 18, No. 2, pp. 107-122, 2007.

[23] Oh H. J., Lee K. M., Lee S. U., “Occlusion Invariant Face Recognition Using Selective Local Non-negative Matrix Factorization Basis Images”, Image and Vision Computing, Vol. 26, Issue 11, pp. 1515-1523, November 2008.

[24] Pan J. Y., Zhang J. S., “Large Margin Based Nonnegative Matrix Factorization and Partial Least Squares Regression for Face Recognition”, Pattern Recognition Letters, Vol. 32, pp. 1822-1835, 2011.

[25] Liu H., Wu Z., Li X., Cai D., Huang T. S., “Constrained Nonnegative Matrix Factorization for Image Representation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 7, July 2012.

[26] Ensari T., Chorowski J., Zurada J. M., “Occluded Face Recognition Using Correntropy-based Nonnegative Matrix Factorization”, International Conference on Machine Learning and Applications (ICMLA), Boca Raton, Florida, USA, December 12-15, 2012.

8th International Symposium
Advances in Artificial Intelligence and Applications

THE AAIA'13 will bring researchers, developers, practitioners, and users to present their latest research, results, and ideas in all areas of artificial intelligence. We hope that theory and successful applications presented at the AAIA'13 will be of interest to researchers and practitioners who want to know about both theoretical advances and latest applied developments in Artificial Intelligence. As such AAIA'13 will provide a forum for the exchange of ideas between theoreticians and practitioners to address the important issues.

Papers related to theories, methodologies, and applications in science and technology in this theme are especially solicited. Topics covering industrial issues/applications and academic research are included, but not limited to:

-
- Knowledge Management
- Decision Support Systems
- Approximate Reasoning
- Fuzzy Modeling and Control
- Data Mining
- Web Mining
- Machine Learning
- Combining Multiple Knowledge Sources in an Integrated Intelligent System
- Neural Networks
- Evolutionary Computation
- Nature Inspired Methods
- Natural Language Processing
- Image Processing and Interpreting
- Applications in Bioinformatics
- Hybrid Intelligent Systems
- Granular Computing
- Architectures of Intelligent Systems
- Robotics
- Real-world Applications of Intelligent Systems
- Rough Sets

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 19–25

Underdetermined Blind Separation of an Unknown Number of Sources Based on Fourier Transform and Matrix Factorization

Ossama S. Alshabrawy*, Mohamed E. Ghoneim*
Damietta University, faculty of Science, Department of Mathematics & Computer Science, 34517, Damietta, Egypt

A. A. Salama
Port Said University, faculty of Science, Department of Mathematics & Computer Science, Port Said, Egypt

Aboul Ella Hassanien*
Cairo University, faculty of Computers and Information, Department of Information Technology, Cairo, Egypt

* Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net

***Abstract*—This paper presents an approach for underdetermined blind source separation that can be applied even if the number of sources is unknown. Moreover, the proposed approach is applicable in the case of separating I+3 sources from I mixtures without additive noise. This situation is more challenging and suitable to practical real world problems. Also, the sparsity conditions are not imposed unlike to those employed by some conventional approaches. Firstly, the number of source signals are estimated followed by the estimation of the mixing matrix based on the use of short time Fourier transform and rough-fuzzy clustering. Then, source signals are normalized and recovered using modified Lin's projected gradient algorithm with modified Armijo rule. The simulation results show that the proposed approach can separate I+3 source signals from I mixed signals, and it has superior evaluation performance compared to conventional approaches.**

***Keywords*—Underdetermined Blind Source Separation; Rough Fuzzy clustering; Short Time Fourier transform; Lin's Projected Gradient; Armijo rule**

I. Introduction

BLIND Signal Separation (or Blind Source Separation, BSS) has received a great deal of attention in the fields of digital communication systems, speech processing, medical imaging, water marking, biomedical engineering, and data mining [3]-[7] in recent years in combination with artificial neural networks, information theory, and computer science applications. Blindness or blind separation means that no or very little information is known about the source signals or the mixing system [1].

The objective of BSS is to extract original source signals using only the information gathered from observed signals with no or very limited knowledge about the source signals or the mixing system. The approaches developed by researchers in the last few years can be classified into two methodologies, namely over-determined BSS and underdetermined BSS, according to the number of source signals and observable mixed signals [20]. BSS that has fewer sensors or observable mixed signals than source signals is called underdetermined BSS while a BSS that has more sensors than sources is called over-determined BSS, Underdetermined BSS is challenging and is more realistic in practical situations. However, most approaches for BSS rarely involve underdetermined BSS cases. The classical independent component analysis (ICA) approach fails to solve underdetermined BSS problems [10]. Moreover in many practical problems, there are a large number of source signals but a few numbers of sensors that means the underdetermined case. Another major difficulty of ICA is that the mixing matrix and the magnitude of original source signals cannot be estimated due to its ambiguities and that the order, sign, and the variances of the independent components cannot be determined [2].

Most of the current traditional BSS methods assume that the source signals are as statistically independent as possible given the observed data and that the mixing matrix is of full column rank. In many real-world situations, however, this hypothesis is not valid. Consequently, recovering the source signals by multiplying the observable data mixtures by the pseudo inverse of the mixing matrix cannot be used. This makes recovering the source signals a very challenging task [8]. In practical terms, the over-determined mixture assumption does not always hold (e.g., in radio communications the probability of receiving more sources than sensors increases with increase of reception bandwidth), thus it is necessary to solve the problem of underdetermined blind source separation (UBSS) [9].

Nonnegative Matrix Factorization (NMF) has been widely applied to BSS problems. However, the separation results are sensitive to the initialization of parameters, also the additive parts by NMF are not necessarily localized, and consequently the solution is not unique. Avoiding the subjectivity of choosing parameters, we use general matrix factorization (GMF), which completely relaxes the non-negativity constraints from its factors with the Alternative Least Squares (ALS) method as an initialization to the

Ossama S. Alshabrawy. Corresponding autor.
Tel. +2-012-211-76423
Email. ossama_alshabrawy87@yahoo.com

source signals instead of random initial values. GMF is a generalization of the well-known NMF where the NMF is constrained by non-negativity on all its factors, is not necessarily localize, has low convergence and, does not provide a unique solution in some cases without additive constraints and parameters. However, GMF has no constraints of non-negativity and is fast convergent with the ALS method used for initialization and improvement.

The motivation of this research is to separate sparse, and super and sub-Gaussian signals in the underdetermined case with an unknown number of source signals without resorting to any sparsity conditions, and to increase the performance of the separation.

The rest of the paper is organized as follows. Section II formulates the problem. In Section III, we introduce an overview, background, and the basic concepts of Projected Gradient and GMF, alternative least square, and rough fuzzy clustering. In Section IV, we present the details of the proposed approach. In section V, we show the analysis of typical experiments and the results obtained by different BSS methods, where the simulation results show the effectiveness and high performance of the proposed algorithm. Finally, a short conclusion and future work are presented in Section 6.

II. Problem Statement

The problem considered in this paper is an underdetermined instantaneous BSS with an unknown number of source signals but without background noise, which can be mathematically formulated as follows:

Assume that for I unobservable components $X(t) = tr[X_1(t), X_2(t), ..., X_J(t)]$, where J is the number of source signals, and *X(t)* is a zero-mean vector. The available sensor vector $Y(t) = tr[Y_1(t), Y_2(t), ..., Y_I(t)]$, where I is the number of sensors and *tr* is the transpose of the vector, is given by

$$Y(t) = AX(t).$$

Here $A \in R^{I \times J}$ is a non-singular and unobservable matrix and has a non-zero determinant, and the rank of A is I. $X \in R^{J \times T}$, $Y \in R^{I \times T}$. t=0,….,T-1 are the sampling instant time points.

III. Preliminary Topics

This section provides a brief explanation of the basic technologies used in this paper including projected gradient and GMF, alternative least squares, and rough fuzzy clustering.

A. Projected Gradient and General Matrix Factorization

GMF is a generalization of NMF where there are no nonnegative constraints on all of the factors [12] and is the focus of a great deal of attention in Mathematics and Computer Science. NMF has been widely used in many areas including BSS [11], [13], [14]. However, the solution is not unique since NMF is non-convex programming, and in most algorithms it frequently converges to local optima. Unlike NMF, GMF is convergent and has good local optima avoidance when initialized with ALS. In this paper, GMF is regarded as a good tool for solving the problem of UBSS. The novelty in this paper is that GMF is to solve the UBSS problem for the first time.

The basic GMF decomposition model for BSS is as follows:

$$Y = AX \qquad (1)$$

where, $Y \in R_{\pm}^{I \times T}$ represents the observable mixtures, $A \in R_{\pm}^{I \times J}$ is the mixing matrix, and $X \in R_{\pm}^{J \times T}$ is the source signals matrix. Hence, Y, A, and X have both signs unlike NMF where Y, A, and X are non-negative. For BSS, I is the number of mixtures or sensors, T is the number of sample time points, and J is the number of sources. With only the data observable mixtures (Y) as the only known variable, the mixing matrix A and the source signals X are estimated using Equation (2).

We will use the projected gradient based update rules in GMF. These updates take the following generalized form of iterative rules [11]:

$$X^{(n+1)} = X^{(n)} - \alpha_X P_X \qquad (2)$$

$$A^{(n+1)} = A^{(n)} - \alpha_A P_A \qquad (3)$$

where P_A and P_X are the descent directions, and α_A and α_x are the learning rates, of A and X respectively.

The projected gradient algorithms for GMF are based on the alternating minimization technique which can be written in the matrix form as follows:

$$\min_{x_{ij}} Cost(Y \,||\, AX) = \frac{1}{2}||Y - AX||_F^2 \qquad (4)$$

$$\min_{a_{ij}} Cost(Y^T \,||\, X^T A^T) = \frac{1}{2}||Y^T - X^T A^T||_F^2 \qquad (5)$$

Basically, the matrix A is assumed to be full rank. Consequently, this provides the existence of a unique solution $X^* \in R^{J \times T}$. The gradient matrix for A and X is given by the following equations:

$$Grad_X(X) = \nabla_X Cost(Y \,||\, AX) = A^T(AX - Y) \qquad (6)$$

$$Grad_A(A) = \nabla_A Cost(Y^T \,||\, X^T A^T) = (AX - Y)X^T \qquad (7)$$

One of the projected gradients based approaches, and will be applied in this paper in a modified version, is Lin's projected gradient algorithm [15]. Lin's projected gradient (LPG) algorithm can be induced by the iterative formulas (2) and (3) with P_A and P_X expressed by the equations (6) and (7). Moreover, the projection on the subspace of non-negative real numbers is not considered.

B. Alternative least squares (ALS)

The minimization of cost function in equations (4) and (5) which represent the standard squared Euclidean distance can be formulated as follows:

$$\begin{aligned} Cost(Y \parallel AX) &= \frac{1}{2}\|Y - AX\|_F^2 \\ &= \frac{1}{2} tr(Y - AX)(Y - AX) \end{aligned} \tag{8}$$

where *tr* stands for the transpose of the matrix. The above cost function can be alternately minimized with respect to the two factors A and X [11]. Moreover, each time during the optimization process of one factor while keeping the other one fixed [18],[19] and finding the stationary or critical points, which are obtained by equating the gradients to zero. This corresponds to the following two minimization problems:

$$\begin{aligned} &A^{(k+1)} = \min_A \|Y - AX^{(k)}\|_F^2, \text{ and} \\ &X^{(k+1)} = \min_X \|Y^T - X^T[A^{(k+1)}]^T\|_F^2 \end{aligned} \tag{9}$$

The gradients after equating them by zero according to the Karush-Kuhn-Tucker (KKT) optimality conditions are:

$$\frac{\partial D_F(Y \parallel AX)}{\partial a_{ij}} = [-YX^T + AXX^T]_{ij} = 0,$$

$$\frac{\partial D_F(Y \parallel AX)}{\partial x_{ij}} = [-A^T Y + A^T AX]_{ij} \; \forall ij. \tag{10}$$

Consequently,

$$A = YX^T(XX^T)^{-1} \text{ and, } X = (A^T A)^{-1} A^T Y. \tag{11}$$

This method will be used as an initialization in our proposed system.

C. Rough fuzzy clustering

In fuzzy c-means (FCM) algorithm developed by Dunn in 1973, improved by Bezdek in 1981, and is the best known method for fuzzy clustering, based on optimizing objective function, the concept of traditional k-means clustering algorithm is extended which for each data point a degree of membership or membership function $\zeta_{ij} \in [0,1]$ of clusters is calculated.

$$\zeta_{ij} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/\delta - 1}} \tag{12}$$

, where δ is the degree of fuzziness.

In contrast to fuzzy clustering, in rough c-means (RCM), the concept of k-means is extended by considering each cluster as an interval or rough set Y [16]. It is characterized by the lower approximation $\underline{BY}$ and the upper approximations $\overline{BY}$ with the following properties: (i) an object or a sample y_k can be part of at most one lower approximation; (ii) if $y_k \in \underline{BY}$ of cluster X, then simultaneously $y_k \in \overline{BY}$; and (iii) if y_k is not a part of any lower approximation, then it belongs to two or more upper approximations. This permits overlaps between clusters.

A rough–fuzzy c-means algorithm which involves the integration of fuzzy and rough sets has been developed [17]. This allows incorporating the fuzzy membership value ζ_{ij} of a sample y_k to cluster center β_i. Moreover, instead of absolute individual distance d_{ik} from the centroid, the membership to the cluster center β_i is relative to the other centers $\beta_j \; \forall i \neq j$. Consequently, the robustness of the clustering will be enhanced with respect to different choices of the parameters. The centroid β_i of cluster U_i can be determined by the following equation:

$$\beta_i = \begin{cases} Z & \text{if } \overline{BU_i} - \underline{BU_i} \neq \varnothing, \; \underline{BU_i} \neq \varnothing, \\ \dfrac{\sum_{y_k \in (\overline{BU_i} - \underline{BU_i})} y_k}{|\overline{BU_i} - \underline{BU_i}|} & \text{if } \overline{BU_i} - \underline{BU_i} \neq \varnothing, \underline{BU_i} = \varnothing, \\ \dfrac{\sum_{y_k \in (\underline{BY})} y_k}{|\underline{BU_i}|} & \text{otherwise}. \end{cases} \tag{13}$$

where,

$$Z = W_{upper} \frac{\sum_{y_k \in (\overline{BU_i} - \underline{BU_i})} y_k}{|\overline{BU_i} - \underline{BU_i}|} + W_{lower} \frac{\sum_{y_k \in (\underline{BU_i})} y_k}{|\underline{BU_i}|}$$

The algorithm of rough fuzzy c-means is stated below in Algorithm 1.

Algorithm 1 Rough fuzzy c-means clustering

Step1: Assign initial means β_i for c clusters.

Step 2: Compute the fuzzy membership ζ_{ij} for c clusters and N data objects according to equation (12) and Normalize the distances used for fuzzy membership in [0,1].

Step 3: Assign each data object y_k to the lower or upper approximation of cluster pair U_i and U_j.

Step 4: Compute the difference $\zeta_{ik} - \zeta_{jk}$ to cluster centroids β_i and β_j

Step 5: Let ζ_{ik} be maximum and ζ_{jk} be the next to maximum

Step 6: If $abs(\zeta_{ik} - \zeta_{jk})$ is less than some *threshold*

Then,

$y_k \in \overline{BU_i}$ and $y_k \in \overline{BU_i}$ and y_k cannot be a member of any lower approximation,

Else,

$y_k \in \underline{BU_i}$ such that the membership ζ_{ik} is the maximum over c clusters
Step 7: Compute the new centroid for each cluster using equation (13)
Step 8: Repeat Steps 2–7 until convergence or until there are no new assignments.

IV. Proposed UBSS Algorithm

In this section, the proposed approach is presented briefly starting with estimating the mixing matrix knowing only the observable mixtures matrix. Also, a method for GMF gradient-based update rules initialized with ALS is introduced.

A. Mixing matrix estimation based on short time Fourier transform and rough fuzzy clustering

Conventional algorithms estimate the mixing matrix based on clustering algorithms such as the k-means algorithm require that the source signals to be very sparse in the time domain and this is unavailable in many practical real world problems. Other algorithms are based on an assumption that there exist many TF points of single source occupancy (SSO), or require that there exists at least one small region in the TF plain with only a single source and such a TF region must exist for each source. All aforementioned approaches require that for each source there exist many TF points of SSO. However, single source detection (SSD) requires that there exists at least one TF point of SSO and is hence less restrictive than the other approaches [20].

The short time Fourier transform (STFT) of the i^{th} observed signal is defined by the following equation:

$$Y_i^{Fourier}(t,r) = \sum_{l=0}^{\infty} h(l-t) X_i(l) e^{-jrl} \qquad (14)$$

at frame t and frequency bin r where $h(l)$ is a window sequence. In equation (14), $i=1,2,...,I$; $t=0,1,...,T-1$ are the sampling points over the time domain and $r=0,1,...,T-1$ are the sampling points over the frequency domain. The SSD is based on the ratio of the TF transforms and finds a set of TF points where a single source is active for each source. Therefore, for a given $\varepsilon > 0$, a set

$$\chi_F = \left\{ (t,r) \mid \| \mathrm{Im}\left[\frac{Y^S(t,r)}{Y_1^S(t,r)} \right] \|_F < \varepsilon,\ Y_1^S(t,r) \neq 0 \right\} \qquad (15)$$

where, Im[ ] denotes the imaginary part. We can choose any of the mixture instead of Y_1.

During clustering the observable mixtures after incorporating STFT, we need to determine the number of source signals. Since there is an overlap between the data objects, estimating the number of sources require an efficient validity index [21] and can be given by the following equation:

$$V(\beta, c) = Scat(c) + \frac{Sep(c)}{Sep(C_{\max})} \qquad (16)$$

where, β is the cluster centers, c is the number of clusters, and C_{max} is the chosen maximum number of clusters. Here,

$$Scat(c) = \frac{\frac{1}{c}\sum_{i=1}^{c} \| \sigma(\beta_i) \|}{\| \sigma(Y) \|} \qquad (17)$$

Also, the value of *Scat(c)* varies from 0 to 1. The term that represents the separation between clusters is defined by

$$sep(c) = \frac{D_{\max}^2}{D_{\min}^2} \sum_{i=1}^{c} \left(\sum_{j=1}^{c} \| \beta_i - \beta_j \|^2 \right)^{-1} \qquad (18)$$

where,

$D_{\min} = \min_{i \neq j} \| \beta_i - \beta_j \|, D_{\max} = \max_{i,j} \| \beta_i - \beta_j \|$

After clustering, and determining the number of source signals, the ith column vector of A, denoted as $\hat{a}_i$, is estimated as

$$\hat{a}_i = \frac{1}{|\chi_{C_i}|} \sum_{(t,r) \in \chi_{C_i}} \mathrm{Re}[Y^F(t,r)]. \qquad (19)$$

Here $|\chi_{C_i}|$ represents the number of TF points in cluster C_i for $i=1,2,...,J$.

Algorithm 2 Mixing matrix estimation and determining the number of source signals

Input: the observable mixtures $Y = [Y_1, Y_2, \ldots\ldots, Y_I]$
Output: number of source signals, the mixing matrix A

Step1: Calculate STFT Y using equations (14)
Step 2: Calculate χ_F using equation (15)
Step 3: Cluster χ_S using rough fuzzy c-means clustering stated in Algorithm 1 for different number of clusters by choosing C_{min}, C_{max} (i.e. min and max chosen number of clusters, respectively) using equations (16)-(18) and the cluster number that minimizes V is considered to be the optimal value for number of source signals.
Step 4: Determine the TF points and their quantity in each cluster.
Step 5: Calculate the columns of the mixing matrix A using equation (19)

B. Lin's Projected Gradient (LPG) with Armijo rule based GMF

In Lin's projected gradient algorithm the learning rates α_A and α_x are not fixed diagonal matrices in the inner iterations but are scalars. These learning rates are computing by inexact estimation techniques. Lin considered two options to estimate the learning rates. The first option is the Armijo rule along the projective arc of the algorithm proposed by Bertsekas [23]. The value of the learning rate α_x, for every iterative step of the algorithm, is given by:

$$\alpha_x^{(k)} = \rho^{m_k}, \qquad (20)$$

where m^k is the first non-negative integer m for which

$$Cost(Y \| AX^{(k+1)}) - Cost(Y \| AX^{(k)}) \leq \upsilon tr\left\{ \nabla_X Cost(Y \| AX^{(k)})^T \left(X^{(k+1)} - X^{(k)} \right) \right\} \qquad (21)$$

with $\rho \in (0,1) and\ \upsilon \in (0,1)$. The value of the learning rate α_A is computed in a similar way.

The second option is the modified Armijo rule. Lin and More [24] noticed that α_A and α_x might be very similar, and they proposed to start from $\alpha_x^{(k-1)}$ and to increase or decrease the learning rate according to condition (20). Here in this paper LPG algorithm with Armijo rule is extended, different from [11], for general matrix factorization relaxing the non-negativity constraints. Moreover the value of ρ and υ are changed to be $\rho \in (-1,1) and\ \upsilon \in (-1,1)$. The algorithm of the modified LPG algorithm with Armijo rule is listed below in Algorithm 3.

Algorithm 3 Modified LPG based GMF

Input: the observable mixtures Y= [Y1, Y2,……..,Yt], number of components (source signals) J, Maximum number of iterations N, the mixing matrix A estimated from algorithm 2

Output: the estimated source signals X

Step 1: Initialize the matrix X by ALS according to equation (11)

Step 2: set α_x =1

Step 3: Assign $X^{(n+1)} = X^{(n)} - \alpha_X P_X$

Step 4: Repeat 4-10 until the stopping criteria is met

Step 5: If condition in (21) is met, then

Step 6: Repeat steps 5,6 until condition (21) does not hold

Step 7: Assign $\alpha_x = \frac{\alpha_x}{\rho}$

Step 8: Update $X^{(n+1)} = X^{(n)} - \alpha_X P_X$

Step 9: Else

Repeat steps 8,9 until condition (21) is met

Step 10: Set $\alpha_x = \alpha_x \rho$

Step 11: Update $X^{(n+1)} = X^{(n)} - \alpha_X P_X$

The source signals are then rescaled and normalized

V. Experiments and Simulation

In this section, the effectiveness of the proposed approach will be discussed by comparing results of experiments and stimulations. Experiments and simulations were performed on synthetically generated signals using the proposed approach and other conventional approaches. In the simulations, sparse, super- and sub-Gaussian signals were separated from the underdetermined mixtures in the challenging case where the true number of source signals is unknown.

The parameter inputs of the modified LPG algorithm are the observable mixtures matrix Y, and the mixing matrix A obtained from algorithm 2. We choose the maximum number of iterations to be only 25 iterations. We investigate the performance of the proposed UBSS approach in the above mentioned cases by comparing its results with the results of approaches in Khor (2006) [22], Kim and Yoo (2009) [20], Xiang and Peng (2010) [8]. Here, the simulation of the separation of sparse and Gaussian signals is provided followed by some discussion. Then, the cases of a variety of sparse, non-sparse, and super- and sub-Gaussian signals are stated.

A. Sparse and Gaussian signals

The separation of J synthetic Gaussian and sparse signals from I=3 mixtures was performed in the time domain for J=4, 5, and 6 source signals. In this simulation, the mixing matrix was estimated using algorithm 2. The proposed approach was compared to the abovementioned algorithms. The simulation settings were as follows. Synthetic sparse signals were generated by generating 5000 Gaussian samples using the *randn* command of Matlab and substituting 80% of the samples chosen randomly by zeros for each source. The results show that the proposed approach can separate I+3 source signals from I mixtures, unlike the previous approaches. This is confirmed in the next simulations. The analysis aims at comparing mainly the reconstruction index Signal-to-Interference Ratio (SIR) to evaluate the performance of the proposed approach. Given original source signals X and its estimations $\hat{X}$ obtained by the proposed approach, SIR in decibels is defined as

$$SIR = -10\log\left(\frac{\| \hat{X}_i - X_i \|_F^2}{\| X_i \|_F^2}\right), \quad i = 1,2,....,J \qquad (15)$$

Fig. 1 illustrates the averaged SIRs when the number of the sources increases from 4 to 6 signals.

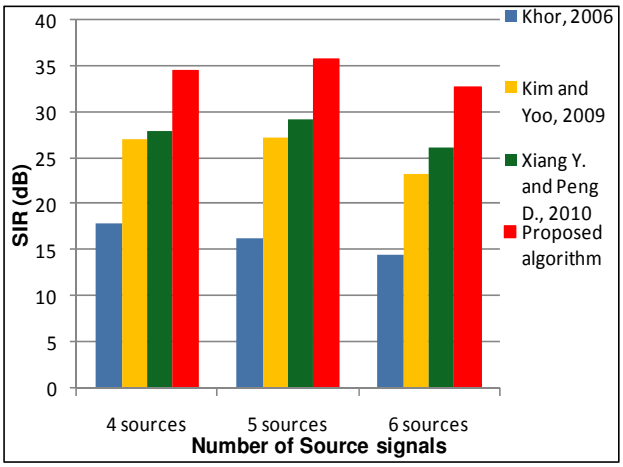


Fig. 1. performance estimation of the source signals according to the number of sources from 3 observable mixtures (in case of Sparse and Gaussian sources).

B. Synthetic signals

We investigated the effectiveness of the proposed UBSS approach by comparing with the methods mentioned above. We chose the number of mixtures to be only 2 and the number of sources to be 5 in order to prove that the proposed algorithm can separate I+3 source signals from I mixtures. The mixture signals that we perform our experiments on are mixed by the following randomly generated mixing matrix:

$$\begin{bmatrix} 0.5377 & -2.2588 & 0.3188 & -0.4336 & 3.5784 \\ 1.8339 & 0.8622 & -1.3077 & 0.3426 & 2.7694 \end{bmatrix}$$

The mixing matrix was once again estimated using algorithm 2. The true and estimated values of A are shown in Table 1. The six source signals, two observable mixtures, and estimated source signals are plotted in Fig. 2. The number of sampling time points is 10,000. The simulation results of the proposed approach in addition to those of the five different UBSS methods are shown in Fig.4. The performance of the source recovery method can be evaluated by Eqs. (15) and (16).

$$SNR = \frac{1}{J}\sum_i 10\log\left(\frac{\| X_i \|_F^2}{\| \hat{X}_i - X_i \|_F^2}\right) \qquad (16)$$

Where J is the number of sources and $\|.\|_F^2$ is the Frobenius norm. The efficiency of the separation results is good when $SNR \geq 25$ [10].

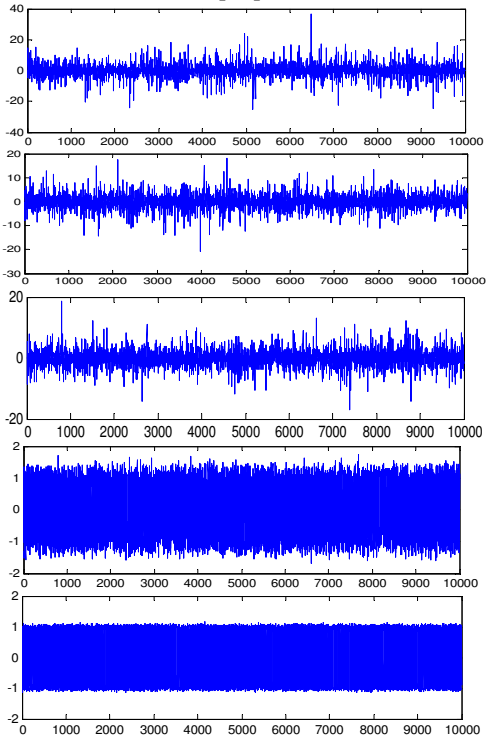
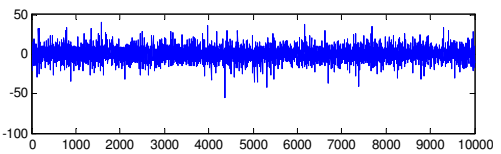


Fig. 2. (a) Source signals



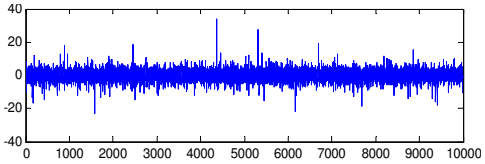


Fig. 2. (b) Observable mixtures

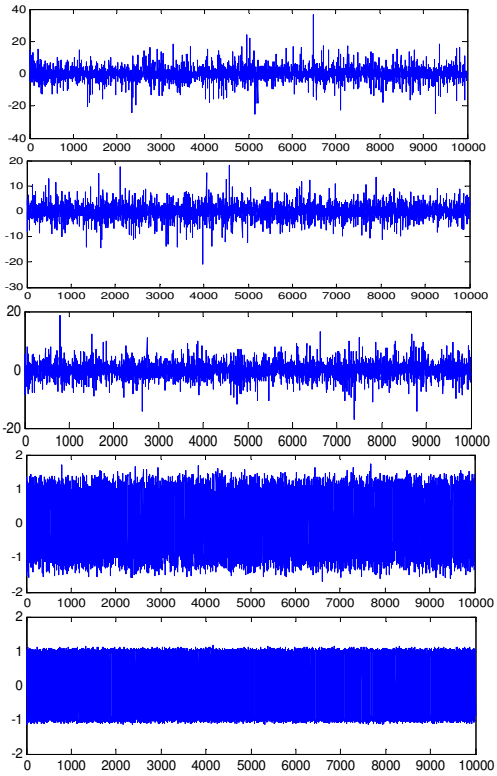


Fig. 2. (c) Estimated source signals

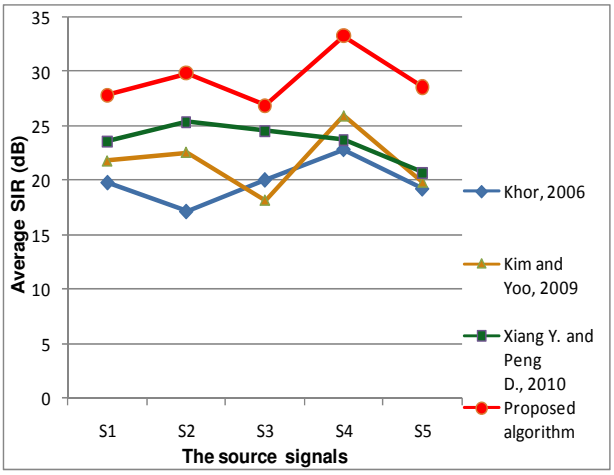


Fig. 3. Performance estimation of the source signals from 2 observable mixtures.

We note from Fig. 3 that the proposed approach achieves about 5.7 dB higher SIR for J=5 sources with only two mixtures than the highest performance algorithm among the other three approaches. Another comparison of the proposed approach with the other five approaches is presented using

the following examples with 2 observable mixtures where X_i is the chosen source signal shown in Fig. 4.

Example 1. $X=\{X_1, X_2, X_5\}$

Example2. $X=\{X_1, X_2, X_3, X_5\}$

Example 3. $X=\{X_1, X_2, X_3, X_4, X_5\}$

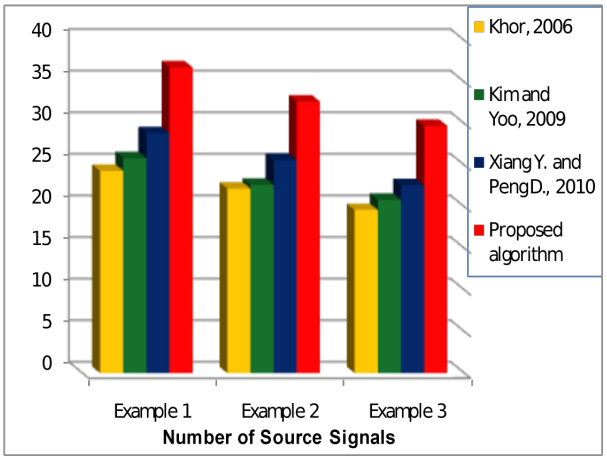


Fig. 4. Performance estimation of Examples 1–3 from 2 observable mixtures.

From the results in Figs. 3and 4, we can conclude that the separation performance of the proposed approach is very high, has faster convergence, and can separate I+3 source signals from I mixed signals.

II. Conclusion

In this paper, we addressed the problem of underdetermined blind source separation with the challenging case that the true number of source signals is unknown. A new two-step approach for optimum estimation of the source signals without additive noise. In this approach, STFT is combined with rough fuzzy c-means clustering to estimate the mixing matrix and determine the number of source signals. Then the source signals are estimated by a modified LPG algorithm with Armijo rule based general matrix factorization. Simulation experiments demonstrated the validity and superior performance of the proposed approach.

References

[1] Ossama S. Alshabrawy, M. E. Ghoniem, W. A. Awad and Aboul ella hassanien, Underdetermined Blind Source Separation based on Fuzzy C-Means and Semi-Nonnegative Matrix Factorization. *IEEE FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS*, Wroclaw, Poland, 9-12 September, 2012, pp. 695-700

[2] A. Hyvarinen, J. Karhunen, E. Oja, "Independent Component Analysis, Wiley," New York, 2001.

[3] Yadong, Liu, Zongtan, Zhou, Dewen, Hu. "A novel method for spatio temporal pattern analysis of brain fMRI data," Science in China Series F: Information Sciences, vol. 48, no. 2, pp. 151–160, 2005.

[4] Araki, S., Makino, S., Blin, A., "Underdetermined blind separation for speech in real environment with sparseness and ICA," In Proceedings of the ICASSP'04, Montreal, Canada, pp.881–884, 2004.

[5] Ohnishi, Naoya, Imiya, Atsushi, "Independent component analysis of optical flow for robot navigation," Neurocomputing, vol.7, nos. 10–12, pp. 2140–2163, 2008.

[6] Tonazzini, Anna, Bedini, Luigi, Salerno, Emanuele, "A Markov model for blind image separation by a mean-field EM algorithm," IEEE Transactions on Image Processing, vol. 15, no.2, pp.473–482, 2005.

[7] Er-Wei, Bai, QingYu, Li, Zhiyong ,Zhang, "Blind source separation channel equalization of nonlinear channels with binary inputs," IEEE Transactions on Signal Processing, vol. 53, no. 7, pp.2315–2323, 2005.

[8] Dezhong Peng, Yong Xiang, "Underdetermined blind separation of nonsparse sources using spatial time-frequency distributions," Digital Signal Processing, vol. 20, pp. 581–596, 2010.

[9] Fengbo Lu, Zhitao Huang,Wenli Jiang, "Underdetermined blind separation of non-disjoint signals in time–frequency domain based on matrix diagonalization," Signal Processing, vol. 91, pp. 1568–1577, January 2011.

[10] Chaozhu Zhang, Cui Zheng, "Underdetermined Blind Source Separation Based on Fuzzy C-Means Clustering and Sparse Representation," International Conference on Graphic and Image Processing (ICGIP), Proc. of SPIE vol. 8285, 2011.

[11] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, Shun-ichi Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley, 2009

[12] Vladimir Nikulin, Tian-Hsiang Huangb, Shu-Kay Ngc, Suren I. Rathnayake, Geoffrey J. McLachlan, "A very fast algorithm for matrix factorization," Statistics and Probability Letters, vol. 81, pp. 773–782, February 2011.

[13] Tuomas Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization with Temporal Continuity and Sparse Criteria," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 3, March 2007.

[14] A. Cichocki, R. Zdunek, S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation, " In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2006, vol. 5, pages 621–624, Toulouse, France, May 14–19, 2006.

[15] Ch-J. Lin. Projected gradient methods for non-negative matrix factorization. Neural Computation, 19(10), pp. 2756–2779, October 2007.

[16] Sushmita Mitra, Witold Pedryczb, Bishal Barman, "Shadowed c-means: Integrating fuzzy and rough clustering," Pattern Recognition, vol. 43, pp. 1282–1291, 2010.

[17] S. Mitra, H. Banka, W. Pedrycz, "Rough–fuzzy collaborative clustering, IEEE Transactions on Systems," Man, and Cybernetics—Part B36, pp. 795-805, 2006.

[18] A. Cichocki, S. Amari. *Adaptive Blind Signal and Image Processing*, John Wiley & Sons Ltd, New York, 2003.

[19] D.D. Lee, H.S. Seung. "Algorithms for Nonnegative Matrix Factorization," vol. 13, MIT Press, 2001.

[20] SangGyun Kim, Chang D. Yoo, "Underdetermined Blind Source Separation Based on Subspace Representation," IEEE Transaction on Signal Processing, vol. 57, no. 7, July 2009.

[21] Sun, Haojun, Wang, Shengrui, Jiang, Qiangshan. FCM-based model selection algorithms for determining the number of clusters. Pattern Recognition 37 (10), pp. 2027–2037, 2004.

[22] Khor, L.C. "Robust adaptive blind signal estimation algorithm for under- determined mixture," IEE Proceedings—Circuits, Devices and Systems, vol. 153, no. 4, pp. 320–331, 2006.

[23] D. P. Bertsekas. Nonlinear Programming. Athena Scientific, Belmont, MA, 1999.

[24] Ch. J. Lin and J.J. Mor´e. Newton's method for large bound-constrained optimization problems. SIAM Journal on Optimization, 9(4), pp. 1100–1127, 1999.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 27–34

The Multiple Pheromone Ant Clustering Algorithm and its application to real world domains

Jan Chircop
Aston University,
Birmingham, United Kingdom
Email: janchircop@gmail.com

Christopher D. Buckingham
Aston University,
Birmingham, United Kingdom
Email: C.D.Buckingham@aston.ac.uk

***Abstract*—The Multiple Pheromone Ant Clustering Algorithm (MPACA) models the collective behaviour of ants to find clusters in data and to assign objects to the most appropriate class. It is an ant colony optimisation approach that uses pheromones to mark paths linking objects that are similar and potentially members of the same cluster or class. Its novelty is in the way it uses separate pheromones for each descriptive attribute of the object rather than a single pheromone representing the whole object. Ants that encounter other ants frequently enough can combine the attribute values they are detecting, which enables the MPACA to learn influential variable interactions. This paper applies the model to real-world data from two domains. One is logistics, focusing on resource allocation rather than the more traditional vehicle-routing problem. The other is mental-health risk assessment. The task for the MPACA in each domain was to predict class membership where the classes for the logistics domain were the levels of demand on haulage company resources and the mental-health classes were levels of suicide risk. Results on these noisy real-world data were promising, demonstrating the ability of the MPACA to find patterns in the data with accuracy comparable to more traditional linear regression models.**

I. Introduction

CLUSTERING is the task of partitioning data sets into categories of common likeness. It can be a complex problem to unravel because the boundaries between classes are often ambiguous and non-linear. If the data set has high dimensionality, it can be extremely difficult to understand the inherent structure and exploit it with an appropriate clustering algorithm. This has led to a large variety of approaches seeking to optimise cluster analysis, including ones modelled on insect behaviour.

This paper investigates how computer models of ants can help humans sort data into meaningful classes using cluster analysis. A brief review of relevant ant models is provided before explaining how the MPACA works. The main aim of the paper is to show how it can provide meaningful results with real-world data and examples from transport logistics and health are used. The paper concludes with a discussion of the model, its effectiveness, and how it can be applied to additional data sets.

II. Background

Swarm intelligence (SI) is the phenomenon whereby intelligent behaviour emerges from the interactions of numerous separate entities with low-level cognitive capacities [1], [4], [5], [6]. There are many examples in the insect world but the focus of this paper is on ants and specifically ant colony optimisation (ACO). Two ant behaviours have fuelled many of the computer models, one for sorting larvae or corposes and the other foraging for food. The so-called Basic Model (BM) [8] comes from the sorting of ant bodies into piles and is often referenced as the Standard Ant Clustering Algorithm (SACA). It works by assessing the similarity of bodies with others in the same location. In contrast, ant foraging depends on laying down pheromone trails that guide other ants towards objects in which they are interested. It is used to optimise paths between objects, either to link similar ones or to find the shortest paths. The MPACA is based on this type of ACO algorithm.

Using scents or pheromone to form paths is a form of stigmergy, where information is placed in the environment for communication purposes [7], [24]. Shorter paths have ants returning to them more quickly and the pheromone is less affected by evaporation. Together, these phenomena attract ants to locations containing objects with similar attributes and are the driving forces for cluster formation. For the MPACA model applied to real-world domains in this paper, objects are placed within a multidimensional graph space, as others have done [20], [26]. Its main innovation is by having multiple pheromones that distinguish ants within colonies rather than more normally between them [10]. The next section summarises the latest version of the MPACA, which was introduced as a clustering algorithm in [41] and applied to some standard data sets. The goal of this paper is to show how the MPACA can be used to learn class assignations and be applied to noisy, diverse real-world data in the domains of mental-health risk assessment and predicting resource requirements for logistics companies.

III. Overview of the MPACA

The Multiple Pheromone Ant Clustering Algorithm, MPACA, is not unique by having many different pheromones laid down on trails for objects. However, no previous models attach a specific pheromone to each particular value of every descriptive attribute of an object. Pheromones encourage other ants to follow them via a scent. In the MPACA model, each pheromone type indicates paths towards a specific feature value in the given search space. It is applicable to multiple dimensions and can accommodate both discrete and continu-

978-83-60810-53-8/$25.00

ous data of any type. Ordinal dimensions are used to set up the hyperdimensional problem space but are first normalised to help prevent bias due to types of distributions. The values are measured in the number of standard deviations (SD) from the mean, z, where

$$z = \frac{(x - \mu)}{SD}, \tag{1}$$

x, is the original value and μ, is the mean.

The values of each object along the dimensions of the hyperdimensional space determine its location in the graph and objects are linked by edges to all other objects if the Euclidean distance is within a parameterised maximum. Non-ordinal features are not part of the hyperdimensional space but still participate in learning by having ants leave pheromone traces along the edges corresponding to these features and their particular values. Ants are then placed on every object, with each ant assigned to one attribute and responding to the particular value the object has for that attribute. The ant's own attribute value becomes the distinctive pheromone it deposits, which it lays whenever leaving an object with a matching value and which it follows if laid by another ant. The upshot is that there will be as many pheromones in the domain as there are distinct attribute values, including nominal features and ordinal dimensions.

Learning takes place by ants following trails matching their own feature value and depositing pheromone from objects if they also have that value. Paths from an object are chosen stochastically based on the amount of matching pheromone compared to the alternatives. The resulting stigmergy means objects with similar feature values will have higher levels of pheromone connecting them because ants will be depositing in both directions along the path compared to ants travelling along edges with different values at each end. Also, the evaporation of pheromone means longer paths tend to have lower levels of pheromone.

The MPACA has a mechanism for ants to learn feature combinations or interactions that might be important for cluster analysis and classification. Pheromone trails for different feature values can draw ants into the same locations if they regularly co-occur with the same objects. If the encounters between the ants exceed a parameterised threshold, the ants will combine each other's features which means they will now only respond to objects having both features. It enables ants to detect feature combinations and thus to pick up non-linear interactions.

Merging colonies is similarly driven by the frequency of ant encounters. Both feature and colony merges are operationalised by recording ant meetings. These take place when an ant, which we will call the "focus" ant for referential clarity, has reached an object and only if its feature set matches that of the object because this means it is in an area of interest to it. The "encounters" data structure of the focus ant is updated at this point by finding all other ants within the vicinity that are also in an area of interest to them, which is the case if they are in deposit mode on a path away from the object that has the focus ant on it or are coming towards the object. If the number of encounters for the focus ant go above a threshold for colony or feature merging (they have different thresholds), then the focus ant updates its feature and colony properties accordingly. The thresholds have to be exceeded within a certain time span. The time span is measured in the number of steps, where each cycle of the system moves an ant one step along an edge and encounters recorded on a step that goes outside the time window are removed.

A. MPACA parameters

Although the general idea and philosophy of the MPACA has been described, much of the detail resides with how it is parameterised. This will be summarised here so that the actual values used when applying the model to the data can be understood.

- **Edges joining objects** Only ordinal dimensions are used to set up the hyperdimensional problem space. They are normalised as already explained by Equation 1. This gives the same units to all dimensions in the hyperdimensional space and object values are likewise normalised so that they can be placed in their appropriate location in the space. The resulting graph, G, has a vertex, v, for every object and all objects are connected to other objects by an edge, e, but only if is is within a distance parameter, d: relationships between objects further apart are therefore ignored.
- **Step size** The granularity for measuring differences between objects depends on how small the steps are along the edges. The assumption was made that plus or minus 2 SDs from the mean covers most population values on that dimension except the outliers. A step size of 0.1 SDs gives 40 steps along each dimension, which is enough for meaningful distinctions between objects.
- **Pheromone deposition, evaporation, and path choice** Pheromone is laid by ants when they leave an object with matching values and the same parameterised amount, ph, is laid for all ants and all features. A percentage is removed from paths by evaporation on each step and a maximum amount, $ph.max$, prevents paths increasing levels of pheromone indefinitely, which would overwhelm the influence of other paths.
  A residual parameter r, determines the percentage of total matching pheromone on all edges that is placed on each of them by default. It adds uncertainty by allowing ants to go down paths with little or no scent and explore new areas. Given N potential paths from a vertex with pheromone scent s on the first step of each path, where s is the pheromone matching the features of the ant, the probability of selecting a particular path, p, is given by
  $$P(p) = \frac{(s+r)}{\sum_{i=1}^{N} s_i + (r \times N)}.$$
- **Detection range for continuous dimensions** Ants responding to a dimension of an object (e.g. length) are given a range around the exact value of their "home"

object (the one they are placed on at the start and that defines their feature value). They respond to any value within this range, which is based on the step size for the dimension.

- **Ant complement**
  The ant complement, ac, determines how many ants are placed on each feature of an object at the start. It defines the population size and influences sensitivity of cluster analysis by increasing encounters between ants. Greater computational load is an inevitable consequence and the balance will depend on the density of objects and dimensionality of space.
- **Merging thresholds**
  The colony threshold, ct, determines when the population density of ants is high enough to trigger the ant joining a colony. The feature threshold, ft, is linked to the number of times a particular feature has been seen in other ants. Both are driven by ant encounters. On each encounter, the ant records the following information of the other ant: the ant identifier (id), the colony id, the carried feature id, the timestep, and a boolean flag holding the deposit mode of the encountered ant at that time stamp. This is put into the*AntSeenRecord*, within the *AntSeenList*. The size of the list structure is kept in check by the time stamp which is placed on it. On exceeding the time-window parameter, this encounter is removed.
- **Time-window**
  The time window, tw, defines the maximum number of steps that can be remembered for ant encounters. It helps prevent over-fitting and enables the ACO model to learn new patterns over time if the domain structure changes.
- **Visibility** The number of steps within which an ant encounter is counted. Any ant within this distance of the ant whose encounters are being calculated (the focus ant in the earlier description) becomes eligible for being recorded as an encounter.

B. Ant movement

Ants move one step at a time and each movement is recorded as one timestep for the whole system. The path or edge to follow is chosen as a probabilistic function of the strength of matching pheromone on the first step of each edge leading from the vertex: the higher the strength, the more likely the path will be chosen, which distinguishes it from [7]. This mechanism does not require any foresight about the potential vertices that can be visited, and has the single restriction that ants cannot go back along an edge they have just traversed.

C. The MPACA Algorithm

Require: Graph space with connecting edges and ants assigned to each feature.
while ($Termination\ not\ met$) **do**
  for ($Each\ ant\ in\ antlist$) **do**
    Increment $StepNumber$ against all encounters in AntSeenList by one
    if ($StepNumber > threshold$) **then**
      Remove encounter from AntSeenList
    end if
    if ($Ant\ at\ vertex$) **then**
      Update $AntSeenList$ counts;
      if ($Ant\ features\ match\ object$) **then**
        Activate pheromone deposition mode;
        Process $AntSeenList$ for colony and feature merging
      else
        Deactivate pheromone deposition mode;
      end if
      Choose next edge stochastically taking pheromone values into account;
    end if
    $EdgeTraversal \leftarrow EdgeTraversal - 1$;
    if ($Ant\ in\ deposition\ mode$) **then**
      deposit pheromone for each feature;
    end if
  end for
  if ($Stopping\ criterion\ reached$) **then**
    Output cluster definitions;
  else
    Perform system wide evaporation;
  end if
end while

In the MPACA, each step of the ants is a single time interval so edges which are n steps long will take n timesteps to traverse. The MPACA terminates when ants reach a stable dynamic equilibrium in the colonies they form. This is indicated by a consistent number of colonies and a stable population number in each one.

IV. Evaluation and results

The main aim of this paper is to determine the potential of the MPACA for analysing diverse real-world data sets. Two example domains have been chosen, mental-health risk assessment and hub-and-spoke logistics. The domains have extremely high dimensions (over 200 for the mental-health one) and extremely high numbers of cases (many millions for the logistics domain). These present serious challenges for the tractability of the MPACA but the rewards are high. If the MPACA can form accurate clusters, these will have ant populations that represent a detailed analysis of the relative importance of features and feature combinations required for cluster membership.

In each domain, one of the authors is creating a cognitive model of decision making based on human expertise [42], [43] [44]. The aim is to use it within an Intelligent Knowledge-Based System that helps end users optimise their decisions based on the input information and by exploiting mathematical analysis of the underlying database. The MPACA can provide a useful alternative method that analyses the ant population demographics in each colony to form rules about class membership that can complement the cognitive model. The Ant-Miner algorithm [27] and its derivatives have shown how this approach can work and provide data representations that are

more comprehensible to users. The main loop of the Ant-Miner algorithm consists of three key steps: rule construction, rule pruning, and pheromone updating. Results show that Ant-Miner has good classification performance on test data sets and the ability to constrain the number of rules required [27], [28]. The MPACA rules would be constructed from a detailed understanding of how ant features and their combinations differ within the learned classes.

A. Application of the MPACA to hub-and-spoke logistics networks

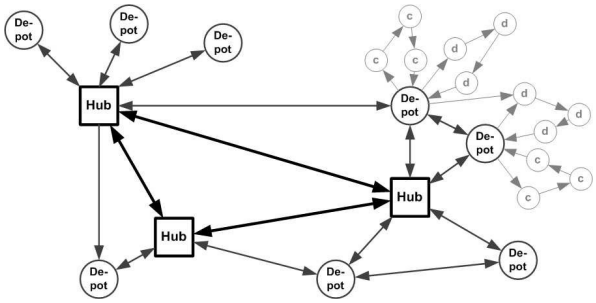


Fig. 1. Transportation in a multiple hub-and-spoke logistics system.

Hub-and-spoke logistics networks have a standard modus operandus [45]. They consist of a number of haulage depots which collect and deliver shipments to and from one or more central hubs. Figure 1 shows a simplified diagram of these activities for a network with 3 hubs and 8 depots. In reality, the networks are much larger than this, with over 100 depots feeding the main hub for the one used to evaluate the MPACA in this paper. The idea is that a depot takes its own customers' shipments to the hub and brings back shipments from any of the other depots that require delivery to the depot's assigned delivery area.

The problem depots have is predicting how many shipments will be at the hub by the end of the day that they are required to deliver. If they take too many lorries to the hub, they will have wasted space on the return trip; if they take too few, they will have to leave shipments behind with costly penalties if the network has to deploy alternative resources to deliver them. In short, if depots could be informed early in the day about the total demand (number of shipments) they will have in the day, this will help decision making to optimise their resources.

Clearly some form of automated analysis is required to enable decision makers in a hub-and-spoke model make sense of the available information [46] and companies have been investing in information technology to this effect [47]. It is a key subject of the EU FP7 co-funded project ADVANCE [48], where various machine learning approaches are being studied with regard to their appropriateness for providing reliable predictions. The MPACA will be applied to the same data to compare the performance of ACO with more traditional machine learning.

Field work conducted for ADVANCE shows that fluctuations in the numbers of shipments (pallets, in this domain) have

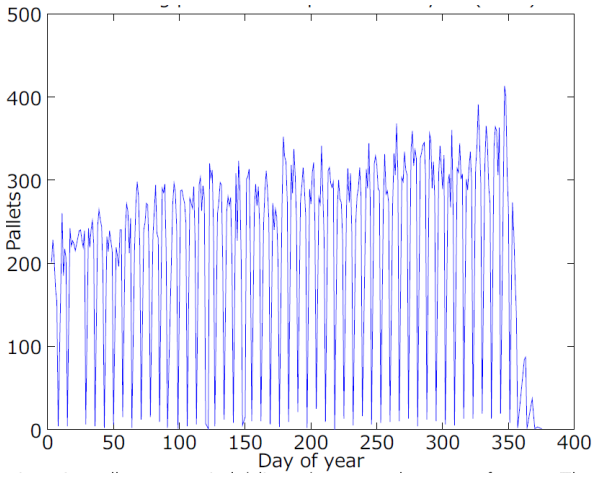


Fig. 2. Fluctuations in the number of pallets each day for a specific depot in the ADVANCE project (the regular very low troughs represent the weekends).

a deleterious impact on operational performance (Figure 2). Such peaks and troughs may appear over the whole network, where the total number of shipments passing through a hub varies widely, as well as on a local level where individual depots experience large changes in numbers from day to day even though the overall network numbers may remain stable. Interviews with depot managers revealed a desire for knowing whether they would have more than or less than the expected number of shipments on a particular day. They could then gear up for additional resources or offer to take on other depots' shipments if spare capacity was likely. To explore the potential of the MPACA in supporting hub-and-spoke decision making, the first step was to find out how well it could predict whether the demand was above or below the mean and compare this with the machine learning program chosen for ADVANCE [40].

1) Predicting shipments: The machine learning program used to compare with the MPACA consists of two main processes: select the most appropriate attributes for a depot and then learn the accompanying linear regression model for predicting the number of shipments or total demand at the end of the day [40]. The attributes used to predict demand include the known current demand (what has already been committed to the hub) and a number of other variables to do with stages of shipment orders, when they were made, and so on. These numbers obviously change as the day progresses so models were learned for separate time points. In fact, a separate regression model was learned for each depot at selected times of each day for each day of the week.

The attribute-selection process picked out 15 of the most influential variables from sixty potential ones and these were used to learn the regression model. The same ones, including the known end-of-day demand, were used by the MPACA to set up the hyperdimensional graph space. Each object (or day in this domain) was assigned to one of two classes: "above" if the known demand was above the mean and "below"

TABLE I
PARAMETER SETTINGS FOR THE MPACA. THE START VALUE IS THE ONE SET AT THE BEGINNING OF LEARNING AND THE MEAN AND STANDARD DEVIATION (SD) ARE THE AVERAGE VALUES AS THESE PARAMETERS WERE MANUALLY VARIED OVER THE 13 TRAINING CYCLES

Parameter	start	mean	SD
Max Edge Length	7	8.4	1.6
Step size	0.1	0.1	0
Phereomone evaporation	0.01	0.05	0
Pheromone deposition	100	100	0
Detection range	2	2	0
Ant complement	1	1.3	0.5
Feature merging threshold	5	5	0
Time window	55	63	8
Visibility	4	4	0

otherwise. At the start of learning, the ants were assigned to the colony matching the class of their starting object. The ants then moved around the graph according to the algorithm described earlier until they had formed population clusters,

Testing was conducted by putting the unknown objects into the hypergraph but with the known-demand dimension removed. In other words, the outcome information about these unknown objects was not included in the domain. They were assigned to the colony that had the nearest centroid (multidimensional mean), measured as the Euclidean distance from the object to that point. This provided the MPACA with the ability to predict whether the demand was going to be greater than or less than normal for the day depending on whether it was in the colony for total known demand above the mean or below the mean. The method differs from the MPACA's origins in cluster analysis [41] by exploiting known outcomes through supervised learning: the actual number of shipments required for delivery is made part of the hyperdimensional space for learning and then removed when classifying unknown cases.

2) Results: Four depots were tested at two different times of the day, 12.00 and 15.00, on a Wednesday. The mean number of shipments for the depots was around 100 (which equates to between two and three lorry loads). Thirteen separate training and testing cycles were conducted for the MPACA and the results were compared with the machine-learning regression model using precision, which is the percentage of outcomes and predictions agreeing with each other with respect to the total sample size of predictions. The sample for each depot consisted of 206 days and these were randomly divided into two equal sets for training and testing.

Table I shows the parameter settings at the beginning and end of learning, where the parameters are in the same order as described in Section III-A. Automated search was not conducted over the parameter space so these are manual settings based on estimates of the optimal initial settings. The mean and SD show that little variation was used to improve the results but this is mainly due to each cycle being set manually. It is likely that a hill-climbing parameter search would produce better results but it is computationally extremely time-consuming and requires optimising the MPACA experimental code.

TABLE II
RESULTS FOR PREDICTING WHETHER DEMAND WILL BE ABOVE OR BELOW THE AVERAGE FOR THAT DAY FOR FOUR DEPOTS AT TWO TIMES. ML GIVES THE MACHINE-LEARNING REGRESSION MODEL PREDICTION AND THE MPACA PRECISION IS ITS MEAN FOR 13 LEARNING AND TESTING CYCLES. THE FINAL STANDARD DEVIATION (SD) COLUMN GIVES THE SD OF THE MEAN ACROSS THE CYCLES.

		Precision		
Depot	**Time**	**ML**	**MPACA**	**SD**
2	12:00	79	68	0.01
3	12:00	83	68	0.01
5	12:00	60	72	0.01
7	12:00	77	74	0.01
2	15:00	74	74	0.02
3	15:00	75	79	0.01
5	15:00	52	80	0.02
7	15:00	65	78	0.00
MPACA mean		71	74	0.01

Table II compares the prediction precision of the MPACA with the machine-learning regresson program produced by ADVANCE [40]. These are preliminary results that are designed to provide an indication of the MPACA's potential for application to real-world data and clearly there are many more sophisticated ways of testing it. Nevertheless, the outcome is promising, with the MPACA having a mean precision equal to the standard regression method. The variation for which of the two models is better for a particular depot and time is probably due to using categorical outcomes where outcome demands only marginally above or below the mean are equally weighted with those having much larger deviations.

B. Mental health risk assessment

The mental-health risk-assessment data relates to the development of the Galatean Risk and Safety Tool, GRiST [49]. GRiST helps mental-health practitioners assess patients' risks of suicide, self-harm, harm to others, self-neglect, and vulnerability. It is based on the assessment knowledge of multidisciplinary practitioners working in all areas of mental health and was designed to disseminate their expertise to services where people did not have a specialist mental-health training.

The MPACA will be tested on the suicide risk data collected by GRiST. The input patient information consists of 138 individual items of information or patient cues. Each of these patient vectors has a clinical risk evaluation given to it by the assessor and the database contains more than 50,000 patient records. However, the data varies in its completeness because the circumstances of assessment often mean only some areas are of interest at any particular time. Therefore, clinical judgements are not based on full vectors, and may have less than 50 per cent of the values present. The output risk judgements are along a sliding scale from 0 (minimum risk) to 10 (maximum risk), which means there are no output classes for categorical assignment. Instead, the judgements map to fuzzy risk labels such as minimum, low, medium, high, and maximum.

The aim of analysing the GRiST suicide data is to determine whether input data can predict clinical judgements accurately.

TABLE III
PARAMETER SETTINGS FOR THE MPACA AS APPLIED TO THE GRIST DATASET. THE START VALUE IS THE ONE SET AT THE BEGINNING OF LEARNING AND THE MEAN AND STANDARD DEVIATION (SD) ARE THE AVERAGE VALUES OVER THE 49 TRAINING CYCLES

Parameter	start	mean	SD
Max Edge Length	9	9.77	0.79
Step size	0.1	0.1	0
Phereomone evaporation	0.005	0.018	0.014
Pheromone deposition	100	148.9	50.51
Detection range	1	1.39	0.49
Ant complement	1	1.45	0.51
Feature merging threshold	5	5	0
Time window	50	53	5.18
Visibility	4	4	0

If so, then the decision support system can provide advice to assessors based on the clinical consensus of the several thousand expert mental-health practitioners who provided the judgements in its database.

The most important pragmatic objective for GRiST predictions is to minimise the numbers of patients who are placed in either the high-risk category when they are low risk or placed in the low-risk category when they are high risk. To test the ability of the MPACA for doing this, two classes of patients were extracted: those with clinical judgements below 4 and those with judgements above 6 on an integer scale from 0 to 10.

Random-forest classification [50] was one of the most successful methods applied to the GRiST data. Its precision for predicting a judgement within plus or minus one of the clinician's judgement was 87%. This was based on 25 of the most important variables and where missing variables had imputed values. For testing the potential of the MPACA, the task was made considerably easier by predicting the most important errors: patients stated to be high risk when they were low or vice versa. However, it was based on a smaller sample using only 13 independent variables and there was no necessity to handle missing data.

The same learning and testing approach was used for the MPACA on the risk data as for the logistics data. A sample of 232 cases were used that were randomly split into 50% training and 50% test cases. The training objects were placed in the hyperdimensional space of 13 variables where the training cases also had the known clinical judgement given as an extra dimension. Ants were placed on each object and if the object was in one of the categories to learn, because the clinical-judgement value was below 4 or above 6, then the ants were assigned to the colony associated with that object. After completion of learning, the test cases were added to the hyperdimensional graph but with the clinical judgement dimension removed. Objects were assigned to the class that had the nearest centroid, as for the logistics domain.

Table III displays the initial parameter values for the 49 cycles of training and testing. Once again, the manual manipulation of parameters from the start value to improve classification did not alter them very much, demonstrated by the very low standard deviation across the 49 cycles. Improvements are obviously possible if automated optimisation was used but these preliminary results show the potential for the MPACA to learn risk judgements. The mean precision, where the MPACA predictions correctly placed test objects into the low clinical risk or high clinical risk categories, was 91.2% with a standard deviation of 0.01. Although this looks like a very good result, it was made easier by only trying to detect gross errors where high and low risks are confused. Attempting to predict the exact judgement between 0 and 10 would obviously be harder but enough encouragement has been given with these initial results to make it worth pursuing.

V. CONCLUSION

This paper has described a new Ant Colony Clustering model called the Multi-Pheromone Ant Clustering Algorithm, MPACA. It was introduced in [41] as a clustering method and was tested on three data sets from the Machine Learning Repository [3]: the Iris, Wine, and Wisconsin Breast Cancer data. This paper gave an overview of the latest incarnation of the MPACA including a detailed description of the algorithm and its parameters. It is unique by having a pheromone for every attribute value of the objects in the domain space. The ants are able to link similar features of objects, to combine the features they detect depending on the frequency with which they meet other ants with the same features, and to form colonies based on local ant population densities. Together, these enable ants to learn the feature profile for different clusters and for mapping colony membership onto those clusters. Where this paper differs from the earlier one is by extending the model to classification learning as well as cluster analysis. In other words, it shows how the MPACA can be adapted for supervised learning and that it should perhaps be renamed a classification/clustering algorithm. Secondly, the paper explored how useful and effective the approach might be when applied to noisy and heterogeneous real-world data sets. These create interesting problems and this paper conducts experiments that determine whether innovations of the MPACA translate into useful outcomes.

Two data sets were used, one for logistics and one for mental health. The structure, dimensionality, and classification objectives differed widely between the two sets but the results show that the MPACA can induce and utilise patterns to produce helpful classification advice. A more stringent test was given to the algorithm for the logistics domain than the mental-health one and the application to both domains could be improved. For the logistics data, having classification decisions based on such broad categories as either above or below the mean does not provide the most interesting output to end users. They need to know how large is the deviation from the mean. In fact, the most important information is whether there will be a peak or trough in demand and the MPACA could easily be adapted to test for these by redefining classes into those where the demand exceeds a given threshold value above or below the mean. This is rather like its application to the mental-health risk data where high and low risk patients were being discriminated. Of course, this leaves patients with

judgements in between these classes without a colony and it would be useful to predict their category as well.

The machine learning regression approach in each domain predicts the actual values of outputs, not just class membership, which makes it more informative. Further work on the MPACA will be on how to translate the colony memberships into a similar prediction. Even with the crude assignment mechanism used in this paper, where unknown objects were classified in the class associated with the colony having the nearest multidimensional mean (centroid), it is possible to translate the relative distances from colonies into the degree of membership in the colony. The more membership in a class above or below the mean, the higher the difference between the predicted demand and the mean.

The most productive way of immediately improving the classification output of the MPACA is by using more sophisticated assignments of unknown objects to classes after learning. Methods currently under investigation include variants of nearest-neighbour analysis where the number of ants from different colonies is calculated for all nodes within a given radius of the object to be classified. The relative proportions of colony populations can be translated into a probability of class membership using a simple Bayes equation. Alternatively, sophisticated probability density functions could be used as input to the Bayesian probability calculations.

There are many avenues requiring exploration for the MPACA model itself, both with the general mechanism and its parameterisation. At the time of writing, there are problems with merging colonies because domains with multiple clusters eventually merge into just two. Somewhere in the learning and merging process, an optimal configuration will have been achieved but it is not easy to know when; some form of dynamic equilibrium should happen and it should also be detectable so that it is clear when learning has reached an optimum end point.

Parameters are an important influence on the model's operation and more needs to be discovered about how they exert their influence so that performance can be improved. The current method is slow and cumbersome, requiring manual setting of parameters, observation of performance, and a new run with adjusted parameters in accordance with conclusions from the observations. A hill-climbing approach where parameters are systematically adjusted to reduce classification errors after learning is the obvious next step. The problem is that ACO methods are computationally expensive and time consuming, requiring careful optimisation of the MPACA code to generate the necessary execution speed.

An important guideline to remember for future research on the MPACA is to avoid chasing performance optimisation without understanding how it is being achieved. Otherwise the particular qualities of the MPACA could be lost or diluted, with improvements failing to come from the metaphor that has motivated the research in the first place.

ACKNOWLEDGEMENT

This research was partly supported by the European Commission through the 7th FP project ADVANCE (http://www.advance-logistics.eu) under grant No. 257398.

REFERENCES

[1] C. Blum and D. Merkle, eds, "Swarm intelligence: Introduction and applications". Springer, 2008.

[2] J.R.J. French and B. M. Ahmed, "The challenge of biomimetic design for carbon-neutral buildings using termite engineering," *Insect Science*, vol. 17, no. 2, pp. 154-162, Feb 2010.

[3] K. Bache and M. Lichman. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

[4] S. Gueron, S. A. Levin, and D. I. Rubenstein, "The dynamics of herds: From individuals to aggregations," *Journal of Theoretical Biology*, vol. 182, no. 1, pp. 85-98, Sep. 1996.

[5] J. K. Parish and W.M. Hamner, eds. "Animal Groups in Three Dimensions, How Species Aggregate". Cambridge University Press, 1997.

[6] J. D. Murray, *Mathematical Biology: I. An Introduction (Interdisciplinary AppliedMathematics) (Pt. 1)*, 3rd ed., New York, Springer, Jan. 2007.

[7] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28-39, Nov. 2006

[8] J. L. Deneubourg, S. Goss, N. Franks, A. S. Franks, C. Detrain, and L. Chrétie, "The dynamics of collective sorting robot-like ants and ant-like robots," in *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats*. Cambridge, MA, USA: MIT Press, 1990, pp. 356-363.

[9] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. Thesis, Politecnico di Milano, Italy, 1992

[10] A. Dussutour, S.C. Nicolis, G. Shephard, M. Beekman, and D.J.T. Sumpter, "The role of multiple pheromones in food recruitment by ants," *The Journal of Experimental Biology*, vol. 212, no. 15, pp. 2337-2348, Aug. 2009.

[11] W. Ngenkaew, S. Ono and S. Nakayama, "Pheromone-based concept in Ant Clustering," in *Proceedings 3rd International Conference on Intelligent System and Knowledge Engineering, ISKE 2008*, vol. 1, 17-19 Nov. 2008, pp. 308-312.

[12] M. Middendorf, F. Reischle, and H. Schmeck, "Multi colony ant algorithms," in emphJournal of Heuristics, vol. 8, no. 3, pp. 305-320, May 2002.

[13] M. Guntsch "Ant Algorithms in Stochastic and Multi-Criteria Environments," Ph.D. Thesis, Institut AIFB, University of Karlsruhe, Germany, 2004

[14] O. A. Mohamed Jafar and R. Sivakumar, "Ant-based Clustering Algorithms: A Brief Survey," *International Journal of Computer Theory and Engineering*, vol. 2, no. 5, pp. 1793-8201, October 2010.

[15] N. Labroche, N. Monmarché and G. Venturini, "A New Clustering Algorithm Based on the Chemical Recognition System of Ants," in *Procceedings of 15th European Conference on Artificial Intelligence (ECAI2002)*, Lyon France, 2002, pp. 345-349.

[16] N. Labroche, F.J. Richard, N. Monmarché, A. Lenoir and G. Venturini, "Modelling of the Chemical Recognition System of Ants." [Online]. Available: http://hant.li.univ-tours.fr/webhant/pub/LabRicMonLenVen02a.iwsoesb.pdf

[17] D. Zaharie and F. Zamfirache "Dealing with noise in ant-based clustering," in *Proceedings of IEEE Congress on Evolutionary Computation*, Edinburgh, UK, 2-5 Sept. 2005, pp. 2395-2401.

[18] X-C. Liang, S-F. Chen and Y. Liu, "The study of small enterprises credit evaluation based on incremental AntClust," in *Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services, GSIS 2007*, Nanjing, China, 18-20 Nov. 2007, pp. 294-298.

[19] H. H. Inbarani and K. Thangavel, "Clickstream Intelligent Clustering using Accelerated Ant Colony Algorithm," in *Proceedings of the International Conference on Advanced Computing and Communications, ADCOM 2006* Surathkal, India, 20-23 Dec. 2006 pp. 129-134.

[20] C. Bertelle, A. Dutot, F. Guinand and D. Olivier, "Organization Detection Using Emergent Computing," in *International Transactions on Systems Science and Applications (ITSSA)*, vol. 2, no. 1, pp. 61-69, 2006.

[21] V. Ramos, F. Muge and P. Pina. "Self-Organized Data and Image Retrieval as a Consequence of Inter-Dynamic Synergistic Relationships in Artificial Ant Colonies," in *Proceedings of Soft Computing Systems: Design, Management and Applications, 2nd Int. Conf. on Hybrid Intelligent Systems, AEB02,* IOS Press, Frontiers of Artificial Intelligence and Applications, vol. 87, pp. 500-509, Amsterdam, Dec 2002.

[22] I. El-Feghi, M. Errateeb, M. Ahmadi and M. A. Sid-Ahmed, "An adaptive ant-based clustering algorithm with improved environment perception," in *Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, SMC 2009, San Antonio, Texas, USA, 11-14 Oct. 2009, pp. 1431-1438.

[23] M. Kothari, S. Ghosh and A. Ghosh, "Aggregation Pheromone Density Based Clustering," in *Proceedings of the 9th International Conference on Information Technology (ICIT '06)*, IEEE Computer Society, Washington, DC, USA, 18-21 Dec. 2006, pp. 259-264.

[24] P. S. Shelokar, V.K. Jayaraman, and B.D. Kulkarni, "An ant colony approach for clustering," in *Analytica Chimica Acta* vol. 509, no. 2, pp. 187-195, 2004

[25] H. Jiang and S. Chen, "A new ant colony algorithm for a general clustering," in *Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services 2007, GSIS 2007*, Nanjing, China, 18-20 Nov. 2007, pp. 1158-1162.

[26] H. Yang, X. Li, C. Bo and X. Shao, "A Graphic Clustering Algorithm Based on MMAS," in *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2006* Vancouver, BC, Canada, 11 September 2006 , pp. 1592-1597.

[27] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," in *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 321-332, Aug 2002.

[28] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "An Ant Colony Based System for Data Mining: Applications To Medical Data," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H. M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, Eds. San Francisco, California, USA: Morgan Kaufmann, Jul-Nov 2001, pp. 791-797

[29] D. Martens, M. De Backer, R. Haesen, J. Vanthienen, M. Snoeck, and B. Baesens, "Classification with ant colony optimization," in *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 5, pp. 651-665, 2007.

[30] D. Elizondo, "The linear separability problem: some testing methods," in *IEEE Transactions on neural networks*, vol. 17, no. 2, pp. 330-344, 2006.

[31] J. Handl, J. Knowles, and M. Dorigo, "On the performance of ant-based clustering," in *Proceedings of the Third International Conference on Hybrid Intelligent Systems*, IOS Press, Frontiers in Artificial Intelligence and Appliations, vol. 104, 2003, pp. 204-213.

[32] Y. Sasaki (2007, October) "The truth of the F-measure," [Online]. Available: http://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/index-e.html

[33] X. F. Huang, Y. Yang and X. Niu "Towards Improving Ant-Based Clustering - An Chaotic Ant Clustering Algorithm," in *Procedings of the International Conference on Computational Intelligence and Security Workshops*, CISW '07, Harbin, China, 15-19 Dec. 2007, pp. 421-424.

[34] Y. Yang and M. Kamel, "Clustering ensemble using swarm intelligence," in *Proceedings of the 2003 IEEE Swarm Intelligence Symposium*, SIS '03, 24-26 April 2003, pp. 65-71.

[35] Q-M. Rong, W-C. Wu and L. Li, "Research on Hybrid Clustering Based on Density and Ant Colony Algorithm," *Proceedings of the Second International Workshop on Education Technology and Computer Science*, ETCS '10, Wuhan, China, 6-7 March 2010, vol. 2, pp. 222-225. doi: 10.1109/ETCS.2010.42

[36] R. Chandrasekar, V. Vijaykumar and T. Srinivasan, "Probabilistic ant based clustering for distributed databases," in *Proceedings of the 3rd International IEEE Conference Intelligent Systems*, 2006, London, UK, Sept 2006, pp. 538-545.

[37] C.D. Buckingham, "Psychological cue use and implications for a clinical decision support system," *Medical Informatics and the Internet in Medicine*, vol. 27, no. 4, pp. 237-251, 2002.

[38] C. D. Buckingham, A. Ahmed, and A.E. Adams, "Using XML and XSLT for flexible elicitation of mental-health risk knowledge," *Medical Informatics and the Internet in Medicine*, vol. 32, no. 1, pp. 65-81, 2007.

[39] S. E. Hegazy and C.D. Buckingham, "A Method for Automatically Eliciting node Weights in a Hierarchical Knowledge Based Structure for Reasoning with Uncertainty," *International Journal On Advances in Software*, vol. 2, no. 1, pp. 76-85, 2009.

[40] P. G. Welch and Z. Kemeny and A. Ekart and E. Ilie-Zudor, "Application of model-based prediction to support operational decisions in logistics networks". In Proc. AILog, ECAI, 2012.

[41] J. Chircop and C. D Buckingham. A Multiple Pheromone Ant Clustering Algorithm. Proceedings of NICSO 2013, to be published in Studies in Computational Intelligence, Springer, 2013.

[42] C. D. Buckingham and P. Buijs and P. G. Welch and A. Kumar and A. Ahmed, "Developing a cognitive model of decision-making to support members of hub-and-spoke logistics networks". In Proceedings of the 14th International Conference on Modern Information Technology in the Innovation Processes of the Industrial Enterprises, Eds. Elisabeth Ilie-Zudor, Zsolt Kemény, László Monostori, pp 14-30, 2012. [Online]. Available: igor.xen.emi.sztaki.hu/mitip/media/MITIP2012 proceedings.pdf

[43] C. D. Buckingham, A. E. Adams, and C. Mace, "Cues and knowledge structures used by mental-health professionals when making risk assessments". *Journal of Mental Health*, vol. 17, no. 3, pp. 299–314, 2008.

[44] C. D. Buckingham and A. Adams, "The grist web-based decision support system for mental-health risk assessment and management". In *Proceedings of the First BCS Health in Wales/ehi2 joint Workshop*, pp. 37–40, 2011.

[45] G. Zapfel and M. Wasner, "Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers". *International Journal of Production Economics*, 78(2), pp. 207–220, 2002. DOI: 10.1016/S0925-5273(00)00152-3.

[46] B. MacCarthy and J. Wilson, eds, "Human Performance in Planning and Scheduling". Taylor & Francis, URL: http://books.google.com/books?id=0wBLHGX1_WIC, 2001.

[47] J. Schumacher and K. Feurstein, "Proceedings of the 3rd European conference on ICT for Transport Logistics". URL: http://www.ecitl.eu/. 2010.

[48] ADVANCE (last accessed, July 2013), ADVANCE: Advanced predictive-analysis-based decision-support engine for logistics. URL: http://www.advance-logistics.eu/.

[49] GRiST (last accessed, July 2013), Galatean Risk and Safety Tool. URL: http://www.egrist.org/.

[50] L. Breiman, "Random forests". *Machine Learning*, 45(1), pp 5–32, 2001.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 35–41

Fuzziness in Partial Approximation Framework

Zoltán Ernő Csajbók
Department of Health Informatics
Faculty of Health, University of Debrecen
Nyíregyháza, Hungary
Email: csajbok.zoltan@foh.unideb.hu

Tamás Mihálydeák
Department of Computer Science
Faculty of Informatics, University of Debrecen
Debrecen, Hungary
Email: mihalydeak.tamas@inf.unideb.hu

***Abstract*—In partial approximation spaces with Pawlakian approximation pairs, three partial membership functions are generated. These fuzzy functions rely on the lower and upper approximations of a set. They provide special type of fuzziness on the universe: all of them are partial functions and derived from the observed data relatively to available knowledge about the objects of the universe. With the help of these functions, three new approximation pairs are generated and so new approximation spaces appear effectively. Using not Pawlakian approximation pairs gives a special insight into the nature of general set approximations, and so new models of necessity and possibility can be given.**

I. Introduction

SET approximations were invented by Pawlak in the early 1980's which is known as rough set theory [1], [2], [3]. Its general scheme may be outlined as follows. Let a beforehand predefined family of subsets of the universe of objects be given. It is called the base system from which definable sets may be derived. Next, so-called lower and upper approximations can be formed with the help of definable sets via beforehand fixed rules in order to approximate any sets in the universe.

The starting point of rough set theory is a nonempty *finite* set U of objects and an equivalence relation ε on U [3]. The equivalence classes are called ε-elementary sets.

Definable sets are any unions of ε-elementary sets. Any set $S \subseteq U$ can be naturally approximated by the lower and upper ε-approximations of S which are denoted by $\underline{\varepsilon}$ and $\overline{\varepsilon}$, respectively. The former is the union of all ε-elementary sets which are the subsets of S, whereas the latter is the union of all ε-elementary sets which have a nonempty intersection with S.

A number of studies deal with the relationship between rough set theory and fuzzy set theory [4], [5], [6], [7], [8]. A detailed discussion of their connections and differences can be found, e.g., in [9], [10], [11].

There are many possibilities to establish a relationship between them [12], [13], [14], [15].

Just until now it has been generally accepted that the two theories are related but distinct and complementary to each other. Recently, however, Chakraborty has proposed a common ground relying on the *classical rough membership function* [16].

The classical rough membership function quantifies the degree of the relative overlap between a set $S \subseteq U$ and an ε-elementary set [10].[1] As usual, it is defined by

$$\mu_S^{\varepsilon}(u) = \frac{|[u]_{\varepsilon} \cap S|}{|[u]_{\varepsilon}|},$$

where $|\cdot|$ is the cardinality of a set, and $[u]_{\varepsilon}$ denotes the ε-elementary set to which a $u \in U$ belongs.

Hence, we just obtain a fuzzy membership function $\mu_S^{\varepsilon} : U \to [0,1]$ with

$$\begin{aligned} \mu_S^{\varepsilon}(u) &= 1 \text{ if and only if } [u]_{\varepsilon} \subseteq S; \\ \mu_S^{\varepsilon}(u) &> 0 \text{ if and only if } [u]_{\varepsilon} \cap S \neq \emptyset; \\ \mu_S^{\varepsilon}(u) &= 0 \text{ if and only if } [u]_{\varepsilon} \cap S = \emptyset. \end{aligned}$$

Thus, the rough membership function can be seen as a fuzzyfication of rough approximation, and μ_S^{ε} is a fuzzy subset of U induced by S.

One of the main features of μ_S^{ε} is that it relies on the system of base sets, the system of equivalence classes. In other words, *rough membership functions are generated by our knowledge* (appearing, e.g., in an information system). This is a distinctive feature of rough membership functions in contrast with fuzzy membership functions [18]. Furthermore, following from the definition of μ_S^{ε}, there are many constraints on the values of rough membership functions [12], [20], [21].

An important observation is that the Pawlakian lower and upper approximation pair can be reconstructed by employing the rough membership function. The well-known formulae are the following:

$$\begin{aligned} \underline{\varepsilon}(S) &= \{u \in U \mid \mu_S^{\varepsilon}(u) = 1\}, \\ \overline{\varepsilon}(S) &= \{u \in U \mid \mu_S^{\varepsilon}(u) > 0\}. \end{aligned}$$

In the terminology of fuzzy set theory, the lower and upper approximations $\underline{\varepsilon}$ and $\overline{\varepsilon}$ are the *core* and the *support* of the fuzzy set μ_S^{ε}, respectively.

Nevertheless, Pawlakian set approximation has some very strong theoretical requirements:

- the system of base sets are total, i.e., their union gives back the universe;
- base sets are pairwise disjoint.

[1] Note that the notion of a classical rough membership function was explicitly introduced by Pawlak and Skowron in [10]. Nevertheless, it had been used and studied earlier by many authors. For more historical remarks, see [17]. Moreover, such a coefficient has already been considered by Łukasiewicz in 1913 [18], [19].

978-83-60810-53-8/$25.00

In many cases, however, our knowledge does not fulfill these requirements:

- The partition shows the limit of our knowledge about the objects of the universe in the sense that two objects are indistinguishable if they belong to the same base set. On the other hand, it makes explicit our knowledge because we do distinguish two objects belonging to different base sets. Giving up the requirement of the pairwise disjoint property, the so–called *covering–based rough set theory* is obtained [22], [23], [24], [25], [26], [27].
- The universe may involve some objects without any information, i.e., base sets are not total. For instance, information systems often contain $NULL$ values. In the papers [28], [29], the authors give a very general system of the set approximation giving up both the pairwise disjoint property and the covering of the universe. It is called the (general) *partial approximation framework*.

In this paper, the above procedure is transferred to a partial set approximation context:

1) First, in a partial approximation space with a Pawlakian approximation pair, three *partial* membership functions are defined in the style of the classical rough membership function.
2) Then, three approximation pairs are generated with the help of partial membership functions. The question is whether these approximation pairs meet (at least) the minimum requirements of approximation pairs, i.e., these pairs actually form approximation pairs in partial approximation spaces.

The rest of the paper consists of three parts. In Section 2, the basic notions and notations of partial approximation spaces are summarized. In Section 3, three approximation pairs are generated as outlined above, and it is shown that they meet the minimum requirements prescribed for approximation pairs in partial approximation spaces. Section 4 consist of some remarks on the logical application of partial membership functions.

II. Partial approximation of sets

A. Basic notions and notations

Let U be a nonempty finite set and $\mathfrak{B} \subseteq 2^U$ be a nonempty family of nonempty subsets of U. U is the *universe of objects*, $\mathfrak{B}$ is the *base system* and its members are $\mathfrak{B}$-*sets* or *base sets* [30], [29], [31], [32], [33].

If $B \in \mathfrak{B}$ is a union of a family of sets $\mathfrak{B}' \subseteq \mathfrak{B} \setminus \{B\}$, B is called *reducible* in $\mathfrak{B}$, otherwise B is *irreducible* in $\mathfrak{B}$.

A base system $\mathfrak{B}$ is *single–layered* if every base set is irreducible, and *one–layered* if the base sets are pairwise disjoint. Of course, a one–layered base system is single–layered. From any base systems, single–layered and one–layered base systems can be constructed [31].

By formulae, a base system $\mathfrak{B}$ is single–layered, if

$$\forall B \in \mathfrak{B}\ \ \forall \mathfrak{B}' \subseteq \mathfrak{B} \setminus \{B\}\, (B \cap \bigcup \mathfrak{B}' \neq B),$$

and one–layered, if

$$\forall B \in \mathfrak{B}\ \ \forall \mathfrak{B}' \subseteq \mathfrak{B} \setminus \{B\}\, (B \cap \bigcup \mathfrak{B}' = \emptyset).$$

Informally, a base system $\mathfrak{B}$ is single–layered if every nonempty union of base sets has at least one member which belongs to exactly one base set, whereas $\mathfrak{B}$ is one–layered if all members of every nonempty union of base sets belong to exactly one base set.

During the approximation process, a family of sets $\mathfrak{D}_{\mathfrak{B}} \subseteq 2^U$ are applied. In the most general case, it is supposed only just that

1) $\mathfrak{D}_{\mathfrak{B}}$ is an extension of $\mathfrak{B}$, i.e., $\mathfrak{B} \subseteq \mathfrak{D}_{\mathfrak{B}}$;
2) $\emptyset \in \mathfrak{D}_{\mathfrak{B}}$.

Let $\mathsf{l}, \mathsf{u} : 2^U \to 2^U$ be an ordered pair of mappings and denoted it by $\langle \mathsf{l}, \mathsf{u} \rangle$.

The intended meaning of l and u is to express the lower and upper approximations of any subsets of U. Hence, it is called an *approximation pair*. The next definition specifies its *minimum requirements*.

Definition 1. An approximation pair $\langle \mathsf{l}, \mathsf{u} \rangle$ is a *weak approximation pair* if

(C0) $\mathsf{l}(2^U), \mathsf{u}(2^U) \subseteq \mathfrak{D}_{\mathfrak{B}}$ (*definability* of l and u);
(C1) l and u are monotone, i.e. for all $S_1, S_2 \in 2^U$ if $S_1 \subseteq S_2$ then $\mathsf{l}(S_1) \subseteq \mathsf{l}(S_2)$ and $\mathsf{u}(S_1) \subseteq \mathsf{u}(S_2)$ (*monotonicity* of l and u);
(C2) $\mathsf{u}(\emptyset) = \emptyset$ (*normality* of u);
(C3) if $S \subseteq U$, then $\mathsf{l}(S) \subseteq \mathsf{u}(S)$ (*weak approximation property*).

Clearly, l and u are many-to-one and $\mathsf{u}(2^U) \neq \mathsf{l}(2^U) \subseteq \mathfrak{D}_{\mathfrak{B}}$ in general.

Informally, definable sets represent our available knowledge about the the objects of the universe. They can be though of as *tools*, in more detail, base sets as *primary tools* and definable sets as *derived tools*. An approximation pair prescribes the *utilization* of tools in approximation processes.

It is reasonable that base sets as primary tools are exactly approximated from "lower side". In classical rough set theory, however, definable sets are exactly approximated from "lower side" as well.

Definition 2. A weak approximation pair $\langle \mathsf{l}, \mathsf{u} \rangle$ is

(C4) *granular* if $B \in \mathfrak{B}$, then $\mathsf{l}(B) = B$ (l is *granular*),
(C5) *standard* if $D \in \mathfrak{D}_{\mathfrak{B}}$, then $\mathsf{l}(D) = D$ (l is *standard*).

Of course, if l is standard, the granularity of l also holds.

The following proposition summarizes some simple consequences of the minimum requirements *(C0)–(C3)* in Definition 1 and the conditions *(C4)–(C5)* in Definition 2.

Proposition 1. *Let $\langle \mathsf{l}, \mathsf{u} \rangle$ be a weak approximation pair on U.*

1) $\mathsf{l}(\emptyset) = \emptyset$ *(normality of* l*).*
2) l *is idempotent, i.e.,* $\mathsf{l}(\mathsf{l}(S)) = \mathsf{l}(S)$ *for all* $S \in 2^U$*, and* $\mathsf{l}(2^U) = \mathfrak{D}_{\mathfrak{B}}$ *if and only if* l *is standard.*

3) a) *If* $\mathsf{l}(S) = S$*, then* $S \in \mathfrak{D}_{\mathfrak{B}}$.
   b) *Let* l *be standard. Then,* $\mathsf{l}(S) = S$ *if and only if* $S \in \mathfrak{D}_{\mathfrak{B}}$.
4) a) *If* $\mathsf{l}(U) = \bigcup \mathfrak{D}_{\mathfrak{B}}$*, then* $\bigcup \mathfrak{D}_{\mathfrak{B}} \in \mathfrak{D}_{\mathfrak{B}}$.
   b) *Let* l *be standard. Then,* $\mathsf{l}(U) = \bigcup \mathfrak{D}_{\mathfrak{B}}$ *if and only if* $\bigcup \mathfrak{D}_{\mathfrak{B}} \in \mathfrak{D}_{\mathfrak{B}}$.

The next definition deals with the question how lower and upper approximations relate to the approximated sets.

Definition 3. A weak approximation pair $\langle \mathsf{l}, \mathsf{u} \rangle$ is

(C6) lower semi–strong if $\mathsf{l}(S) \subseteq S$ for all $S \in 2^U$ (i.e., l is contractive);
(C7) upper semi–strong if $S \subseteq \mathsf{u}(S)$ for all $S \in 2^U$ (i.e., u is extensive);
(C8) strong if it is lower and upper semi–strong, i.e., each subset $S \in 2^U$ is bounded by $\mathsf{l}(S)$ and $\mathsf{u}(S)$: $\mathsf{l}(S) \subseteq S \subseteq \mathsf{u}(S)$.

Proposition 2.

1) *If* $\langle \mathsf{l}, \mathsf{u} \rangle$ *is an upper semi–strong approximation pair on* U*, then* $\mathsf{u}(U) = U$ *(co–normality of* u*).*
2) *If* $\langle \mathsf{l}, \mathsf{u} \rangle$ *is an upper semi–strong approximation pair on* U *and* l *is standard, then* $\mathsf{l}(U) = U$ *(co–normality of* l*).*

Based on the foregoing, a general set–theoretic partial approximation framework can be defined as follows.

Definition 4. The ordered 5–tuple $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}, \mathsf{l}, \mathsf{u} \rangle$ whose components are defined as before, is called a *(general) approximation space*.

Definition 5. $\mathsf{GAS}(U)$ is a *(general) total approximation space* or simply *total*, if $\mathfrak{B}$ covers the universe, i.e., $\bigcup \mathfrak{B} = U$, otherwise $\mathsf{GAS}(U)$ is a *(general) partial approximation space* or simply *partial*.

Definition 6. $\mathsf{GAS}(U)$ *relies on Pawlakian base*, if $\mathfrak{B}$ is a partition of U.

Corollary 1. $\mathsf{GAS}(U)$ *relies on Pawlakian base if and only if its base system is total and one–layered.*

Definition 7. The general approximation space $\mathsf{GAS}(U)$ is a *weak/standard/lower semi–strong/upper semi–strong/strong approximation space*, if the approximation pair $\langle \mathsf{l}, \mathsf{u} \rangle$ is weak/ standard/lower semi–strong/upper semi–strong/strong, respectively.

B. Exactness in general approximation spaces

In classical rough set theory, the notions of "crisp" and "definable" are inherently one and the same. In general approximation spaces, however, they can be differentiated.

Definition 8. Let $\mathsf{GAS}(U)$ be a weak approximation space and $S \subseteq U$.
S is *crisp*, if $\mathsf{l}(S) = \mathsf{u}(S)$, otherwise
S is *rough*.

If a set is crisp, its lower and upper approximations coincide with the approximated set only in strong approximation spaces. Furthermore, a crisp set is necessarily definable only in strong approximation spaces as well. However, it can easily be shown that a definable set is not necessarily crisp even in strong approximation spaces ([33], Example 8). Consequently, in general approximations spaces, the notions of "crisp" and "definability" are generally not synonymous to each other.

C. General approximation spaces with Pawlakian approximation pairs

Definition 9. $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}, \mathsf{l}, \mathsf{u} \rangle$ is a *approximation space with a Pawlakian approximation pair*, if

1) U is a finite nonempty set;
2) $\mathfrak{D}_{\mathfrak{B}}$ is *strict finite union type*, i.e., it is given by the following inductive definition:
   a) $\emptyset \in \mathfrak{D}_{\mathfrak{B}}$;
   b) $\mathfrak{B} \subseteq \mathfrak{D}_{\mathfrak{B}}$;
   c) if $B_1, B_2 \in \mathfrak{B}$, then $B_1 \cup B_2 \in \mathfrak{D}_{\mathfrak{B}}$;
3) $\langle \mathsf{l}, \mathsf{u} \rangle$ is a *Pawlakian approximation pair*, i.e.,
   a) $\mathsf{l}(S) = \bigcup \mathbb{L}(S)$, where $\mathbb{L}(S) = \{B \in \mathfrak{B} \mid B \subseteq S\}$;
   b) $\mathsf{u}(S) = \bigcup \mathbb{U}(S)$, where $\mathbb{U}(S) = \{B \in \mathfrak{B} \mid B \cap S \neq \emptyset\}$.

Proposition 3. *Let* $\mathsf{GAS}(U)$ *be an approximation space with a Pawlakian approximation pair.*

1) $\mathsf{GAS}(U)$ *is a standard lower semi–strong approximation space.*
2) $\mathsf{GAS}(U)$ *is an upper semi–strong approximation space if and only if* $\mathfrak{B}$ *covers the universe.*

Definition 10. Let $\mathsf{GAS}(U)$ be an approximation space with a Pawlakian approximation pair and $S \subseteq U$. Then

$$\mathsf{b}(S) = \bigcup (\mathbb{U}(S) \setminus \mathbb{L}(S))$$

is called the *boundary* of S.

Clearly, $\mathsf{b}(S) \subseteq \mathsf{u}(S)$ for all $S \subseteq U$.

Corollary 2. *Let* $\mathsf{GAS}(U)$ *be an approximation space with a Pawlakian approximation pair.*

1) *In general,* $\mathsf{u}(S) \setminus \mathsf{l}(S) \subseteq \mathsf{b}(S)$ *for any* $S \subseteq U$.
2) *If* $S \subseteq U$*,*

$$\mathsf{b}(S) = \mathsf{u}(S) \setminus \mathsf{l}(S) \Leftrightarrow \mathsf{b}(S) \cap \mathsf{l}(S) = \emptyset.$$

Proof:

1) $u \in \mathsf{u}(S) \setminus \mathsf{l}(S)$

$$\begin{aligned} &\Leftrightarrow \quad u \in \bigcup \mathbb{U}(S) \wedge u \notin \bigcup \mathbb{L}(S) \\ &\Leftrightarrow \quad \exists B \in \mathfrak{B}\,(u \in B \wedge B \in \mathbb{U}(S) \wedge B \notin \mathbb{L}(S)) \\ &\Leftrightarrow \quad \exists B \in \mathfrak{B}\,(u \in B \wedge B \in \mathbb{U}(S) \setminus \mathbb{L}(S)) \\ &\Rightarrow \quad u \in \bigcup (\mathbb{U}(S) \setminus \mathbb{L}(S)) = \mathsf{b}(S) \end{aligned}$$

2) ($\Rightarrow$) $\mathsf{b}(S) \cap \mathsf{l}(S) = (\mathsf{u}(S) \setminus \mathsf{l}(S)) \cap \mathsf{l}(S) = \emptyset$.
   ($\Leftarrow$) $\mathsf{b}(S)$

$$\begin{aligned} &= (\mathsf{b}(S) \cap \mathsf{l}(S)) \cup (\mathsf{b}(S) \cap (\mathsf{l}(S))^c) \\ &= \mathsf{b}(S) \cap (\mathsf{l}(S))^c \\ &\subseteq \mathsf{u}(S) \cap (\mathsf{l}(S))^c = \mathsf{u}(S) \setminus \mathsf{l}(S), \end{aligned}$$

which are compared to (1), we get

$$\mathsf{b}(S) = \mathsf{u}(S) \setminus \mathsf{l}(S).$$

■

III. Fuzziness in partial approximation spaces with Pawlakian approximation pairs

Let $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}, \mathsf{l}, \mathsf{u} \rangle$ be a partial approximation space with a Pawlakian approximation pair. In other words, $\mathsf{GAS}(U)$ is an approximation space with a Pawlakian approximation pair and $\bigcup \mathfrak{B} \subseteq U$.

A. Partial membership functions

If $u \in U$, let $\mathcal{N}_{\mathfrak{B}}(u) = \{B \in \mathfrak{B} \mid u \in B\}$. The family of sets $\mathcal{N}_{\mathfrak{B}}(u)$ may be called the *(reflexive) neighborhood system* of u with respect to the base system $\mathfrak{B}$ [34], and its members are called the *neighborhoods* of u.

Three different partial membership functions are defined in $\mathsf{GAS}(U)$ as follows [32], [35], [36], [38], [20].

Definition 11. Let $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}, \mathsf{l}, \mathsf{u} \rangle$ be a partial approximation space with a Pawlakian approximation pair and $S \subseteq U$.

$\mu_S^o, \mu_S^a, \mu_S^p : U \to [0,1]$ are *optimistic/average/pessimistic partial membership functions*, respectively, if

$$\mu_S^o(u) = \begin{cases} \mathsf{max}\left\{ \frac{|B \cap S|}{|B|} \mid B \in \mathcal{N}_{\mathfrak{B}}(u) \right\}, & \text{if } u \in \bigcup \mathfrak{B}; \\ \mathsf{undefined}, & \text{otherwise;} \end{cases}$$

$$\mu_S^a(u) = \begin{cases} \mathsf{avg}\left\{ \frac{|B \cap S|}{|B|} \mid B \in \mathcal{N}_{\mathfrak{B}}(u) \right\}, & \text{if } u \in \bigcup \mathfrak{B}; \\ \mathsf{undefined}, & \text{otherwise;} \end{cases}$$

$$\mu_S^p(u) = \begin{cases} \mathsf{min}\left\{ \frac{|B \cap S|}{|B|} \mid B \in \mathcal{N}_{\mathfrak{B}}(u) \right\}, & \text{if } u \in \bigcup \mathfrak{B}; \\ \mathsf{undefined}, & \text{otherwise.} \end{cases}$$

Remark 1. For the sake of brevity, we will use the symbol "*" in order to denote a member of $\{o, a, p\}$.

In Definition 11, each partial membership function μ_S^* forms a special type of fuzziness on U which is induced by the base system $\mathfrak{B}$, i.e., our available knowledge (primary tools) about the objects of the universe.

An important feature of each μ_S^* is that it is a *partial* function. Clearly, if $\bigcup \mathfrak{B} \subsetneqq U$, $\mu_S^*(u)$ is undefinable for all $u \in U \setminus \bigcup \mathfrak{B}$. In other words, $\mathtt{dom}\ \mu_S^* = \bigcup \mathfrak{B} \subsetneqq U$.[2]

The following statements can easily be checked.

Proposition 4. *Let $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}, \mathsf{l}, \mathsf{u} \rangle$ be a partial approximation space with a Pawlakian approximation pair. Then, for any $S \subseteq U$ and $u \in U$*

1) $\mu_S^o(u) = 1$ *if and only if*

$$\exists B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \subseteq S) \text{ (i.e., } \mathcal{N}_{\mathfrak{B}}(u) \cap \mathbb{L}(S) \neq \emptyset);$$

2) $\mu_S^a(u) = 1$, $\mu_S^p(u) = 1$ *if and only if*

$$\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \subseteq S) \text{ (i.e., } \mathcal{N}_{\mathfrak{B}}(u) \subseteq \mathbb{L}(S));$$

3) $\mu_S^o(u) > 0$, $\mu_S^a(u) > 0$ *if and only if*

$$\exists B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \cap S \neq \emptyset) \text{ (i.e., } \mathcal{N}_{\mathfrak{B}}(u) \cap \mathbb{U}(S) \neq \emptyset);$$

4) $\mu_S^p(u) > 0$ *if and only if*

$$\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \cap S \neq \emptyset) \text{ (i.e., } \mathcal{N}_{\mathfrak{B}}(u) \subseteq \mathbb{U}(S));$$

5) $\mu_S^o(u), \mu_S^a(u) = 0$ *if and only if*

$$\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \cap S = \emptyset) \text{ (i.e., } \mathcal{N}_{\mathfrak{B}}(u) \cap \mathbb{U}(S) = \emptyset).$$

6) $\mu_S^p(u) = 0$ *if and only if*

$$\exists B \in \mathcal{N}_{\mathfrak{B}}(u)\, (B \cap S = \emptyset).$$

Proposition 4 implies the following statements.

Corollary 3. *Let $\mathsf{GAS}(U)$ be a partial approximation space with a Pawlakian approximation pair. Then, for the optimistic partial membership function μ_S^o,*

1) $\mu_S^o(u) = 1$ *if and only if* $u \in \mathsf{l}(S)$,
2) $\mu_S^o(u) > 0$ *if and only if* $u \in \mathsf{u}(S)$,
3) $0 < \mu_S^o(u) < 1$ *if and only if* $u \in \mathsf{u}(S) \setminus \mathsf{l}(S)$,
4) $\mu_S^o(u) = 0$ *if and only if* $u \in \bigcup \mathfrak{B} \setminus \mathsf{u}(S)$,

for any $S \subseteq U$ and $u \in U$.

Corollary 4. *Let $\mathsf{GAS}(U)$ be a partial approximation space with a Pawlakian approximation pair. Then, for the average partial membership function μ_S^a,*

1) *if* $\mu_S^a(u) = 1$, *then* $u \in \mathsf{l}(S)$,
2) $\mu_S^a(u) > 0$ *if and only if* $u \in \mathsf{u}(S)$,
3) *if* $u \in \mathsf{u}(S) \setminus \mathsf{l}(S)$, *then* $0 < \mu_S^a(u) < 1$,
4) $\mu_S^a(u) = 0$ *if and only if* $u \in \bigcup \mathfrak{B} \setminus \mathsf{u}(S)$,

for any $S \subseteq U$ and $u \in U$.

Corollary 5. *Let $\mathsf{GAS}(U)$ be a partial approximation space with a Pawlakian approximation pair. Then, for the pessimistic partial membership function μ_S^p,*

1) *if* $\mu_S^p(u) = 1$ *then* $u \in \mathsf{l}(S)$,
2) *if* $\mu_S^p(u) > 0$, *then* $u \in \mathsf{u}(S)$,
3) *if* $\mu_S^p(u) > 0$ *and* $u \notin \mathsf{l}(S)$,

   then $u \in \mathsf{u}(S)$ *and* $\mu_S^p(u) < 1$,
4) *if* $u \in \bigcup \mathfrak{B} \setminus \mathsf{u}(S)$, *then* $\mu_S^p(u) = 0$.

for any $S \subseteq U$ and $u \in U$.

The different notions of *necessity* and *possibility* can be found in the definitions of partial membership functions μ_S^*.

The values $\mu_S^*(u)$ ($u \in U$) of the partial membership functions defined above informally mean the following.

The case of optimistic partial membership function:

1) if $\mu_S^o(u) = 1$, i.e., u has at least one neighborhood inside S, u can certainly be classified as belonging to S in an optimistic sense;

[2] $\mathtt{dom}\ f$ denotes the domain of the map f.

2) if $\mu_S^o(u) > 0$, i.e., u has at least one neighborhood wholly or partly inside S, u can possibly be classified as belonging to S in an optimistic sense;
3) if $0 < \mu_S^o(u) < 1$, i.e., u does not have any neighborhood inside S but has at least one neighborhood partly inside and partly outside S, u cannot be classified as either belonging to S or does not belonging to S in an optimistic sense.

The case of the average partial membership function:

1) if $\mu_S^a(u) = 1$, i.e., all neighborhoods of u are inside S, u can certainly be classified as belonging to S in average approach;
2) if $\mu_S^a(u) > 0$, i.e., u has at least one neighborhood wholly or partly inside S, u can possibly be classified as belonging to S in average approach;
3) if $0 < \mu_S^a(u) < 1$, i.e., u has a neighborhood not inside S and has at least one neighborhood wholly or partly inside S, u cannot be classified as either belonging to S or does not belonging to S in average approach.

The case of pessimistic partial membership function:

1) if $\mu_S^p(u) = 1$, i.e., all neighborhoods of u are inside S, u can certainly be classified as belonging to S in a pessimistic sense;
2) if $\mu_S^p(u) > 0$, i.e., all neighborhoods of u are wholly or partly inside S, u can possibly classified as belonging to S in a pessimistic sense;
3) if $0 < \mu_S^p(u) < 1$, i.e., u has a neighborhood not inside S and all neighborhoods of u are wholly or partly inside S, u cannot be classified as either belonging to S or does not belonging to S in a pessimistic sense.

Last, for all three partial membership functions,

$$\mu_S^*(u) = \mathsf{undefined}$$

indicates that we do not have any information about u. Consequently, defining membership degree for u should be meaningless with respect to our knowledge about the objects of the universe.

In classical rough set theory, lower and upper approximations and the boundary can be reconstructed setting out from the membership function. In a fuzzy context, the reconstruction can be carried out by means of `core` and `support` of membership functions in a standard way.

As usual, for the partial membership function μ_S^*, the `core` and `support` are the following:

$$\begin{aligned}\mathtt{core}(\mu_S^*) &= \{u \in U \mid \mu_S^*(u) = 1\};\\ \mathtt{support}(\mu_S^*) &= \{u \in U \mid \mu_S^*(u) > 0\}.\end{aligned}$$

Now, $\mathsf{l}^*, \mathsf{u}^* : 2^U \to 2^U$ approximation pair may be defined as usual:

$$\begin{aligned}\mathsf{l}^*(S) &= \mathtt{core}(\mu_S^*) = \{u \in U \mid \mu_S^*(u) = 1\},\\ \mathsf{u}^*(S) &= \mathtt{support}(\mu_S^*) = \{u \in U \mid \mu_S^*(u) > 0\}.\end{aligned}$$

1) The case of optimistic partial membership functions: In the case of the optimistic partial membership function μ_S^o, the *optimistic lower and upper approximation pair* is the following:

$$\begin{aligned}\mathsf{l}^o(S) &= \mathtt{core}(\mu_S^o) = \{u \in U \mid \mu_S^o(u) = 1\}\\ &= \{u \in \mathsf{l}(S) \mid \exists B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq S)\}\\ &= \mathsf{l}(S)\end{aligned}$$

by Corollary 3 (1), and

$$\begin{aligned}\mathsf{u}^o(S) &= \mathtt{support}(\mu_S^o) = \{u \in U \mid \mu_S^o(u) > 0\}\\ &= \{u \in \mathsf{u}(S) \mid \exists B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \cap S \neq \emptyset)\}\\ &= \mathsf{u}(S)\end{aligned}$$

by Corollary 3 (2).

Informally, $\mathsf{l}^o(S)$ is a collection of such $u \in U$ which has at least one neighborhood included in S, and $\mathsf{l}^o(S) = \mathsf{l}(S)$. $\mathsf{u}^o(S)$ is a collection of such $u \in U$ which has at least one neighborhood having nonempty intersection with S, and $\mathsf{u}^o(S) = \mathsf{u}(S)$.

In other words, in the case of optimistic partial membership function μ_S^o, we get back the Pawlakian approximation pair $\langle \mathsf{l}, \mathsf{u}\rangle$. It implies that $\langle \mathsf{l}^o, \mathsf{u}^o\rangle$ meets the minimum requirements *(C0)–(C3)* and the conditions *(C4)–(C5)*.

2) The case of average partial membership functions: In the case of the average partial membership function μ_S^o, the *average lower and upper approximation pair* is the following:

$$\begin{aligned}\mathsf{l}^a(S) &= \mathtt{core}(\mu_S^a) = \{u \in U \mid \mu_S^a(u) = 1\}\\ &= \{u \in U \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq S)\}\\ &\subseteq \mathsf{l}(S)\end{aligned}$$

by Corollary 4 (1), and

$$\begin{aligned}\mathsf{u}^a(S) &= \mathtt{support}(\mu_S^a) = \{u \in U \mid \mu_S^a(u) > 0\}\\ &= \{u \in \mathsf{u}(S) \mid \exists B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \cap S \neq \emptyset)\}\\ &= \mathsf{u}(S)\end{aligned}$$

by Corollary 4 (2).

Informally, $\mathsf{l}^a(S)$ is a collection of such a $u \in U$ whose all neighborhoods included in S, and $\mathsf{l}^a(S) \subseteq \mathsf{l}(S)$. $\mathsf{u}^a(S)$ is a collection of such a $u \in U$ which has at least one neighborhood having nonempty intersection with S, and $\mathsf{u}^a(S) = \mathsf{u}(S)$.

That is, in the case of average partial membership function μ_S^a, we get back the upper Pawlakian approximation map, but the Pawlakian lower approximation map has already changed.

Proposition 5. $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}^a, \mathsf{l}^a, \mathsf{u}^a\rangle$ *is a weak general approximation space provided that* $\mathfrak{D}_1 \setminus \mathfrak{D}_2 \in \mathfrak{D}_{\mathfrak{B}}^a$ *(*$\mathfrak{D}_1, \mathfrak{D}_2 \in \mathfrak{D}_{\mathfrak{B}}$*).*

Proof:

(C0)–(C2) They are straightforward.

(C3) If $u \in \mathsf{l}^a(S)$, then $\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq S)$, and so $\exists B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \cap S \neq \emptyset)$, i.e., $u \in \mathsf{u}^a(S)$.

■

3) The case of pessimistic partial membership functions: In the case of the pessimistic partial membership function μ_S^p, the *pessimistic lower and upper approximation pair* is the following:

$$\begin{aligned} \mathsf{l}^p(S) &= \mathtt{core}(\mu_S^p) = \{u \in U \mid \mu_S^p(u) = 1\} \\ &= \{u \in U \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq S)\} \\ &\subseteq \mathsf{l}(S) \end{aligned}$$

by Corollary 5 (1), and

$$\begin{aligned} \mathsf{u}^p(S) &= \mathtt{support}(\mu_S^p) = \{u \in U \mid \mu_S^p(u) > 0\} \\ &= \{u \in U \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \cap S \neq \emptyset)\} \\ &\subseteq \mathsf{u}(S) \end{aligned}$$

by Corollary 5 (2).

Informally, $\mathsf{l}^p(S)$ is a collection of such $u \in U$ whose all neighborhoods included in S, and $\mathsf{l}^p(S) \subseteq \mathsf{l}(S)$. $\mathsf{u}^p(S)$ is a collection of such $u \in U$ whose all neighborhoods having nonempty intersection with S, and $\mathsf{u}^p(S) \subseteq \mathsf{u}(S)$.

In the case of pessimistic partial membership function μ_S^p, both lower and upper Pawlakian approximation maps have changed.

Proposition 6. $\mathsf{GAS}(U) = \langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}^p, \mathsf{l}^p, \mathsf{u}^p \rangle$ *is a weak general approximation space provided that* $\mathfrak{D}_1 \setminus \mathfrak{D}_2 \in \mathfrak{D}_{\mathfrak{B}}^p$ *(*$\mathfrak{D}_1, \mathfrak{D}_2 \in \mathfrak{D}_{\mathfrak{B}}$*).*

Proof:

(C0)–(C2) They are straightforward.

(C3) If $u \in \mathsf{l}^p(S)$, then $\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq S)$, and so $\forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \cap S \neq \emptyset)$, i.e., $u \in \mathsf{u}^p(S)$.

■

The next proposition deals with the conditions *(C4)–(C5)* of average and pessimistic approximation pairs.

Proposition 7. *Let* $\langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}^a, \mathsf{l}^a, \mathsf{u}^a \rangle$ *and* $\langle U, \mathfrak{B}, \mathfrak{D}_{\mathfrak{B}}^p, \mathsf{l}^p, \mathsf{u}^p \rangle$ *be weak approximation spaces whose components are defined as above.*

If the base system $\mathfrak{B}$ *is one–layered,* $\mathfrak{D}_{\mathfrak{B}}^{\mathfrak{a}} = \mathfrak{D}_{\mathfrak{B}}^p = \mathfrak{D}_{\mathfrak{B}}$ *and the weak approximation pairs* $\langle \mathsf{l}^a, \mathsf{u}^a \rangle$ *and* $\langle \mathsf{l}^p, \mathsf{u}^p \rangle$ *are standard, i.e.,* $\mathsf{l}^a(D) = D$ *and* $\mathsf{l}^p(D) = D$ *for all* $D \in \mathfrak{D}_{\mathfrak{B}}$.

Proof:

Since l is standard, $\mathsf{l}^a(D) \subseteq \mathsf{l}(D) = D$ for all $D \in \mathfrak{D}_{\mathfrak{B}}$.

On the other hand, $\mathfrak{B}$ is one–layered, and so every definable set $D \in \mathfrak{D}_{\mathfrak{B}}$ is a finite union of pairwise disjoint base sets, e.g., $D = B_1 \cup \cdots \cup B_n$, where B_i's are pairwise disjoint. Moreover, for every $u \in D$ there exists exactly one $i \in \{1, 2, \ldots, n\}$ in such a way that $\mathcal{N}_{\mathfrak{B}}(u) = \{B_i\}$.

Hence, we get for all $D \in \mathfrak{D}_{\mathfrak{B}}$,

$$\begin{aligned} \mathsf{l}^a(D) &= \{u \in U \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq D)\} \\ &\supseteq \{u \in D \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq D)\} \\ &= \{u \in B_1 \cup \cdots \cup B_n \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq D)\} \\ &= \{u \in B_1 \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq D)\} \\ &\quad \cup \cdots \cup \{u \in B_n \mid \forall B \in \mathcal{N}_{\mathfrak{B}}(u)\,(B \subseteq D)\} \\ &= B_1 \cup \cdots \cup B_n = D. \end{aligned}$$

Therefore, $\mathsf{l}^a(D) = D$.

The standard property of l^p can be proved similarly.

■

IV. Some remarks on the logical applications

In the previous sections, first, three partial membership functions have been defined in partial approximation spaces with Pawlakian approximation pairs, then three approximation pairs have been generated with the help of them. It has been shown that, among others, they meet the minimum requirements prescribed for approximation pairs in partial approximation spaces, i.e., they actually form approximation pairs.

Optimistic, average and pessimistic partial membership functions have already been studied by the second author from the logical point of view in [38], [32]. It turned out that they are in connection with *decision–theoretic rough set models* (DTRS) which can be considered as the probabilistic extensions of algebraic rough set models [37].

Optimistic, average and pessimistic partial membership functions may serve as a bases of the semantics of a partial first–order logic. In the paper [35], the semantic system of a partial first–order logic with three different types of partial membership functions is presented. The proposed logical system gives an exact possibility to introduce different semantic notions of logical consequence relations which can be used in order to make clear the consequences of our decisions.

V. Conclusion and future work

In this paper, having defined three partial membership functions, three approximation pairs have been generated in partial approximation spaces with Pawlakian approximation pairs. We have investigated how these pairs meet the requirements prescribed for approximation pairs in partial approximation spaces. As a result, in this way we have constructed two not Pawlakian approximation pairs.

In the future, it should be worth performing similar investigations in partial approximation spaces setting out from *arbitrary* approximation pairs, in particular, which have been obtained in this paper.

Acknowledgment

The publication was supported by the TÁMOP–4.2.2.C–11/1/KONV–2012–0001

project. The project has been supported by the European Union, co–financed by the European Social Fund.

The authors are thankful to the anonymous referees for valuable suggestions.

References

[1] Z. Pawlak, "Information systems theoretical foundations," *Information Systems*, vol. 6, no. 3, pp. 205–218, 1981.

[2] ——, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[3] ——, *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.

[4] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[5] R. R. Yager, S. Ovchinnikov, R. M. Tong, and H. T. Nguyen, Eds., *Fuzzy sets and applications*. New York, NY, USA: Wiley-Interscience, 1987.
[6] B. Yuan and G. J. Klir, Eds., *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1996.
[7] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. New Jersey: Prentice Hall, 1995.
[8] D. Dubois and H. Prade, *Fuzzy sets and systems - Theory and applications*. New York: Academic press, 1980.
[9] Y. Y. Yao, "A comparative study of fuzzy sets and rough sets," *Information Sciences*, vol. 109, pp. 21–47, 1998.
[10] Z. Pawlak and A. Skowron, "Rough membership functions," in *Advances in the Dempster-Shafer theory of evidence*, R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Eds. New York, NY, USA: John Wiley & Sons, Inc., 1994, pp. 251–271.
[11] L. Polkowski, *Rough Sets: Mathematical Foundations*, ser. Advances in Intelligent and Soft Computing. Physica-Verlag, A Springer-Verlag Company, 2002.
[12] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, "Rough sets: A tutorial," in *Rough Fuzzy Hybridization. A New Trend in Decision-Making*, S. Pal and A. Skowron, Eds. Singapore: Springer-Verlag, 1999, pp. 3–98.
[13] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 23, pp. 3–18, 1987.
[14] ——, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General Systems*, vol. 17, no. 2-3, pp. 191–209, 1990.
[15] ——, "Putting rough sets and fuzzy sets together," in *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Set Theory*, R. Slowinski, Ed. Kluwer Academic, Dordrecht, 1992, pp. 203–232.
[16] M. Chakraborty, "On fuzzy sets and rough sets from the perspective of indiscernibility," in *Logic and Its Applications. 4th Indian Conference, ICLA 2011 Delhi, India, January 5-11, 2011, Proceedings*, ser. LNAI, M. Banerjee and A. Seth, Eds., vol. 6521. Berlin Heidelberg: Springer-Verlag, 2011, pp. 22–37.
[17] Y. Yao, "Probabilistic rough set approximations," *International Journal of Approximate Reasoning*, vol. 49, no. 2, pp. 255–271, 2008.
[18] Z. Pawlak, L. Polkowski, and A. Skowron, "Rough sets: An approach to vagueness," in *Encyclopedia of Database Technologies and Applications*, L. C. Rivero, J. H. Doorn, and V. E. Ferraggine, Eds. Hershey, PA: Idea Group Inc., 2005, pp. 575–580.
[19] J. Łukasiewicz, "Die logischen grundlagen der wahrscheinlichkeitsrechnung (1913)," in *Jan Łukasiewicz - Selected Works*, L. Borkowski, Ed. Amsterdam, Warsaw: Polish Scientific Publishers and North-Holland Publishing Company, 1970.
[20] Y. Yao, "Semantics of fuzzy sets in rough set theory," in *Transactions on Rough Sets II*, ser. LNCS, J. F. Peters, A. Skowron, D. Dubois, J. W. Grzymała-Busse, M. Inuiguchi, and L. Polkowski, Eds. Springer Berlin Heidelberg, 2005, vol. 3135, pp. 297–318.
[21] A. Skowron and J. Stepaniuk, "Tolerance approximation spaces," *Fundamenta Informaticae*, vol. 27, no. 2-3, pp. 245–253, 1996.
[22] Z. Bonikowski, E. Bryniarski, and U. Wybraniec-Skardowska, "Extensions and intentions in the rough set theory," *Information Sciences*, vol. 107, no. 1-4, pp. 149–167, 1998.
[23] Y. Y. Yao, "On generalizing rough set theory," in *Proceedings of RSFDGrC 2003*, ser. LNAI 2639. Berlin Heidelberg: Springer-Verlag, 2003, pp. 44–51.
[24] Z. Pawlak and A. Skowron, "Rough sets: Some extensions," *Information Sciences*, vol. 177, pp. 28–40, 2007.
[25] W. Zhu and F.-Y. Wang, "On three types of covering-based rough sets," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1131–1144, 2007.
[26] W. Zhu, "Topological approaches to covering rough sets," *Information Sciences*, vol. 177, no. 6, pp. 1499–1508, 2007.
[27] ——, "Relationship between generalized rough sets based on binary relation and covering," *Information Sciences*, vol. 179, no. 3, pp. 210–225, 2009.
[28] Z. Csajbók and T. Mihálydeák, "Partial approximative set theory: A generalization of the rough set theory," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp. 437–444, 2012.
[29] ——, "A general set theoretic approximation framework," in *Proceedings of IPMU 2012, Catania, Italy, July 9-13, 2012, Part I*, ser. CCIS, S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, Eds., vol. 297. Berlin Heidelberg: Springer-Verlag, 2012, pp. 604–612.
[30] Z. Csajbók, "Approximation of sets based on partial covering," *Theoretical Computer Science. Special Issue on Rough Sets and Fuzzy Sets in Natural Computing*, vol. 412, no. 42, pp. 5820–5833, 2011.
[31] Z. E. Csajbók, "Approximation of sets based on partial covering," in *Transactions on Rough Sets*, ser. LNCS, Transactions on Rough Sets XVI, J. F. Peters, A. Skowron, S. Ramanna, Z. Suraj, and X. Wang, Eds., vol. 7736. Heidelberg: Springer, 2013, pp. 144–220.
[32] T. Mihálydeák, "Partial first–order logic with approximative functors based on properties," in *Rough Sets and Knowledge Technology. 7th International Conference, RSKT 2012, Chengdu, China, August 17-20, 2012, Proceedings*, ser. LNAI, T. Li, H. S. Nguyen, G. Wang, J. Grzymala-Busse, R. Janicki, A. E. Hassanien, and H. Yu, Eds., vol. 7414. Berlin Heidelberg: Springer-Verlag, 2012, pp. 514–523.
[33] Z. Csajbók and T. Mihálydeák, "Partial approximative set theory: A generalization of the rough set theory," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp. 437–444, 2012.
[34] Y. Y. Yao, "Granular computing using neighborhood systems," in *Advances in Soft Computing: Engineering Design and Manufacturing. The 3rd On-line World Conference on Soft Computing (WSC3)*, R. Roy, T. Furuhashi, and P. K. Chawdhry, Eds. London: Springer-Verlag, 1999, pp. 539–553.
[35] T. Mihálydeák, "Partial first–order logic relying on optimistic, pessimistic and average partial membership functions." in *Proccedings of the 8th conference of the European Society for Fuzzy Logic and Technology, EUSFLAT-2013*, in the fall of 2013, Forthcoming.
[36] Y. Y. Yao and J. P. Zhang, "Interpreting fuzzy membership functions in the theory of rough sets," in *Rough Sets and Current Trends in Computing*, ser. Lecture Notes in Computer Science, W. Ziarko and Y. Y. Yao, Eds., vol. 2005. Springer, 2000, pp. 82–89.
[37] Y. Y. Yao, "Decision-theoretic rough set models," in *Rough Sets and Knowledge Technology. Second International Conference, RSKT 2007, Toronto, Canada, May 14-16, 2007. Proceedings*, ser. Lecture Notes in Computer Science, J. T. Yao et al., Eds., vol. 4481. Springer, 2007, pp. 1–12.
[38] T. Mihálydeák, "Partial firstorder logical semantics based on approximations of sets," in *Non-classical Modal and Perdicate Logics 2011, Guangzhou (Canton), China, F solutions, Prague*, P. Cintula, S. Ju, and M. Vita, Eds., 2011, pp. 85–90.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 43–49

Comparison of Selected Textural Features as Global Content-Based Descriptors of VHR Satellite Image – the EROS-A Study

Wojciech Drzewiecki
AGH University
of Science and Technology,
Al. Mickiewicza 30,
30-059 Krakow, Poland
Email: drzewiec@agh.edu.pl

Anna Wawrzaszek,
Michał Krupiński,
Sebastian Aleksandrowicz
Space Research Centre of the Polish Academy of Sciences, Bartycka 18A, 00-716 Warszawa, Poland
Email: {sanna, mkrupinski, saleksandrowicz}
@cbk.waw.pl

Katarzyna Bernat
AGH University
of Science and Technology,
Al. Mickiewicza 30,
30-059 Krakow, Poland
Email: kbernat@agh.edu.pl

***Absrtact*—Texture is considered as one of the most crucial image features used commonly in computer vision. It is important source of information about image content, especially for single-band images. In this paper we present the results of research carried out to assess the usefulness of selected textural features of different groups in panchromatic very high resolution (VHR) satellite image classification. The study is based on images obtained from EROS A satellite. The aim of our tests was to estimate and compare the accuracy of main land cover types classification, with a particular focus on determining usefulness of textural features based on multifractal formalism.**

Presented research confirmed that it is possible to use the textural features as efficient global descriptors of VHR satellite image content. It was also prove that multifractal parameters should be considered as valuable textural features in the context of land cover classification.

I. Introduction

TEXTURE as a primary factor of visual perception is a necessary feature of image description. It is usually easy to recognize texture, but it is more difficult to define it, because texture, in contrast to colour, is not determined by a single point, but involve neighbouring area and can be related to a direction or a scale. It was created large spectrum of parameters, due to a lot of possible textural descriptors, to help to extract information about texture (also in the context of satellite images [1]). Textural characteristic can be calculated based on the entire image (global features), fragments of this image delineated by segmentation results or small clusters of pixels formed by moving windows [2], [3], [4]. Different texture analysis techniques, such as Markovian analysis (including Haralick measures), spatial autocorrelation, multi-scale autoregressive models, wavelet transforms or fractals have been successfully used to describe the content of the images [2], [3], [5]. They are considered especially important in case of single-band images, like medical ones.

This work was supported by Polish Natonal Science Centre (NCN) and the Ministry of Science and Higher Education (MNiSW) through grant NN 526 1568 40.

The textural analysis becomes also an important component of the process of information extraction from satellite images, especially in Object-Based Image Analysis approach where textural features supplement the set of typical characteristics obtained from histogram features and image objects' shape. However, it is even more valuable tool when single-band panchromatic images of very high spatial resolution are considered. Textural analysis may facilitate information extraction from such images by enabling the automatic classification of their content. It is also especially important from the content-based image retrieval (CBIR) point of view. Due to the increase of high resolution remote sensing imagery, the developments in this direction are particularly desirable.

In our work we propose multifractal formalism, as a generalization of the fractal geometry, in order to more complete analyze the texture of satellite images [6]. In our previous study [7] we compared efficiency of selected textural features as global content-based descriptors of panchromatic WorldView2 Very High Resolution satellite images. We wanted to investigate how accurately remotely sensed image can be automatically classified to the broad land cover types such as agriculture, urban areas, water bodies, and forest, based on textural information derived from the entire image. We were able to construct decision trees capable for very accurate classification of images from our test image database into these main landuse categories. The research proved that degree of multifractality can be considered as important global image characteristic.

However, tested WorldView2 images were characterised by very homogeneous landuse – over 90% of the image was in the dominating landuse category – and high radiometric quality. In case of the present study we applied the same methodology of analysis to panchromatic EROS A satellite images. This is also a VHR sensor, although older then WorldView2 and acquiring images with a little bit lower spatial resolution (2 m in case of EROS A vs. 0.5 m in case of WorldView2) and higher level of noise. Moreover, images in our EROS A test image database are not such homogeneous regarding the landuse of imaged terrain – the dominat-

ing landuse category covers over 50% of the image. Our previous work [6] showed that selected multifractal parameters can be used as features describing the content of these images. The aim of the present study is to compare their efficiency with other textural parameters. We also intend to carry out experiment to determine the parameters most appropriate as classification features.

II. Data

The test data used in the experiment are the same like used in [6] and consist of two partially overlapping EROS-A scenes (panchromatic high-resolution images, ground resolution – 2 m) of Krakow area (south Poland) acquired on 10th and 15th October 2007. Images were acquired using similar pointing angles (29 and 27.5 degrees) and ground sample distances (2.20 and 2.15 m respectively). The images were orthorectified using orbital model.

We created our testing image database from the 512x512 pixel orthoimage tiles. To make more image tiles available for the study, the ortophotomaps were cut into tiles twice. The second set of tiles was cut with the origin of tiles shifted 256 pixels east and 256 pixels north. Every image in the created database was labeled according to its prevailing land cover category (agriculture, forest, urban), based on photo interpretation done in other studies for the purpose of landscape ecological research. Only images where dominating land cover class covered over 50% of imaged area were used for analysis.

The final database consisted from following sets of images:

- Image set EROS1 – 262 image tiles cut from Scene 1 (agriculture – 199, urban – 40, forest - 23);
- Image set EROS1s – 259 image tiles cut from Scene 1 with shifted origins (agriculture – 204, urban – 35, forest - 20);
- Image set EROS2 – 344 image tiles cut from Scene 2 (agriculture – 298, urban – 25, forest - 21);
- Image set EROS2s – 349 image tiles cut from Scene 1 with shifted origins (agriculture – 308, urban – 24, forest – 17).

III. Methods

The analytical approach adopted in the study is the same as in [7].

A. Textural parameters

Chosen global textural characteristics were calculated for every image chip. As the result every image in the database was described by 295 attributes, which may be grouped into 9 attribute groups (AG):

- AG1: the label (land cover class);
- AG2: four histogram-based characteristics (mean, variance, skewness and kurtosis);
- AG3: six multifractal parameters ($\Delta^{SUM,}$ Δ^{MAX}, Δ^{BCD}, Δ_p^{SUM}, Δ_p^{MAX}, Δ_p^{BCD}); Δ stands for the degree of multifractality and Δ_p for the degree of multifractality for $q > 0$; SUM, MAX and BCD are three different measure types (measure SUM takes sum of pixel values on a given box; measure MAX choses maximum value of pixels in a given box; measure BCD takes deviation of gray levels in a box) [8], [6].
- AG4: 220 co-occurrence matrix-based parameters [9], [10]: angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy; these parameters were computed 20 times, for (d,0), (0,d), (d,d), (d,-d) where the distance d can take values of 1, 2, 3, 4, and 5;
- AG5: 20 run length matrix-based parameters [9], [11]: run length nonuniformity, gray level nonuniformity, long run emphasis moment, short run emphasis inverse moment, fraction of image in runs; these parameters were computed 4 times (for vertical, horizontal, 45-degree and 135-degree directions);
- AG6: 5 absolute gradient-based parameters [9]: mean absolute gradient, variance of absolute gradient, skewness of absolute gradient, kurtosis of absolute gradient, percentage of pixels with nonzero gradient;
- AG7: 5 autoregressive model parameters [12], [13]: θ_1, θ_2, θ_3, θ_4, σ;
- AG8: 20 parameters derived from wavelet analysis [14], [15], [16];
- AG9: fractal dimension determined by using differential boxing-counting (DBC) method [17].

Features from attribute groups AG4 – AG8 were obtained using MaZda software [16]. Histogram-based features, multifractal parameters and fractal dimension were calculated in MatLab.

B. Fractal Dimension and Multifractal Parameters

There are several methods for estimating a fractal dimension (FD) in an image [18], [19]. In the most commonly used Box-Counting methods fractal dimension is calculated by covering an object with boxes of varying size l and is given by the relation:

$$D_F = \lim_{l \to 0} \frac{\ln N(l)}{\ln 1/l} \qquad (1)$$

where $N(l)$ denotes the number of boxes of size l needed to cover considered object. Methods differ mainly in the ways they approximate the quantity $N(l)$. Most of them are applied to images that must be turned into binary images.

In our work we calculate fractal dimension by using differential box-counting (DBC) method [20], [17]. This

method, proposed by Sarkar and Chaudhuri [20], allows working directly on grey-scale images and thus the binarization process is avoided.

In DBC algorithm an image of size $M \times M$ is considered as a three-dimensional spatial surface, where (x,y) denotes pixel position and the third coordinate (z) denotes pixel gray level. The (x,y) plane is partitioned into grids of size $s \times s$, where $M/2 \geqslant s > 1$ and s is an integer. On each grid there is a column of boxes of size $s \times s \times s'$,where s' is the height of each box, $G/s' = M/s$, and G is the total number of gray levels. Let the minimum and maximum gray level of the image in (i,j)-th grid fall into the kth and lth boxes, respectively [20]. Then the contribution of $N(l)$ in the pixel (i,j) of the grid is $n_l(i,j)=l-k+1$. Taking contributions from all grids, we have

$$N(l)=\sum_{i,j} n_l(i,j) \tag{2}$$

Then $N(l)$ is computed to different values of l. Finally, the fractal dimension D_F is estimated from the least square linear fit of log $(N(l))$ against log $(1/l)$ (see Equation 1). It is worth noting that presented DBC methods was compared with other four methods proposed by Peleg [21], Pentland [22], Gangepain and Roques-Carmes [23], and Keller et al. [24], respectively. The DBC method was considered as a better method, as was also supported by the other investigation [25]. Moreover, some modifications of DBC method have been lately proposed [26].

In our research we also consider one of the multifractal functions: generalized dimensions, D_q, as well quantitative parameter strictly connected with this function. The generalized dimensions D_q are calculated as a function of a continuous index q, where $-\infty < q < \infty$ (e.g., see [27], figure 3.1). Index q can be compared to a microscope for exploring different regions of the considered image.

As for FD estimation, many methods exist to obtain the multifractal functions [20]. Here, the Box-Counting based moment method has been applied [28]. In the first step of analysis an image is divided into boxes of size $\delta \times \delta$. Next, for each box following multifractal measure is calculated:

$$\mu_i(\delta)=\frac{p_i(\delta)}{\sum_{i=1}^{N(\delta)} p_i(\delta)}, \tag{3}$$

where $i = 1, ..., N(\delta) = 2^n$ labels the individual boxes of size δ. Here $p_i(\delta)$ denotes three different measures [29], [8]:

$$p_i^{\mathrm{SUM}}(\delta)=\sum_{(k,l)\in\Omega_i} g(k,l) \tag{4}$$

$$p_i^{\mathrm{MAX}}(\delta)=\max_{(k,l)\in\Omega_i} g(k,l) \tag{5}$$

$$p_i^{\mathrm{BCD}}(\delta)=\max_{(k,l)\in\Omega_i} |d(k,l)| \tag{6}$$

where $g(k,l)$ is a gray-scale intensity at point (k,l), Ω_i is a set of all pixels (k,l) in the ith box and $d(k,l)$ denotes the deviation of gray levels in box i.

In the next step of our analysis, a weighted summation is performed over all boxes in a particular grid returning the partition function of order q, i.e.

$$\chi(q,\delta)=\sum_{i=1}^{N(\delta)} (\mu_i(\delta))^q \tag{7}$$

which scales with the box length $\delta \to 0$ and $N(\delta) \to \infty$ according to:

$$\chi(q,\delta) \propto \delta^{D_q(q-1)} \tag{8}$$

From the Equation (8) we obtain generalized dimensions [30]

$$D_q=\lim_{\delta\to 0}\frac{\log(\chi(q,\delta))}{(q-1)\log(\delta)} \tag{9}$$

The difference of the maximum and minimum dimension D_q, associated with the least dense and the most dense regions in the considered measure, is given by

$$\Delta = D_{-\infty} - D_{+\infty} \tag{10}$$

and can be regarded as a degree of multifractality (e.g., [31], [27]). The degree of multifractality Δ is a measure of complexity of considered data; higher values of Δ inform us about greater non homogeneity on image and suggest that different fractals are needed for its full description. In particular, for monofractal scaling the degree of multifractality equals zero.

Finally, as a result of multifractal analysis performed for each image we obtain the following set of six parameters: the degree of multifractality (Δ) for measure SUM (Δ^{SUM}), MAX (Δ^{MAX}) and BCD (Δ^{BCD}), as well the degree of multifractality for positive values of index q (Δ_p^{SUM}, Δ_p^{MAX} and Δ_p^{BCD}). Presented parameters state quantitative and global characteristics used to compare complexity of images.

C. Classification

The decision (classification) tree approach was used for classification. We decided to use this method as it has good computational efficiency and the obtained tree can be presented as a set of easily interpretable rules. It has also been already successfully applied for the semantic labeling of satellite images [32], [33].

In our study, the classification was done using See5 software (Rel. 2.07). The software generates decision trees based on C5.0 algorithm, improved commercial version of well-known C4.5 [24].

Classification was done based on different sets of features (classification features sets, CFS):

- CFS1: all classification features (AG2 – AG9);
- CFS2: all classification features apart from histogram-based ones (AG3 – AG9);
- CFS3: co-occurrence matrix-based features (AG4);
- CFS4: co-occurrence matrix-based and histogram-based features (AG2 and AG4);
- CFS5: run length matrix-based features (AG5);
- CFS6: run length matrix-based and histogram-based features (AG2 and AG5);
- CFS7: absolute gradient-based features (AG6);
- CFS8: absolute gradient-based and histogram-based features (AG2 and AG6);
- CFS9: autoregressive model parameters (AG7);
- CFS10: autoregressive model parameters and histogram-based features (AG2 and AG7);
- CFS11: parameters derived from wavelet analysis (AG8);
- CFS12: parameters derived from wavelet analysis and histogram-based features (AG2 and AG8);
- CFS13: fractal dimension and histogram-based features (AG2 and AG9);
- CFS14: histogram-based features (AG2);
- CFS15: multifractal parameters (AG3);
- CFS16: multifractal parameters and histogram-based features (AG2 and AG3);

Such approach enabled us both, to evaluate the individual performance of each group of textural characteristics (used alone and together with histogram-based features) and to assess the usefulness of combining of features from different groups.

Five approaches with different pruning and thresholding options as well as with or without winnowing of attributes were used for every set of classification features. Boosting with ten trials was used in every classification run.

In the area where satellite scenes overlap, existed some number of 'twin' tiles covering exactly the same area. To eliminate the possibility of using them as training and test data at the same time, in our study we have used the image data sets as shown in Table I.

The average overall classification accuracy was calculated for each classification approach and each set of tested classification features (CFS). The lowest of five classification errors was then assigned as a measure of classification quality for particular tested set of attributes.

TABLE I.
TRAINING AND TEST DATA SETS

	Training data set	Test data set
Classification 1	EROS1	EROS2s
Classification 2	EROS2s	EROS1
Classification 3	EROS1 and EROS2s	EROS1s and EROS2

IV. RESULTS AND DISCUSSION

The results of classification tests are shown in Table II. The best results for each of classification tests gave the classification accuracies in the range from 94 to 96 percent. In two classification tests the best results were obtained when all calculated textural parameters were included in classification feature set. However, in Classification 2 the best result was achieved using the classification feature set consisted only of absolute gradient-based and histogram-based features (CFS8). This kind of textural features is also the best one, when looking into the performance of particular texture attribute groups. This result is surprising as in the previous study for WorlView2 images [7], this attribute group gave rather poor results when compared to other ones. It should be noticed however, that the achieved level of accuracy was quite similar (93 – 95% in case of EROS A and 93% in case of WorldView2). The performance of other textural characteristics was much worse in the actual study. It is especially the case of autoregressive model parameters (CFS9), which performance for EROS A images classification can be pointed as the worst one. For WorldView2 images this group of attributes was between the best ones [7].

The results presented in Table II show that, in general, the classification performance increases when textural features are combined with histogram-based ones. This conclusion is consistent with the results of our previous studies [6], [7].

Multifractal parameters used together with histogram-based features gave the second-best result for Classification 2. However, in the two other tests their performance was rather average.

When features from all attribute groups are combined (CFS1 and CFS2) the classification tree built may be quite complex as many of available features can be used in classification process. In the classification method used in our study, the number of features used for classification may be reduced by using the winnowing approach. We compared the results of the overall accuracy achieved for both options (without winnowing and winnowed) in Table III.

It should be noted, that for two of three tests the classification performance of winnowed set is comparable to the full one and still better then performance of any other classification features set. In these cases the final set of winnowed attributes is the same and consists of two co-occurrence matrix-based parameters, multifractal parameter and skewness of absolute gradient. This result is very similar to achieved for WorldViev2 images, where also the result based on three winnowed parameters was better then any of the results of particular attribute groups [7]. In case of that study the set of winnowed attributes consisted of multifractal parameter (Δ_p-$^{MAX.}$) skewness of absolute gradient and the feature derived from autoregressive model (σ).

TABLE II.
CLASSIFICATION RESULTS

Classification 1		**Classification 2**		**Classification 3**	
Classification feature set	Overall clasification error [%]	Classification feature set	Overall clasification error [%]	Classification feature set	Overall clasification error [%]
CFS 1	5.5	CFS 8	5.8	CFS 2	3.8
CFS 2	5.5	CFS 16	6.9	CFS 1	4.4
CFS 8	7.0	CFS 7	9.3	CFS 3	4.9
CFS 3	7.3	CFS 2	9.7	CFS 11	4.9
CFS 11	7.3	CFS 6	9.7	CFS 8	5.1
CFS 4	7.8	CFS 12	9.7	CFS 12	5.1
CFS 13	7.8	CFS 1	10.0	CFS 4	5.4
CFS 12	8.7	CFS 3	10.0	CFS 10	5.9
CFS 15	8.7	CFS 4	10.0	CFS 6	6.1
CFS 16	8.7	CFS 14	10.0	CFS 13	6.4
CFS 7	9.3	CFS 15	10.4	CFS 16	6.9
CFS 6	9.9	CFS 13	12.2	CFS 15	7.4
CFS 10	10.2	CFS 5	13.5	CFS 7	7.7
CFS 14	11.0	CFS 11	16.6	CFS 5	8.2
CFS 9	14.0	CFS 10	21.6	CFS 14	10.6
CFS 5	14.5	CFS 9	43.6	CFS 9	13.7

TABLE III.
INFLUENCE OF ATTRIBUTES WINNOWING ON CLASSIFICATION RESULTS

	Overall classification error [%] without winnowing	Overall classification error [%] with winnowing	Winnowed attributes
Classification 1	5.5	5.5	S(5, 5) Contrast Δ_p^{SUM} S(1, 0) Difference Variance skewness of absolute gradient
Classification 2	9.7	12.0	S(3, -3) Difference Entropy S(0, 1) Entropy S(5, 0) Sum Entropy Horzl_ Long Run Emphasis Moment 45dgr_ Short Run Emphasis Inverse Moment
Classification 3	3.8	4.7	(5, 5) Contrast Δ_p^{SUM} S(1, 0) Difference Variance skewness of absolute gradient

V. CONCLUSIONS

The aims of the presented study were twofold: (i) to test the usefulness of the selected textural parameters as classification features of panchromatic VHR satellite images and (ii) to compare the efficiency of the multifractal parameters (which we propose for more complete description of the texture of remote sensing images) and other textural features in the context of land cover classification. The present study of EROS A satellite images was a continuation of the research done previously for WorldView2 data [7]. Some results confirmed, but partially the results of this research differ from the earlier ones.

In both studies we prove that for VHR satellite images it is possible to use the textural features as efficient global descriptors of image content. The observed in this study increase in classification accuracy when textural features are supplemented by histogram-based ones was also present in the results of our earlier experiments [6], [7]. Similarly, we noticed earlier the possibility of successful reduction of classification features. It is worth noting, that in both our experiments we were able to reduce the number of classification features from 295 to 3 or 4 with very limited (or even negli-

gible) impact on the overall classification accuracy. This is very important, as calculating of textural parameters for VHR satellite images is very computationally expensive.

When comparing the classification efficiency of different groups of textural parameters, in the present study the best results were obtained for absolute gradient-based features. In the WorldView2 experiment done previously the best accuracy was achieved for multifractal parameters. It is interesting that in the case of the absolute gradient group the level of error obtained in both research is quite similar and for other kinds of textural features the errors are much higher for EROS A classification. In our opinion there are at least two possible sources of such results and much lower value of classification accuracy achieved for EROS A images in general (96.2% comparing to 99.6% for WorldView2).

First of all, the differences are present in the input images themselves. The WorldView2 images have higher spatial resolution (0.5 vs. 2.0 m) and are considered as better radiometrically. The higher level of noise potentially present in EROS A data may deteriorate the quality of textural measures derived from images. It is possible that absolute gradient-based features are less sensitive to the noise present in the imagery.

The second source of differences in the results may be in a different image content. In case of WorldView2 data used in previous experiment, the images were almost entirely (at least in 90%) covered by single land cover class. In present study the EROS A images were labeled based on the prevailing land cover category defined as covering over 50% of imaged area. This could result in much higher complexity of image content, and in turn, in lower classification accuracies.

Both possibilities should be taken into account during further research. Some noise may be added to WorldView2 images, as well as more homogenous EROS A images (or WV2 images of the areas having more complex land cover) may be used.

Presented research prove that for VHR satellite images multifractal parameters should be considered as valuable textural features. Based on this features the second-best classification result was obtained in the one of the three performed tests. For two other tests the multifractal feature (Δ_p^{SUM}) was in the set of the four winnowed attributes, enabling very efficient classification approach. It should be stressed that similar result was obtained also in previous WorldView2 experiment. In both cases, the very limited (and very efficient) sets of textural parameters were chosen by winnowing, containing the multifractal parameter (although not the same one) and the skewness of absolute gradient feature. However, the importance of these parameters for classification of VHR satellite images indicated in reported studies should be proven during further research extended for other VHR satellite sensors and images of different areas.

REFERENCES

[1] P. Howarth, and S. Ruger, "Evaluation of Texture Features for Content-Based Image Retrieval", *Lecture Notes in Compter Science,* vol. 3115, pp. 326-334, 2004.

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", *ACM Computing Surveys,* vol. 40, no.5, pp. 2-60, 2008.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based Image Retrieval at the End of the Early Years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1318, 2000.

[4] C. W. Emerson, D. A. Quattrochi, and N. S. –N. Lam "Spatial Metadata for Remote Sensing Imagery ", in *NASA's Earth Science Technology Conference*, Paolo Alto, 2004.

[5] A. Samal, S. Bhatia, P. Vadlamani, and D. Marx, "Searching satellite imagery with integrated measures", *Pattern Recognition*, vol. 42, pp. 2502-2513, 2009.

[6] W. Drzewiecki, A. Wawrzaszek, S. Aleksandrowicz, and M. Krupinski, „Initial Evaluation of the Applicability of Multifractal Measures as Global Content-based Image Descriptors" in *Proc. of ESA-EUSC-JRC 8th Conference on Information Mining,* Oberpfaffenhofen, 2012.

[7] W. Drzewiecki, A. Wawrzaszek, M. Krupiński, S. Aleksandrowicz, and K. Bernat, "Comparison of Selected Textural Features as Global Content-Based Descriptors of VHR Satellite Image" *IEEE 2013 International Geoscience & Remote Sensing Symposium (IGARSS 2013), Melbourne, Australia,* to be published.

[8] T. Stojić, I. Reljin, and B. Reljin, "Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms" *Physica A,* vol. 367, pp. 494-508, 2006.

[9] R. Haralick, "Statistical and Structural Approaches to Texture" *IEEE Proceedings,* vol. 67, pp. 786-804, 1979.

[10] R. F. Walker, P. T. Jackway, and I. D. Long Staff, "Recent Developments in the Use of the Co-occurrence Matrix for Texture Recognition", in *13th International Conference on Digital Signal Processing Proceedings*, 1997, vol. 1, pp. 63-65.

[11] M. M. Galloway, "Texture Analysis Using Gray Level Run Lengths", *Computer Graphics and Image Processing,* vol. 4, pp. 172-179, 1975.

[12] A. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall International, Englewood Clifs, 1989.

[13] M. Hajek, M. Dezortova, A. Materka, and R. Lerski (eds.), „Texture Analysis for Magnetic Resonance Imaging", Med4publishing, Prague, 2006.

[14] S. G. Mallat, "Multifrequency Channel decompositions of images and wavelet models", *IEEE Transaction on Acoustics, Speech and Signal Processing*, vol. 37, pp. 2091-2110, 1989.

[15] A. Materka, and M. Strzelecki, "Texture Analysis Methods – A Review", Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels, 1998.

[16] P. M. Szczypiński, M. Strzelecki, A. Materka, and A. Klepaczko, „MaZda – A Soffware for image texture analysis", *Computer Methods and Programs in Biomedicine*, vol. 94, pp. 66-76, 2009.

[17] B. B. Chaudhuri, and N. Sarkar, "Texture Segmantation Using Fractal Dimension", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 72-77, 1995.

[18] W. Sun, G. Xu, P. Gong, and S. Liang, "Fractal analysis of remotely sensed images: A review of methods and applications." *International Journal of Remote Sensing,* vol. 27(22), pp. 4963–4990, 2006.

[19] R. Lopes, and N. Betrouni, "Fractal and multifractal analysis: A review.", *Medical Image Analysis,* vol. 13, pp. 634–649, 2009.

[20] N. Sarkar, and B.B. Chaudhuri, "An efficient differential box-counting approach to compute fractal dimension of image" *IEEE Transactions on Systems, Man and Cybernetics,* vol. 24, pp. 115-120, 1994.

[21] S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, pp. 518–523, 1984.

[22] A. P. Pentland, "Fractal-based description of natural scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, pp. 661–674, 1984.

[23] L. Gangepain, and C. Roques-Cames, "Fractal approach to two dimensional and three dimensional surface roughness" Wear 109, pp. 119–126, 1986.

[24] J. Keller, R. Crownover, and S. Chen, "Texture description and segmentation through fractal geometry." *Computer Vision, Graphics, and Image Processing,* vol. 45, pp. 150–166, 1989.

[25] L. Yu , D. Zhang, K. Wang, and W. Yang, "Coarse iris classification using box-counting to estimate fractal dimensions" *Pattern Recognition,* vol. 38, pp. 1791–1798, 2005.

[26] J. Li, Q. Du, and C. Sun, "An improved box-counting method for image fractal dimension estimation", *Pattern Recognition*, vol. 42, pp. 2460 – 2469, 2009.

[27] Wawrzaszek, A., Krupiński M., Aleksandrowicz S., and W. Drzewiecki, "Multifractal formalism in satellite image analysis" (in Polish with English summary) *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, to be published.

[28] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets." *Physical Review A*, vol. 33(2), pp. 1141-1151, 1986.

[29] J. Levy Vehel, and P. Mignot, "Multifractal segmentation of images" *Fractals*, vol. 2(3), pp. 371–378, 1994.

[30] H. Hentschel, and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractor" *Physica D,* vol. 8, pp. 435-444, 1983.

[31] E. Ott, *Chaos in Dynamical Systems,* Cambridge: Cambridge Univ. Press, 1993.

[32] K. Koperski, G. Marchisio, C. Tusk, and S. Aksoy, Interactive models for semantic labeling of satellite images, In *Proc. International Symposium on Optical Science and Technology SPIE's 47th Annual Meeting,* Seattle, July 2002.

[33] C. Tusk, K. Koperski, S. Aksoy, and G. Marchisio, Automated Feature Selection through Relevance Feedback. In *Geoscience and Remote Sensing Symposium, IGARSS '03, Proceedings 2003 IEEE International,* vol. 6, pp. 3691–3693, 2003.

[34] J. R. Quinlan, *C4.5: Programms for Machine Learning,* Morgan Kaufmann, CA: San Mateo, 1993.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 51–54

On the Computer Certification of Fuzzy Numbers

Adam Grabowski
Institute of Informatics, University of Białystok
Akademicka 2, 15-267 Białystok, Poland
Email: adam@math.uwb.edu.pl

***Abstract*—The formalization of fuzzy sets in terms of corresponding membership functions is already available in machine-verified mathematical knowledge base. We show how it can be extended to provide the development of fuzzy numbers fully benefitting from the existing framework. The flexibility which is offered by automated proof-assistants allowed us to overcome some initial difficulties. Although fuzziness stems from the same background as rough set theory, i.e. incomplete or imprecise information, both formal approaches are substantially different.**

I. Introduction

DURING the past decades, mathematics would evolve from the pen-and-paper model in the direction of use of computers. As fuzzy set theory proposed by Zadeh offered new mathematical insight for the real data in the world of uncertain or incomplete information, dealing mainly with those contained in digital archives, it is not surprising that similar methods will be used in order to obtain the properties of objects within the theory itself. The original approach to fuzzy numbers met some criticism and various ways of improvement were offered as yet. But usually computers serve as an assistant offering calculations – why not to benefit from their more artificial intelligence strength? We try to address some issues concerned with the digitization of this specific fragment of fuzzy set theory, representing a path to fuzzy numbers, so it can be considered as a case study in a knowledge management, being a work on fuzzy sets in the same time.

The paper is organized as follows. The next section is devoted specifically to the situation in the area of computer-checked formalization of mathematics and contains a brief primer to formal fuzzy sets; in the third we gave an example of the proof to show how it looks like; fourth is devoted to the connection of our work with classical and the rough set theory. The other two sections explain specific issues we met during our work while the final brings some concluding remarks and the plans for future work.

II. A Formal Primer of Fuzzy Sets

We were surprised that within the rough set theory the notion of a rough set is not formally chosen as unique. On the one hand, it is a class of abstraction with respect to the rough equality, on the other – the pair of approximation operators. As both theories have much in common, we expected the same from fuzzy sets. But – the membership function itself can be just treated as a fuzzy set. Obviously, there is something unclear with the domain vs. support of a function (as what we call 'fuzzy sets' in fact is a fuzzy subset), but it is not that dangerous. As the author developed the formalization of rough sets, he could make the decision of how much of the existing apparatus should be used also in this case. Eventually all relational structure framework [3] was dropped as completely useless here. We could take the Cartesian product of the original set and the corresponding function, but it is enough to deal only with the latter one.

"Computer certification" is a relatively new term describing the process of the formalization via rewriting the text in a specific manner, usually in a rigorous language. Now this idea, although rather old (taking Peano, Whitehead and Russell as protagonists), gradually obtains a new life. As the tools evolved, the new paradigm was established: computers can potentially serve as a kind of oracle to check if the text is really correct. And then, the formalization is not *l'art pour l'art,* but it extends perspectives of knowledge reusing. The problem with computer-driven formalization is that it draws the attention of researchers somewhere at the intersection of mathematics and computer science, and if the complexity of the tools will be too high, only software engineers will be attracted and all the usefulness for an ordinary mathematician will be lost. But here, at this border, where there are the origins of MKM – Mathematical Knowledge Management, the place of fuzzy sets can be also. To give more or less formal definition, according to Wiedijk [9], *the formalization* can be seen presently as "the translation into a formal (i.e. rigorous) language so computers check this for correctness."

In this era of digital information anyone is free to choose his own way; to quote V. Voevodsky, Fields Medal winner's words: "Eventually I became convinced that the most interesting and important directions in current mathematics are the ones related to the transition into a new era which will be characterized by the widespread use of automated tools for proof construction and verification". If we take into account famous Four Colour Theorem, automated tools can really enable making some significant part of proofs, so hard to discuss with this opinion.

Among many available systems which serve as a proof-assistant we have chosen Mizar. The Mizar system [4] consists of three parts – the formal language, the software, and the database. The latter, called Mizar Mathematical Library (MML for short) established in 1989 is considered one of the largest repositories of computer checked mathematical knowledge. The basic item in the MML is called a Mizar article. It reflects roughly a structure of an ordinary paper, being considered at two main layers – the declarative one, where definitions and

978-83-60810-53-8/$25.00

theorems are stated and the other one – proofs. Naturally, although the latter is the larger, the earlier needs some additional care.

As lattice theory and functional analysis are the most developed disciplines within the MML, further codification of fuzzy numbers, including their lattice-theoretic flavour, looks very promising. As a by-product, apart of readability of the Mizar language, also presentation of the source which is accessible by ordinary mathematicians and pure HTML form with clickable links to notions and theorems are available after the acceptance of the development into the library. As far as we know, this is the first attempt to formalize fuzzy sets in such extent using any popular computerized proof assistant.

Recall that a fuzzy set A over a universe X is a set defined as

$$A = \{(x, \mu_A) : x \in X\},$$

where $\mu_A \in [0, 1]$ is membership degree of x in A. Because the notions in the MML make a natural hierarchy (as the base set theory is Tarski-Grothendieck, which is close to ordinary Zermelo-Fraenkel axiomatics, accepted by most mathematicians): functions $\rightarrow$ relations = subsets of Cartesian product $\rightarrow$ sets, so it is a relation. Zadeh's approach assumes furthermore that μ_A is a function, extending a characteristic function χ_A. So, for arbitrary point x of the set A, the pair (x, μ_A) can be replaced just by the value of the membership function $\mu_A(x)$, which is in fact, formally speaking, the pair under consideration. Then all operations can be viewed as operations on functions, which appeared to be pretty natural in the set-theoretic background taken in the MML as the base. All basic formalized definitions and theorems can be tracked under the address http://mizar.org.

```
definition let C be non empty set;
  mode Membership_Func of C is
    [.0,1.]-valued Function of C,REAL;
end;
```

The aforementioned definition introduces membership function just as a function from given non-empty set into a subset of the set of all real numbers, and the values belong to the unit interval. Of course, `Membership_Func` is not uniquely determined for C – the keyword `mode` starts the shorthand for a type in Mizar, that is, in fact C variable can be read from the corresponding function rather than vice versa.

```
definition let C be non empty set;
  mode FuzzySet of C is Membership_Func of C;
end;
```

We collected translations of selected formalized notions in Table I. As we can read from this table, there are standard operations of fuzzy sets available, usually taken componentwise (note that `F.x` stands for the value of the function F on an argument x). Note that the Mizar repository extensively uses a difference between functions and partial functions; (`Function of X, Y` and `PartFunc of X, Y` in Mizar formalism); because in case of partial functions only the inclusion of the domain in the set X is required, hence the earlier type expands to the latter automatically.

TABLE I
FORMALIZED NOTIONS AND THEIR FORMAL TRANSLATIONS

The notion	Formal counterpart
the membership function	`Membership_Func of C`
fuzzy set	`FuzzySet of C`
$\chi_A(x)$	`chi(A,X).x`
α-set	`alpha-set C`
supp C	`support C`
$F \cap G$	`min (F,G)`
$F \cup G$	`max (F,G)`
cF	`1_minus F`

III. AN ILLUSTRATIVE EXAMPLE OF THE PROOF

In this section we focus on the example of formalized theorem about level sets. Before we start, we explain some plain ASCII symbols which will be used as all Mizar articles have the limitation of using only this narrow set of codes (but automated translation enables to use ordinary mathematical notations, usually based on LaTeX). `>=` stands for $\geq$, `c=` is set-theoretic inclusion, `"` means counterimage or the converse of the relation (including function); `in` stands for $\in$. Of course, `[.a,b.]` is a interval of real numbers a and b. `dom` denotes a domain of a function [4]. It appeared to be pretty feasible, because we could formalize natural properties in a rather compact way, as shown below [2]:

```
definition let C be non empty set;
           let F be FuzzySet of C;
           let a be Real;
  func a-cut F -> Subset of C equals
    { x where x is Element of C : F.x >= a };
end;
```

One can easily notice near one-to-one correspondence with the well-known definition of α-cuts (or level sets):

$$A_\alpha = \{x \in X : \mu_A(x) \geq \alpha\}$$

As it is the counterimage of the interval $[\alpha, 1]$, we can prove the following theorem:

```
theorem AlphaCut1:
  for F being FuzzySet of C,
      alpha being Real holds
    alpha-cut F = F " ([. alpha, 1 .])
  proof
    let F be FuzzySet of C,
        alpha be Real;
    thus alpha-cut F c= F " ([. alpha, 1 .])
    proof
      let x be element;
      assume x in alpha-cut F; then
      consider y being Element of C such that
A1:   x = y & F.y >= alpha;
      x in C by A1; then
A2:   x in dom F by FUNCT_2:def 1; then
      F.y in [. 0, 1 .] by A1,PARTFUN1:4; then
      0 <= F.y & F.y <= 1 by XXREAL_1:1; then
      F.y in [. alpha, 1 .] by XXREAL_1:1,A1;
      hence thesis by A1,FUNCT_1:def 7,A2;
    end;
    :: the other inclusion omitted
  end;
```

IV. Rough and Fuzzy Formal Approaches

In the usual informal mathematical jargon it is easy to say that e.g., two objects are identical up to the isomorphism, formal language has to deal somehow with it. In the fuzzy set theory this can be noticed at the very beginning – some people treat fuzzy sets as the pair of the set and corresponding membership function. Fuzzy sets are subsets of ordinary sets; as we can take membership function just as χ of ordinary sets, it clearly shows the feasibility of this approach. Of course, it is impossible then, at least without any additional preparing work, to find the common bottom ground for ordinary sets and fuzzy sets; however all sets can be made fuzzy in view of the simple lemma cited below:

```
theorem
  for C being non empty set holds
    chi(C,C) is FuzzySet of C;
```

Although two widely-known views (rough and fuzzy approaches) for incomplete or imprecise information have much in common in principle, there are essential differences between both of them [10]. In fuzzy sets, every element has its own membership measure. In rough approach, the degree of membership is rather calculated from the set as a whole, so it is pretty close to Bayes' probability theory, as we quote this below.

```
definition let A be finite Tolerance_Space;
           let X be Subset of A;
  func MemberFunc (X, A) ->
    Function of the carrier of A, REAL means
  for x being Element of A holds  it.x =
  card (X /\ Class (the InternalRel of A,x)) /
    (card Class (the InternalRel of A,x));
end;
```

Paradoxically, even the notion of the rough set was defined in two ways, as pairs of the lower and the upper approximations of a set (in the sense of Iwiński), and as classes of abstraction with respect to given reflexive, symmetric, and transitive binary relation (original Pawlak's approach). MML reflects both approaches, concentrating on the properties of approximation operators and various types of binary relations generalizing equivalence relations.

V. Some "Easy" Problems

Virtually any mathematician uses a formal language; as engineers have also a higher math course in his curriculum, it shouldn't be a big problem to validate facts formally. But if we claim the proof is correct, some natural questions arise: the correctness with respect to which assumptions? What about foundations? If it comes to machine, what are properties of the checker? The original de Bruijn's dream was to have a small checker with the transparent kernel. Most of contemporary proof assistants are rather far from this requirement. Although a theorem can look easy, formal mathematics can bring some unpredictable problems; it is enough to mention e.g. Kepler's conjecture about the densest sphere packing (a part of Hilbert's 18th problem) or Jordan curve theorem that any simple closed curve cuts the plane into two disjoint areas. Intuitively, they are nearly trivial and understandable virtually for any human being. However, even at the very foundational level of used logic (constructive proofs do not claim the law of excluded middle) we can find some unexpected difficulties. Especially important example in our fuzzy context is the so-called glueing lemma – the proof of a simple fact about pasting some continuous functions together to make e.g. triangular (or trapezoidal) fuzzy set, intuitively trivial, draws some surprising dependencies.

It is rather hard to approximate the real complexity of a proof; one of the most popular measures is the de Bruijn factor, i.e. the ratio between the formal translation of the mathematical paper and the original (usually after packing the source and corresponding LaTeX file). Although it is claimed to be about 4 in the case of the Mizar library, in our case is about six (i.e. formal proofs are six times longer than their informal counterparts). Such relatively high number is caused by technical calculations in the process of glueing continuous functions.

VI. Towards Fuzzy Numbers

As all Mizar types should have non-empty denotation, it would force us to define both triangular and trapezoidal fuzzy sets. The natural definition is usually written as conditional definition of parts of the function. We used intervals `[.a,b.]` and `AffineMaps` to save some work (e.g., affine maps are proven to be continuous, one-to-one, and monotone real maps under underlying assumptions). The operator `+*` glues two functions if their domains are disjoint; if not, then the ordering of glueing counts.

```
definition let a,b,c be Real;
  assume a < b & b < c;
  func TriangularFS (a,b,c) -> FuzzySet of REAL
    equals
    AffineMap (0,0)
   +* (AffineMap (1/(b-a),-a/(b-a)) | [.a,b.])
   +* (AffineMap (-1/(c-b),c/(c-b)) | ].b,c.]);
end;
```

The assumptions on the ordering of real variables a, b, c are unnecessary here and will be removed; we kept this as needed to prove the continuity of this fuzzy set afterwards. It is worth mentioning here that the proof of correctness of the above definition is 40 lines long – surprisingly long comparing to the popular (of course, false) opinion that definitions don't need proofs. Continuity of this triangular fuzzy set needed much more lines in our Mizar script (90 lines in case of one-point glueing).

Remembering that a fuzzy number is a convex, normalized fuzzy set on the real line $\mathbb{R}$, with exactly one $x \in \mathbb{R}$ such that $\mu_A(x) = 1$ and μ_A is at least segmentally continuous, we defined it as the Mizar type:

```
  mode FuzzyNumber is f-convex continuous
    strictly-normalized FuzzySet of REAL;
```

As all types are constructed as radix types with added optional adjectives, the generalization, especially that automated-

driven (by cutting the adjectives in the assumptions), is possible and quite frequently used. Some of the adjectives are a little bit stronger that others, with the quoted below as example:

```
definition let C be non empty set;
           let F be FuzzySet of C;
  attr F is strictly-normalized means :SNDef:
    ex x being Element of C st
    F.x = 1  &  for y being Element of C st
      F.y = 1 holds y = x;
end;
```

Observe that this adjective means that a fuzzy set is also normalized in normal sense. Due to automatic clustering of attributes after *registering* this quite natural and easy property any additional reference won't be needed.

```
registration let C be non empty set;
  cluster strictly-normalized ->
    normalized for FuzzySet of C;
end;
```

In our opinion, we made some significant progress on the certification of fuzzy sets and numbers, but our primary aim was to get the formal net of notions correct and reusable and we hope to benefit from it in our future work.

VII. Conclusion and Further Work

The primary aim of using computers in the process of the formalization was to provide its undoubtful correctness. One can argue however that also careful human review should do the same work. The famous exception is the publication of the proof of the Kepler conjecture by Hales in "Annals of Mathematics"; referees cannot be fully sure of the correctness of computer programs and tedious, extremely long computer-driven calculations. But things are different when it comes to program themselves; hardly readable, looking like computer code, proof of Four Color Theorem is verified formally; it sheds some new light for the verification of program libraries – e.g. there is significant progress made with the computer certification of Java or C libraries or even compilers themselves. But here readability is of minor interest; also proofs and the content itself are rather routine. Once the topic is formalized in the machine-understandable language, automated provers can be applied to obtain new results automatically. Based on computer-certified content, further automatic semantical investigations can be made [5], as, for example, extracting lemmas, annotating technical proofs or investigating direct corollaries, automated translation, and fast unification. Furthermore, MML is a subject of continuous changes called *revisions* which can be the result of software upgrades, generalizations, theory merging, introducing new language constructions etc. Also the original first formal approach for fuzzy sets which is dated back to 2001 [6], was thoroughly revised by the author to improve its reuse (e.g., a fuzzy set was primarily defined as the Cartesian product of the set C and the image of the membership function applied to C).

Computer certification of proofs seems to be an emerging trend and some corresponding issues can be raised. We are assured that there are some visible pros of our approach, as for example, automated removal of repetitions, and also the need of writing a sort of preliminary section vanishes in the Mizar code. The type system enables us to search for possible generalizations (including a kind of reverse mathematics at the very end); the use of automated knowledge discovery tools is much easier due to internal information exchange format, which at the same time offers direct translations for a number of formats (e.g, close to the English-like human-oriented language), not limited to the Mizar source code. There are of course drawbacks we should remember of: first of all, the syntax. The Mizar language, although pretty close to natural language, is still an artificial language. Of course, main problem with the formalization is making proper formal background – lemmas and theorems – which can be really time-consuming, hence the stress on reusability of available knowledge.

We argue that the formalization itself can be very fruitful and creative as long as it extends the horizons of the research and make new results possible. Furthermore, the more the database larger is, the formalization can be more feasible. Even if the formalized content concerning fuzzy sets is not that big as of now (there is only about 9000 lines of Mizar code on fuzzy sets comparing with 2.5 million of lines in the whole MML), the basics are already done, and it can serve both as a good starting point for further development, including rough-fuzzy hybridization, as well as from translated existing content we can try to obtain new results. Regardless of the gains of the availability of the topic to majority of popular proof assistants one can ask a question of assurance of the correctness of the proofs; Urban's [8] tools translating Mizar language into the input of first-order theorem-provers or XML interface providing information exchange between various math-assistants are already in use, so not only proof-checkers other than the Mizar verifier can analyze it, but additionally it can allow for some "dirty work" to be done by computer.

References

[1] D. Dubois, H. Prade, Operations on fuzzy numbers, *International Journal of System Sciences,* 9(6), 613–626, 1978.

[2] A. Grabowski, The formal construction of some special fuzzy sets, to appear in *Formalized Mathematics,* 2013.

[3] A. Grabowski, M. Jastrzębska, Rough set theory from a math-assistant perspective, *Lecture Notes in Artificial Intelligence*, 4585, 152–161, 2007.

[4] A. Grabowski, A. Korniłowicz, A. Naumowicz, Mizar in a nutshell, *Journal of Formalized Reasoning,* 3(2), 153–245, 2010.

[5] A. Grabowski, Ch. Schwarzweller, Towards automatically categorizing mathematical knowledge, M. Ganzha, L. Maciaszek, and M. Paprzycki (Eds.), *FedCSIS 2012 Proceedings,* 63–68, 2012.

[6] T. Mitsuishi, N. Endou, Y. Shidama, The concept of fuzzy set and membership function and basic properties of fuzzy set operation, *Formalized Mathematics*, 9(2), 351–356, 2001.

[7] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data,* Kluwer, Dordrecht, 1991.

[8] J. Urban, G. Sutcliffe, Automated reasoning and presentation support for formalizing mathematics in Mizar, *LNCS,* 6167, 132–146, 2010.

[9] F. Wiedijk, Formal proof – getting started, *Notices of the American Mathematical Society,* 55(11), 1408–1414, 2008.

[10] Y.Y. Yao, A comparative study of fuzzy sets and rough sets, *Information Sciences*, 109(1-4), 227–242, 1998.

[11] L. Zadeh, Fuzzy sets, *Information and Control,* 8(3), 338-353, 1965.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 55–61

Cardiac disorders detection approach based on local transfer function classifier

Ahmed Hamdy[1,*], Nashwa El-Bendary[2,*], Ashraf Khodeir[4,*], Mohamed Mostafa M. Fouad[2,*],
Aboul Ella Hassanien[3,*], Hesham Hefny[1]
[1]Department of Computer Sciences and Information, ISSR, Cairo University, Egypt
[2]Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
[3]Faculty of Computers and Information, Cairo University, Egypt
[4]Faculty of Science, Ain Shams university, Egypt
[*]Scientifc Research Group in Egypt (SRGE)
http://www.egyptscience.net

***Abstract*—Truly, heart is successor to the brain in being the most significant vital organ in the body of a human. Heart, being a magnificent pump, has his performance orchestrated via a group of valves and highly sophisticated neural control. While the kinetics of the heart is accompanied by sound production, sound waves produced, by the heart, are reliable diagnostic tools to check heart activity. Chronologically, several data sets have been put forward to sneak on the heart performance and lead to medical intervention whenever necessary. The heart sounds data set, utilized in this paper, provides researchers with abundance of sound signals that was classified using different classification algorithms; decision tree, rotation forest, random forest are few to mention. This paper proposes an approach based on local transfer function classifier as a new model of neural networks for heart valve diseases detection. In order to achieve this objective, and to increase the efficiency of the predication model, boolean reasoning discretization algorithm is introduced to discretion the heart signal data set, then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. Then, the rough sets dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes. Finally, a local transfer function classifier was employed to evaluate the ability of the selected descriptors for discrimination whether they represent healthy or unhealthy. The experimental results obtained, show that the overall accuracy offered by the employed local transfer function classifier was high compared with other techniques including decision table, rotation forest, random forest, and NBtree.**

***Index Terms*—Cardiac disorders, LTF-C, machine learning, feature selection**

I. Introduction

HEART sounds automated diagnosis in recent years became very important to determine condition of the patient (healthy or unhealthy) and determine type of the disease (valvular disease or not), since heart diseases are identified by sounds produced by the heart [1], [9]. Most of heart valve diseases have an effect on the heart sound of patients [2]. Operation of auscultation of heart sounds by the Stethoscope require a professional person to recognize the sounds then detect whether the subject is patient or not and also can detect the type of the heart disease in patients [3], [5]. Junior physicians can't easy detect type of heart disease from the it sound. Using artificial intelligent tools for remote classification of heart sound signal is a useful technique to avoid the needing to the experience physician and expensive equipments such as Echocardiography (ECG), Magnetic Resonance Imaging, (MRI), etc., which used to recognize heart diseases in accurate manner than heart auscultation. In [1] a different classification algorithms using support vector machine (SVM) with different parameters have been applied to find the best classification accuracy. But due to the high number of features, $100 features$, the classification accuracy could be enhanced if irrelevant and noisy features are removed. Discretization or feature selection or both should be prior the classification operation by most of the classification techniques which could lowers the classification performance and accuracy under many conditions [4]. The discretization method should be a supervised manner to satisfy nature of the classification problem, then feature selection method should be applied after the discretization, that demonstrates the dependence on such method for producing an appropriate results, the successful performance of the two pre-processing steps mean successful classification results. Finally a classification technique should be applied to perform the class label, disease type, prediction. Every classification technique has its own strong and weak points [5]. The most important preprocessing step is the feature reduction of the input data set. The data set contain features that are considered as noisy or irrelevant features, these features could have a negative impact on the classification accuracy of the instances, patients. Feature reduction methods are either feature extraction or feature selection method. Feature extraction method applies operation on the original features and extracts a lower number of features that carries the same characteristics. Feature selection methods has two advantages, the first advantage is rank and select the most important features, where if only a subset of features with the highest rank are used in classification, high classification accuracy could be achieved. The extracted heart sound data are three different data sets, each of 100 features where they are slitted into six different parts. The first data set is required to classify

978-83-60810-53-8/$25.00

whether the heart of the patients are normal or not. The second and third data set are required for the detection of the heart valve disease. The heart valve diseases under investigation in this paper are aortic stenosis *AS*, aortic regurgitation *AR*, mitral stenosis *MS* and mitral regurgitation *MR*. This disease classification is performed in two steps where the first step is applied on the second data set for determining the type of the systolic murmur which means *AS* or *MR*, and the second step is applied on the third data set of a diastolic murmur diseases which means *AR* or *MS*. The second advantage of feature selection method is to determine which stage of the heart sound could have the greatest indication to heart valve disease in the case of each murmur type. The four stages of a heart sound are the first heart signal *S*1, the systolic period, the second heart signal and the diastolic period [1].

This paper proposes an approach based on local transfer function classifier as a new model of neural networks for heart valve diseases detection. In order to achieve a good detection, and to increase the efficiency of the predication model, boolean reasoning discretization algorithm is introduced to discretion the heart signal data set, then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. Then, the rough sets dependency rules are generated directly from all generated reducts. Rough confusion matrix is used to evaluate the performance of the predicted reducts and classes. Finally, a local transfer function classifier was employed to evaluate the ability of the selected descriptors for discrimination whether they represent healthy or unhealthy.

The rest of this paper is structured as follows: Section II gives a brief introduction of the heart signals data collection and its characteristics. Section III shows an overview of rough set approach to features selection and reduction methods. The proposed approach is given in section IV. The experimental results and conclusions are presented in Section V and VI respectively.

II. Heart sound signals Data Set and its features

A lot of researches have been applied on heart sound for the detection of heart valve disease. Features are extracted from the heart sound signal into a data set that is composed of 100 features. Then, a classification algorithm is applied on such data set for detection of heart valve disease. Features are extracted in three phases, segmentation [6] [7], transformation and extraction. These extracted features represent the four stages of a heart signal which are *S*1 signal, systolic period, *S*2 signal and diastolic period as shown in figure 1. These features are divided into six groups as follows: (1) **att0:att3** are the standard deviation of all heart sounds, *S*1, *S*2 and average heart rate; (2) **att4:att11** represents signal *S*1; (3) **att12:att35** represents the systolic period; (4) **att36:att43** represents signal *S*2; (5) **att44:att91** represents the diastolic period, and (6) **att92:att99** the four stages of a heart signals are passed from four band-pass frequency filters. The energy of each output is calculated to form these last 8 features.

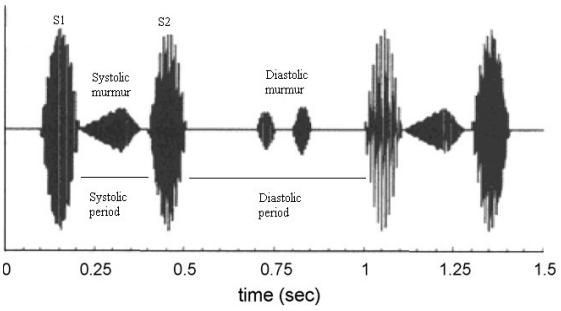


Fig. 1. Heart signal: systolic period and diastolic period [1]

III. Preliminaries

This section provides a brief explanation of the basic framework of rough sets and local transfer function neural network classifier, along with some of the key basic concepts.

A. Rough sets

Rough set theory [17], [16], [15], [14] is a fairly new intelligent technique for managing uncertainty that has can be used for the discovery of data dependencies, evaluation of the importance of attributes, discovery of patterns in data, reduction of attributes, and the extraction of rules from databases. Such rules have the potential to reveal new patterns in the data and can also collectively function as a classifier for unseen data sets. Unlike other computational intelligence techniques, rough set analysis requires no external parameters and uses only the information present in the given data. One of the interesting features of rough sets theory is that it can tell whether the data is complete or not based on the data itself. If the data is incomplete, it suggests more information about the objects to be collected in order to build a good classification model. On the other hand, if the data is complete, rough sets can determine the minimum data needed for classification. This property of rough sets is important for applications where domain knowledge is limited or data collection is very expensive/laborious because it makes sure the data collected is just good enough to build a good classification model without sacrificing the accuracy of the classification model or wasting time and effort to gather extra information about the objects [17], [16], [15], [14].

In rough set theory, data is collected in a table, called a decision table (DT). Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, we assume that the set of examples with a class label to indicate the class to which each example belongs are given. We call the class label the decision attributes, and the rest of the attributes the condition attributes. Rough sets theory defines three regions based on the equivalent classes induced by the attribute values: *lower approximation*, *upper approximation* and *boundary*. Lower approximation contains all the objects, which are classified surely based on the data collected, and upper approximation contains all the objects which can be classified probably, while the boundary is the difference between the

upper approximation and the lower approximation. So, we can define a rough set as any set defined through its lower and upper approximations. On the other hand, indiscernibility notion is fundamental to rough sets theory. Informally, two objects in a decision table are indiscernible if one cannot distinguish between them on the basis of a given set of attributes. Hence, indiscernibility is a function of the set of attributes under consideration. For each set of attributes we can thus define a binary indiscernibility relation, which is a collection of pairs of objects that are indiscernible to each other. An indiscernibility relation partitions the set of cases or objects into a number of equivalence classes. An equivalence class of a particular object is simply the collection of objects that are indiscernible to the object in question.

B. Local transfer function neural network classifier

A new model based on artificial neural network, called Local transfer function classifier produces encouraging results for many data sets, it is virtually the same architecture as Radial Basis Function Neural Network (RBFNN), its used in supervised learning [13].

Let the training set be composed of N pairs of the form: $(X^{(i)}, c^{(i)})$, where $X^{(i)} = \left[x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)}\right]$ is the i-th input pattern belonging to the $c^{(i)}$-th class $(c^{(i)} = 1, 2, ..., k)$. Vectors $X^{(i)}$ can be treated as points in the n-dimensional space X. Close neighborhood of the point $X^{(i)}$ should belong to the same class as $X^{(i)}$, therefore the space X can be divided into finite number of *decision regions*-areas of the same value of classification.

IV. THE PROPOSED HEART VALVE DISEASES ANALYSIS

One way to construct a simple model computed from data, easier to understand and having good predictive power, is to create a set of minimal number of rules. Some condition values may be unnecessary in a decision rule produced directly from the data set. Such values can then be eliminated to create a more comprehensible minimal rule preserving essential information. The proposed heart valve diseases detection approach is comprised of the following three fundamental building phases: (1) Pre-processing including a Discretiztion of the attributes; (2) Generate the reducts with minimal number of attributes along with significant of the attributes; (3) Rule generation for the classification: generate a list of rules, compute the overall accuracy of the generated rules; this phase utilizes the rules generated from the previous phase to predict the classification accuracy. These three phases are described in detail in the following section along with the steps involved and the characteristic features for each process. Fig. 2 illustrates the general architecture of the proposed of heart valve disease analysis.

A. Pre-processing phase: Rough Discretization process

When dealing with attributes in concept image classification, it is obvious that they may have varying degree of importance in the problem being considered, importance can be pre-assumed using auxiliary knowledge about the problem,

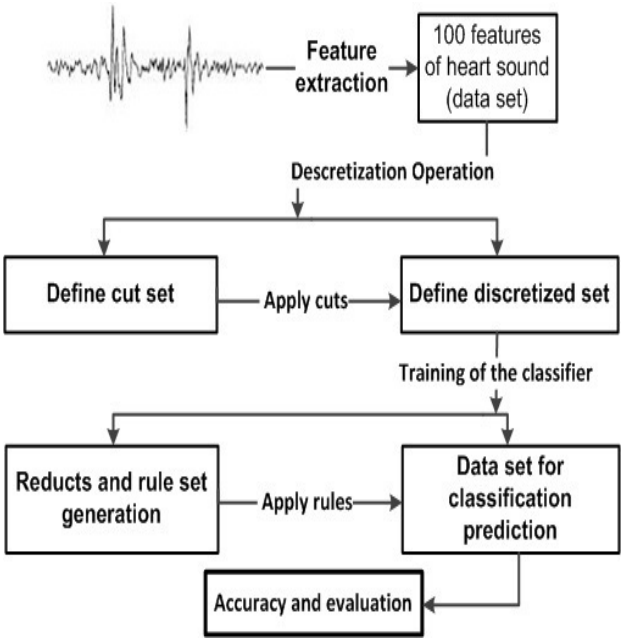


Fig. 2. The general architecture of the proposed of heart valve disease analysis

properly chosen weights. However, in the case of using the rough set approach to concept classification, it avoids any additional information aside from what is included in the information table itself. Basically, the rough set approach tries to determine from the data available in the information table whether all the attributes are of the same strength and, if not, how they differ in respect of the classifier power. Therefore, some strategies for discretization of real valued features must be used when we need to apply learning strategies for data classification (*e.g.*, equal width and equal frequency intervals). It has been shown that the quality of learning algorithm is dependent on this strategy, which has been used for real-valued data discritization [10].

Many classification algorithms such as rough set theory, require that training data contain only discretized feature values. Otherwise, too many equivalent classes will be produced and the algorithms will be over sensitive to noise. To use such an algorithm when there are numeric-valued features, all numeric values must first be converted into discrete values - a process called discretization[11]. This process is performed by dividing the values of a continuous attributes into a small number of intervals, where each interval is mapped to a discrete categorical, nominal, symbolic symbol. Discretization can significantly influence the effectiveness of a classification algorithm.

Medical data sets contains continues and discrete valued data in real world data set. The discretization process divides the attributes value into intervals[12]. The discretization based on RS and Boolean Reasoning (RSBR) shows the best results in the case of heart valve disease data set. In the discretization of a decision table $S = (U, A \bigcap \{d\})$, where U is a non-empty

finite set of objects and A is a non-empty finite set of attributes. And $V_a = [x_a, x_a)$ is an interval of real values x_a, w_a in attribute a. The required is to a partition P_a of V_a for any $a \in A$. Any partition of V_a is defined by a sequence of the so-called cuts $x_1 < x_2 < .. < x_k$ from V_a. The main steps of the RSBR discretization algorithm are provided in algorithm 1.

Algorithm 1 RSBR discretization algorithm

Input: Information system table (S) with real valued attribute A_{ij} and n is the number of intervales for each attribute.
Output: Information table (ST) with discretized real valued attribute

1: **for** $A_{ij} \in S$ **do**
2: Define a set of boolean variables as follows:

$$B = \{\sum_{i=1}^{n} C_{ai}, \sum_{i=1}^{n} C_{bi} \sum_{i=1}^{n} C_{ci}, ..., \sum_{i=1}^{n} C_{ni}\} \quad (1)$$

3: **end for**
Where $\sum_{i=1}^{n} C_{ai}$ correspond to a set of interval defined on the variables of attributes a
4: Create a new information table S_{new} by using the set of intervals C_{ai}
5: Find the minimal subset of C_{ai} that discerns all the objects in the decision class D using the following formula:

$$\Upsilon^u = \wedge\{\Phi(i,j) : d(x_i \neq d(x_j))\} \quad (2)$$

Where $\Phi(i,j)$ is the number of minimal cuts that must be used to discern two different instances x_i and x_j in the information table.

B. Reducts with minimal number of attributes process

Reduct is an important concept in rough sets theory and data reduction is a main application of rough set theory in pattern recognition and data mining. As it has been proven that finding the minimal reduct of an information system is a NP hard problem [10].

The computation of the reducts from a decision table is a way of selecting relevant features [18]. It is a global method in the sense that the resultant reducts represent the minimal sets of features which are necessary to maintain the same classification accuracy given by the original and complete set of attributes. A straight manner for selecting relevant features is to assign a measure of relevance to each attribute and choose the attributes with higher values. Based on the reduct system, we generate the list of rules that will be used for building the classifier model for the new objects. In decision tables, there often exist conditional attributes that do not provide any additional information about the objects. So, we should remove those attributes since it reduces complexity and cost of decision process [18]. A decision table may have more than one reduct. Anyone of them can be used to replace the original table. Finding all the reducts from a decision table is NP-complete. Fortunately, in applications, it is usually not necessary to find all of them. Few of them are sufficient. A natural question is, which reducts are the best. The selection depends on the optimality criterion associated with the attribute. If it is possible to assign a cost function to attributes, then the selection can be naturally based on the combined minimum cost criteria. In the absence of an attribute cost function, the only source of information to select the reduct is the content of the table.

We present a reduct algorithm based on the entropy information measure introduced in [18]. Algorithm-2 shows the main steps of the reduct algorithm.

Algorithm 2 Reduct-based on entropy algorithm

Input: Rough Sets Decision System (RSDS)
Output: One reduct of RSDS

1: $\forall a \in A$ compute the equivalence relation
2: $\phi \leftarrow reduct$;
3: **for** $a_i \in A - reduct$ **do**
4: Compute $H_i = H(a_i \mid reduct)$ {Where H_i is the information quantity of the attribute set, R is a equivalence relation matrix}

$$H = -\frac{1}{n}\sum_{i=1}^{n} log\lambda_i \quad (3)$$

$$\lambda_i = \frac{|[x_i]R|}{n} \quad (4)$$

5: **end for**
6: Compute the significance of attribute a (SIG) in attribute set A using the following equations:

$$SIG(a,A) = H(A) - H(A-a) \quad (5)$$

$$H(a|reduct) = \max(SIG(a_i, reduct)) \quad (6)$$

7: Select attribute which satisfies Equation(20)
8: **if** $H(a|reduct) > 0$ **then**
9: $reduct \cup a \longrightarrow reduct$
10: **end if**
11: Go to Step 3

C. Rule generation for the classification process

The generated reducts are used to generate decision rules. The decision rule, at its left side, is a combination of values of attributes such that the set of (almost) all objects matching this combination have the decision value given at the rule's rough side. The rule derived from reducts can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data. The main steps of the rule generation and classification algorithm are provided in Algorithm 3 (cf. [18]).

V. Experimental results and discussions

A. The heart sound signals data set characteristics

Cardiac disorders of heart diseases in the proposed approach were applied on three different data set of cardiac disorders

Algorithm 3 Rule generation for the classification

Input: reduct sets $R_{final} = \{r_1 \cup r_2 \cup .... \cup r_n\}$
Output: Set of rules

```
for each reduct r do
    for each correspondence object x do
        Contract the decision rule (c1 = v1 ∧ c2 = v2 ∧ .... ∧ cn = vn) ⟶ d = u
        Scan the reduct r over an object x
        Construct (ci, 1 ≤ i ≤ n)
        for every c ∈ C do
            Assign the value v to the correspondence attribute a
        end for
        Construct a decision attribute d
        Assign the value u to the correspondence decision attribute d
    end for
end for
```

with different number of instances in every class. The first data set is healthy and unhealthy persons "$HS-H-U$" contains 70 instances, where 38 instances represent healthy persons and the other 32 instances represents unhealthy patients. The second data set represents 84 instances systolic diseases such that 41 instances aortic stenosis and 43 instances mitral regurgitation "$HS-AS-MR$". Finally the third data "$HS-AR-MS$" set represents 76 instances diastolic diseases, it consists of 38 instances aortic regurgitation and 38 instances mitral stenosis.

TABLE I
MINIMAL REDUCT SETS OF THE THREE DATA SETS

Data type	Reduct sets
$HS-H-U$	att0, att2, att33, att87, att93, att96
$HS-AS-MR$	att0, att2, att31, att87, att89, att99
$HS-AR-MS$	att1, att4, att12, att35, att37

Table I shows the generated reducts that contains minimal number of attributes. While Table II, III, and IV show the generated rules set of the three data sets of heart valve disease signals. As an explanation for some of rules, should be first clearing some terms, "att3" is feature of standard deviation of all heart sounds, S1, S2 and average heart rate, "att36" is feature of signal S2, "att92 and att96" are represent the four stage of the heart signal. If att36= 0.15405 and att92=0.06865 and att3=0.32355 then this patient is normal, about 16 cases match this rule. If att96=0.04485 and att3=0.32355 and att92=0.06865 then this patient is up normal, about 8 cases match this rule.

B. Results analysis and discussion

In this approach, local transfer function neural network classifier (LTF-C) has been applied on three types of heart valve murmurs data sets. It shows the highest classification results as shown by figures 3,4 and 5. The best classification in the three data sets is achieved by LTF-C classifier, comes

TABLE II
GENERATED RULES FOR THE $HS-H-U$ DATA SET

Matches	Decision rules	Class
16	(att36="(0.15405,Inf)") & (att92="(-Inf,0.06865)") & (att3="(0.32355,Inf)")	1
6	(att96="(-Inf,0.04485)") & (att0="(-Inf,0.05875)") & (att3="(-Inf,0.32355)")	1
3	(att96="(-Inf,0.04485)") & (att92="(0.06865,Inf)") & (att36="(0.15405,Inf)") & (att0="(0.05875,Inf)") & (att3="(-Inf,0.32355)")	1
2	(att3="(0.32355,Inf)") & (att36="(-Inf,0.15405)") & (att0="(-Inf,0.05875)") & (att96="(-Inf,0.04485)") & (att92="(-Inf,0.06865)")	1
2	(att0="(0.05875,Inf)") & (att36="(-Inf,0.15405)") & (att92="(-Inf,0.06865)") & (att3="(-Inf,0.32355)")	1
3	(att3="(0.32355,Inf)") & (att92="(0.06865,Inf)") & (att94="(0.2325,Inf)") & (att96="(-Inf,0.04485)")	1
9	(att96="(0.04485,Inf)") & (att36="(0.15405,Inf)") & (att0="(0.05875,Inf)") & (att92="(0.06865,Inf)")	2
3	(att96="(0.04485,Inf)") & (att3="(-Inf,0.32355)") & (att0="(-Inf,0.05875)") & (att36="(0.15405,Inf)") & (att92="(-Inf,0.06865)") & (att94="(-Inf,0.2325)")	2
8	(att96="(0.04485,Inf)") & (att3="(-Inf,0.32355)") & (att92="(0.06865,Inf)")	2
2	(att94="(-Inf,0.2325)") & (att3="(0.32355,Inf)") & (att36="(-Inf,0.15405)") & (att0="(-Inf,0.05875)") & (att92="(-Inf,0.06865)") & (att96="(0.04485,Inf)")	2
3	(att0="(0.05875,Inf)") & (att96="(0.04485,Inf)") & (att92="(-Inf,0.06865)") & (att3="(-Inf,0.32355)") & (att36="(0.15405,Inf)")	2
1	(att3="(0.32355,Inf)") & (att0="(0.05875,Inf)") & (att36="(-Inf,0.15405)") & (att96="(0.04485,Inf)") & (att92="(-Inf,0.06865)") & (att94="(-Inf,0.2325)")	2

TABLE III
GENERATED RULES FOR THE $HS-AR-MS$ DATA SET

Matches	Decision rules	Class
28	(att12="(-Inf,5.0E-5)")&(att35="(-Inf,0.19665)") &(att4="(0.22775,Inf)")	2
22	(att35="(0.19665,Inf)")&(att37="(0.01215,Inf)")	1
15	(att1="(-Inf,0.23755)")&(att4="(-Inf,0.22775)") &(att37="(0.01215,Inf)")&(att12="(5.0E-5,Inf)")	1
7	(att12="(5.0E-5,Inf)")&(att37="(-Inf,0.01215)")	1
5	(att37="(-Inf,0.01215)")&(att1="(-Inf,0.23755)") &(att4="(-Inf,0.22775)")&(att35="(-Inf,0.19665)")	1
4	(att12="(-Inf,5.0E-5)")&(att4="(-Inf,0.22775)") &(att35="(-Inf,0.19665)")&(att37="(0.01215,Inf)")	2
2	(att12="(-Inf,5.0E-5)")&(att37="(-Inf,0.01215)") &(att4="(-Inf,0.22775)")&(att1="(0.23755,Inf)")	2

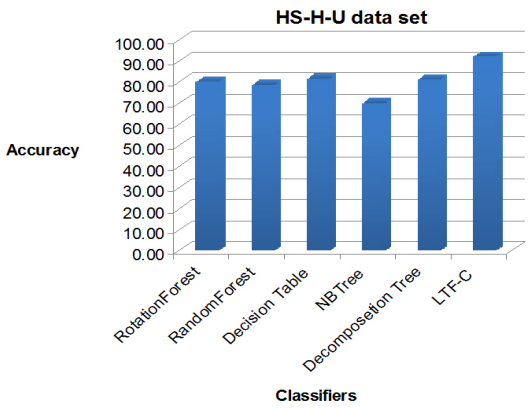


Fig. 3. Classification results for Healthy Unhealthy Data Set

TABLE IV
GENERATED RULES FOR THE $HS-AS-MR$ DATA SET

Matches	Decision rules	Class
23	(att31=”(0.01485,Inf)”)&(att0=”(0.0279,Inf)”) &(att99=”(-Inf,67.9889)”)	2
20	(att2=”(-Inf,0.58995)”)&(att87=”(-Inf,0.00105)”) &(att31=”(-Inf,0.01485)”)	1
8	(att87=”(-Inf,0.00105)”)&(att99=”(67.9889,Inf)”) &(att31=”(0.01485,Inf)”)	1
6	(att0=”(0.0279,Inf)”)&(att31=”(-Inf,0.01485)”) &(att87=”(0.00105,Inf)”)&(att89=”(0.00895,Inf)”)	1
6	(att99=”(67.9889,Inf)”)&(att87=”(0.00105,Inf)”) &(att0=”(0.0279,Inf)”)&(att31=”(0.01485,Inf)”) &(att89=”(0.00895,Inf)”)	2
5	(att31=”(0.01485,Inf)”)&(att2=”(-Inf,0.58995)”) &(att0=”(-Inf,0.0279)”)&(att87=”(-Inf,0.00105)”) &(att89=”(-Inf,0.00895)”)&(att99=”(-Inf,67.9889)”)	2
4	(att0=”(0.0279,Inf)”)&(att31=”(-Inf,0.01485)”) &(att2=”(0.58995,Inf)”)&(att87=”(-Inf,0.00105)”)	1
4	(att87=”(0.00105,Inf)”)&(att0=”(0.0279,Inf)”) &(att2=”(0.58995,Inf)”)&(att31=”(-Inf,0.01485)”) &(att89=”(-Inf,0.00895)”)&(att99=”(-Inf,67.9889)”)	2

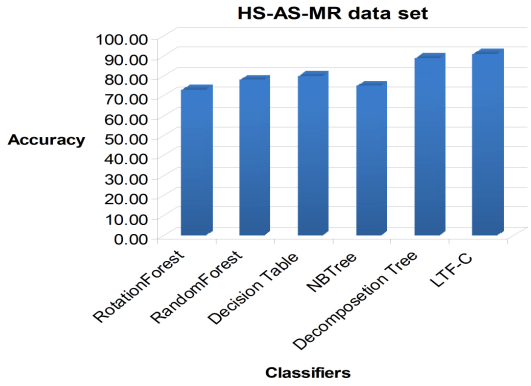


Fig. 4. Classification results for AS-MR Data Set

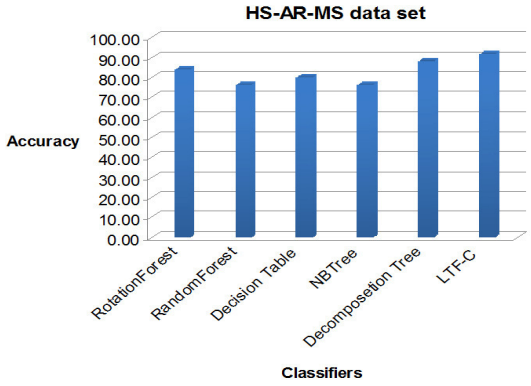


Fig. 5. Classification results for AR-MS Data Set

after it in the order Decision Table regarding the healthy unhealthy data set “$HS-H-U$” and Decomposition Tree classifier regarding both other data sets the systolic data set “$HS-AS-MR$” and the diastolic data set “$HS-AR-MS$”. The following table V shows the collection results of the three data sets.

Table V illustrates the overall local transfer function neural network classifier accuracy in terms of sensitivity and specificity compared with decision table, rotation forest, random forest, and NBtree. Empirical results reveal that the proposed local transfer function neural network approach performs better than the other classifiers.

TABLE V
ACCURACY RESULTS FOR THREE HEART VALVE DISEASES DATA SETS

Classifier	$HS-H-U$	$HS-AS-MR$	$HS-AR-MS$
LTF-C	**92.00**	**91.00**	**91.70**
Decomposetion Tree	81.00	89.00	88.00
NBTree	70.00	75.00	76.00
Decision Table	81.43	80.00	80.00
RandomForest	78.57	78.00	76.00
RotationForest	80.00	73.00	84.00

VI. CONCLUSIONS AND FUTURE WORKS

The heart sounds data set, utilized in this paper, provides researchers with abundance of sound signals that was classified using different classification algorithms; decision tree, rotation forest, random forest are few to mention. Such algorithms were of disputable performance if compared with the classification algorithm adopted in this paper, i.e., the ”local transfer function neural network classifier (LTF-C)”. Discretization of analogue heart sounds was a preparatory step to apply LTF-C classification technique. Consequently, discretized data were classified into several domains and Rough Confusion Matrix was used to produce reducts out of them. The purpose of such data manipulation is to reach a state of subjecting features to discernability, so classes of distinct features can fuel proper decision for a cardiologist, to which class of cardiac disorder this patient belongs. Classes were meant to touch upon crucial cardiac diseases and to aid in diagnosis and prognosis as well. The LTF-C achieved a high accuracy classification compared with other machine learning techniques such as Decomposition Tree, NBTree, Decision Table, RandomForest in addition to RotationForest.

REFERENCES

[1] Maglogiannis I, Loukis E, Zafiropoulos E and Stasis A., “Support vectors machine-based identification of heart valve diseases using heart sounds”, Journal of Computer Methods and Programs in Biomedicine, Elsevier, North-Holland, vol. 95, pp. 47–61, (2009).
[2] T. Chen, K. Kuan, L. Celi and G. Clifford, “Intelligent heart sound diagnostics on a cellphone using a hands-free kit”, Proceedings of AAAI Artificial Intelligence for Development (AI-D’10). Stanford University, California, (2010).
[3] J.E. Hebden and J.N. Torry, “Neural network and conventional classifiers to distinguish between first and second heart sounds”, Artificial Intelligence Methods for Biomedical Data Processing IEE Colloquium (Digest), vol 3, pp. 1–6, (1996).
[4] Kavita Das, and Vyas O. P., A suitability study of discretization methods for associative classifiers. International Journal of Computer Applications, vol. 5(10), pp. 0975–8887, (2010).

[5] Hamdy, A., Hefny, H., Hassanien, A.E., Salama, M. A., and Kim, T. (n.d.), "The importance of handling multivariate attributes in the identification of heart valve diseases using heart signals". Proceedings of the the Second International Federated Conference on Computer Science and Information Systems (FedCSIS), pp.75,79, 9-12 Sept. (2012).

[6] H. Liang, S. Lukkarinen and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram", Computers in Cardiology, pp. 105–108. (1997).

[7] H. Liang, S. Lukkarinen and I. Hartimo, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction", Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, USA, vol. 4, pp. 1630–1633, (1997).

[8] D. Kumar, P. Carvalho, M. Antunes, R.P. Paiva and J. Henriques, "Heart murmur classification with feature selection", In proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, vol. 1, pp. 4566–4569, (2010).

[9] Plamena Andreeva, et al., "Data Mining Learning Models and Algorithms for Medical Applications", 18 Conference Systems for Automation of Engineering and Research (SEAR 2004), Varna, BG, pp. 148–152, (2004).

[10] Aboul Ella Hassanien, Mohamed E. Abdelhafez, Hala S. Own: Rough Sets Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic Children Patients. Adv. Fuzzy Systems 2008, (2008).

[11] HS. Nguyen, SH. Nguyen, "Discretization methods in data mining", Rough sets in knowledge discovery, pp. 451-482, (1998).

[12] HS. Nguyen, "Approximate boolean reasoning: Foundations and applications in data mining", Transactions on Rough Sets V, (2006).

[13] M. Wojnarski, "LTF-C: Architecture, Training Algorithm and Applications of New Neural Classifier", Fundamenta Informaticae, Vol.54(1), pp. 89–105. IOS Press, (2003).

[14] Zdzisiaw Pawlak, Rough set approach to knowledge-based decision support. European Journal of Operational Research, vol. 99, pp. 48-57, (1997).

[15] Pawlak Z. (1991) Rough Sets- Theoretical aspect of Reasoning about Data. Kluwer Academic Publishers, 1991.

[16] Pawlak Z., Rough sets. *International J. Comp. Inform. Science*, vol. 11, pp.341-356, (1982).

[17] Pawlak Z., Grzymala-Busse J., Slowinski R., and Ziarko W., Rough Sets. Communications of the ACM, vol.38, no.11, pp.88-95, (1995).

[18] Aboul Ella Hassanien. Intelligence techniques for prostate ultrasound image analysis. Int. J. Hybrid Intell. Syst. 6(3): 155-167, (2009).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 63–67

A Human Inspired Collision Avoidance Strategy for Moving Agents

Pejman Kamkarian
Electrical and Computer Engineering Department,
Southern Illinois University, Carbondale, IL 62901, USA
Email: pejman@siu.edu

Henry Hexmoor
Computer Science Department, Southern Illinois
University, Carbondale, IL 62901, USA
Email: hexmoor@cs.siu.edu

Abstract- This paper presents an approach for controlling collision avoidance among a group of moving multi-agents such that they are not able to communicate with each other and hence, cannot share information. The basics and key features of our collision control algorithm are discussed to include practical examinations. Our approach is based on multi-agent systems and help moving agents to pursue their goals using collision free routes. In terms of validating our solution, we plan to apply into a configuration set of agents located in our experimental space. We also explain our solution algorithm that we have developed, along with the examination that we subjected it to, as well as sketching some of the most important challenges that remain to be addresses in our future researches.

Keywords- Multi-agents; Moving agents; Agents collision control; Intelligent agents; Agents decision making system.

I. INTRODUCTION

In one class of mobile robotics research heralded by [1], sets of rules are used for collision avoidance. Rules correspond to sensing abilities of the mobile robots. Assuming that each robot has abilities to detect collisions and can measure the distance to a colliding robot and its velocity in the forward direction of the avoiding robot, the following two rules are used. In these rules, the robot can know the location of the colliding robot by using sensors for detecting collisions and can judge whether it is approaching or leaving by measuring its velocity. At first, the robot will attempt to apply the rules without communications. If the rules do not apply because of an environmental restriction, the robot stops and starts communicating to determine a way out of the deadlock.

(1) If the colliding robot is located in front and near and it is approaching, then avoid it from the left.

(2) If the colliding robot is located in front and near and it is departing, then stop locomotion for predetermined time duration.

When rules are not applicable, in another class of mobile robotics research, collision avoidance is a negotiated activity between pairs of robots that have potential collisions. During the message exchange of warning and its reply, priorities of each robot are reported to each other. Mobile Robot 1 detects the collision, and after it takes into account the priorities, it determines if it is reasonable for robot 1 to avoid collision due to high priority. If robot 2 detects the collision, reverse processes are executed. If the priority of robot 2 is found to be higher than robot 1 as a result of negotiation, robot 1 sends to robot 2 a declaration to proceed instead of a command to wait for robot 1. Then, robot 2 moves to avoid collision and sends a command to restart to robot1. Alternatively, the coordination algorithm is useful when groups of robots move in opposite directions and while navigating or when a specific region is a target for many robots, in particular [9]. The main goal in the coordination algorithm is that it forces robots to wait while the other robots continue to move to their target, and then allows the remaining robots to move. Consequently, with this coordination, the congestion problem will be decreased, and at the same time, the percentage of reaching robots to their target will be increased. We used a version of this priority scheme in an earlier paper [6].
There have been several attempts to use human behavior as inspiration. Human inspired methodology is summarized in the following three steps [8].

(1) Keep your direction and velocity of motion if there is minor possibility of a collision.

(2) Else, if a major possibility of a collision exists, and there are no ways to manoeuver around it; then, stop to let the other person to continue to move in their direction and velocity,

(3) Else, if a possibility of a collision exists, and there is a way to manoeuver around it; then Change your direction of movement with slightly changing speed to around the other person, and joining back to your original path of motion.

The strategy proposed in this paper is also largely human inspired and it can be applied to robots in crowded environments. We begin by an outline of conditions and premises in section 2. We then examine a magnified scenario of encounters between a pair of individuals in collision sites in section 3. Spiral orbits as a strategy for collision aversion is discussed in section 4. Experimental results are discussed in section 5 followed by conclusions in section 6.

978-83-60810-53-8/$25.00

II. CONDITIONS AND PREMISES

There are several terms and conditions that need to be considered in regards to developing a suitable strategy that is able to avert collisions among multiple moving agents as they move toward their goals located in a site. There are two entities available on our two dimensional space; agents and goals.

Below, we outline the most important features that are directly and indirectly pertinent for elaborating our solution.

(1) Agents can be located at arbitrary position on our feasible two dimensional spaces. There are limitations on their numbers; however, the total number of agents at any cycle of experiment will not exceed the total number of goals. Our experimental space, has the following default criteria: $EC = \{EC_1, EC_2, \dots, EC_i\}$, $EC_i = \{A_i, G_j\}, A_i = \{a_1, a_2, \dots, a_i\}, G_j = \{g_1, g_2, \dots, g_j\}, \{\forall(i,j) \in \mathbb{N} \mid (i \in A, j \in G) : i \geq j\}$, where EC, A_i and G_j present experimental cycles, agents and goal respectively.

(2) There is a one to one correspondence between agent groups and the goal collection, which means that each goal will be assigned to a unique, single agent. As a matter of course, each agent maintains its goal during cycles of experimentation. This is captures in the mapping$f: X \to Y$,
where set X indicates the agent set and Y represents the goal set; $f(A_i) = G_i$.

(3) Each agent starts moving at any time during the experiment. $T = \{\Delta t_{a_1}, \Delta t_{a_2}, \dots, \Delta t_{a_i}\}$; $\forall(\Delta t_{a_{i-1}}, \Delta t_{a_i}) \in T, \Delta t_{a_{i-1}} \neq \Delta t_{a_i}$, where Δt_{a_i} , is the start time for a_i. In other words, there is no common time for each agent to start moving toward its goal within experiment cycles.

(4) All agents are assumed to have the same uniform size and shape as pointed out in many other sources such as [2], [5], [11].

(5) Each agent has its own distinctive communication protocol; however they are all use the same collision control method to handle collision sites, if occurred. As a matter of fact, as opposed to those who employ a method of communications among multiple moving agents for their verifications, such as [2], [5], we assumed that agents do not have any connection to each other and hence are not able to disseminate information among themselves during a period of experiment.

(6) Moving toward goals is not necessarily bounded on a straight path, which means, each agent is able to select a route with any speed based on the situation.

(7) Goals can be located at any arbitrary location in two-dimensional space; however, they cannot be relocated and changed their positions during each cycle of an experiment. Agents know their exact location as well as their goal locations; $a_i = \{c_{(x,y)_i}, g_{(x',y')_i}\}$ where $c_{(x,y)_i}$is the coordinate of the agent and $g_{(x',y')_i}$ is the coordinate of the goal location. Each agent is capable of calculating its location and also its corresponding goal location at any needed time during each experimental cycle. An effectual way to achieve this goal is to equip agents with a vision sensor device which is already demonstrated in former publications such as optical motion planning mapping in [7], [10], or motion sensor themselves that are presented in [4], [3] and [12].

III. COLLISION SITE

Each agent is able to recognize a limit site (i.e., a region) located in front of its vision sensor, which is computed by $vs = \int_0^\theta \int_0^r dS = \int_0^\theta \int_0^r \tilde{r} d\tilde{r} d\tilde{\theta} = \int_0^\theta \frac{1}{2} r^2 \theta$, and based on that, it decides and determines the best possible path toward its goal at any time. In this formula, r and θ denote maximum vision depth and vision range angle for each agent vision sensor, respectively. We define a collision site form, if an agent detects another agent in its vision site range. In such situations, orientations of those agents participating to form collision sites are not taken into account. Figure 1 shows a prototypical collision site.

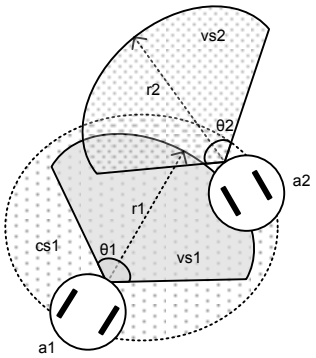


Fig 1. a_1 detects a_2 on its vision site (vs_1), and hence collision site (cs_1) is formed

All agents that are forming such collision sites will use our solution as a potential collision control strategy to enter and exit those sites. The process of handling a collision site consists of two phases. The first phase is forming a distance set $cs_i = \{D_j a_{j(x,y)_j}, D_k a_{k(x,y)_k}, \dots\}$, where $D_j a_{j(x,y)_j}$, indicates distance between a_j and a_i. The second phase is to form the smallest circular area that contains the agent that has the nearest distance from a_i where both agents located on the perimeter of it. For instance, a_i, forms $(x-l)^2 + (y-k)^2 = r^2$, where x, y, l, k, and r, are coordinates for a_i and a_j respectively, if $\forall(a_j, a_k) \in cs_i$, $D_j a_{j(x,y)_j} < D_k a_{k(x,y)_k}$. a_i, then starts moving toward a temporary expansion spiral route computed by $r = pe^{q\theta}$, where e indicates the base of natural logarithms, and p and q are parametric positive real constant values respectively, as polar coordinates, (r, θ), shown as $\overrightarrow{VR_1}$ route, in Figure 2, around a virtual collision site circle formed by a_i and a_j. We present our general path finder solution along with relative formulas in next section.

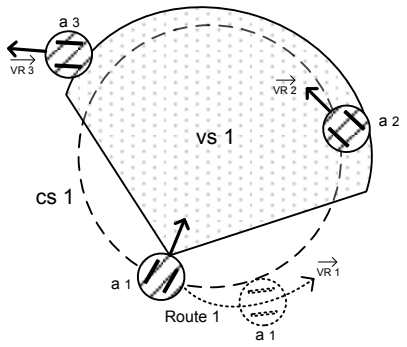


Fig 2. Collision site formed between a_1, a_2, and a_3, once a_1 detected a_2, and a_3 on its vision site. a_1 hence, starts moving on $\overrightarrow{VR_1}$ route

IV. COLLISION AVOIDENCE USING VIRTUAL EXPANSION SPIRAL ORBIT SOLUTION

In this paper, we concentrated on developing a strategy, which is capable of preventing a group of agents to strike each other while moving toward their goals. In order to achieve our objective, we developed an algorithm that each agent in our experimental space uses to determine the safest, short route at each decision making cycle. This consists of two general strategies that will be used in each presumed situation. The path finder algorithm determines a straight line connecting path as the shortest route toward agent goal during the times that agent is not causing or participating in formation of collision sites. The algorithm, however, will determine an expanding spiral path as a temporary route when the agent is still located in the collision site, in terms of preventing potential collisions among moving agents in our two dimensional space. Robots follow those temporary routes until exiting from those collision sites successfully. By analogy, this is a virtual, pivoting dance step between two agents in the collision sites. State diagram in Figure 3 demonstrates general strategies that we used in our route finder algorithm solution.

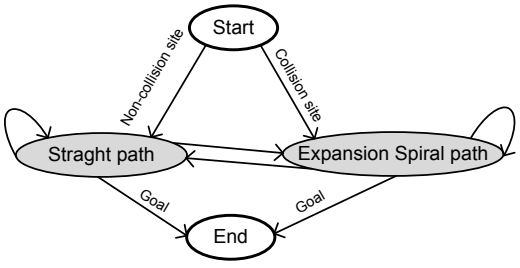


Fig 3. State graph of our path finder solution

Our path finder algorithm including collision avoidance solution inside is shown in Algorithm 1.

Algorithm 1. Path finder algorithm

1. Initialize your current location.
2. Initialize your assigned goal location.
3. Recognize and analyze around yourself by using your vision sensor to determine your current status.
4. If distance from your goal is 0, then end. ($d_{a_i} = 0$).
5. If collision site detected, then jump to step 9.
6. Adjust your direction toward your goal.
7. Move toward goal for one scale.
8. Jump to step 3.
9. Form a group of all agents that are in your vision site, virtually.
10. Calculate distance to each agent from your current coordinates.
11. Form a circle virtually crossing between you and the nearest agent in your vision site.
12. Move toward expansion spiral path around nearest virtual circle, formed in previous step for one scale.
13. Jump to step 3.

We assumed two sets in our two dimension experimental space; $A = \{a_1, a_2, \dots, a_i\}$, and $G = \{g_1, g_2, \dots, g_i\}$, where A, and G are agents and goals set, respectively. Agents know their assigned goal initially, $\forall a_i \in A, \exists g_i \in G, (a_i, g_i) \in I$, where I indicates a one to one correspondence pairs between agent set and goal set. Initially, each agent knows its exact coordinates along with its assigned goal location at the beginning of experimental cycle. $T_0 = \{a_{1(a_{1(x,y)_1}, g_{1(x',y')_1})}, a_{2(a_{2(x,y)_2}, g_{2(x',y')_2})}, \dots, a_{i(a_{i(x,y)_i}, g_{i(x',y')_i})}\}$ Agents start moving toward their goals at random times. Each agent a_i, at the beginning of the process of movement; Δt_{a_i}, a_i evaluates the environment around by analysing data obtained from its vision sensor. This strategy helps them to determine their situation and hence adjust their path accordingly. For instance, a_i is able to analyze the surface of $vs_i = \int_0^{\theta_i} \frac{1}{2} {r_i}^2 \theta$, as its vision site, captured from its vision sensor at any arbitrary time during movement toward its goal. $\bigcap_{p=1}^{m} vs_p = \emptyset \implies cs = 0$, indicates that there is no agent participating to form any collision site and hence a_i concludes to move toward its goal through by $y = y_1 + [(y'_1 - y_1) / (x'_1 - x_1)] \cdot (x - x_1)$, where (x_1, y_1) and (x'_1, y'_1) denote the position of a sample agent and goal on two dimension space respectively. $\exists (a_i, a_j) \in A, (vs_i \cap vs_j) \neq \emptyset \implies cs \neq \emptyset$, however, indicates there is at least one collision site for a_i, and hence it should alter its path through a temporal expansion spiral route. Those agents are located in collision site; follow these temporary spiral routes, until exiting from those sites. During the process of moving out from collision sites, agents continue analysing their environment around to detect any new agents into their current collision sites and hence, form a new temporary path based on the agent that maintains the nearest distance from them, in order to adjust their route, as it expressed in the collision site section.

V. EXPERIMENTAL RESULTS

The presented path finder algorithm, explained in the previous section, was implemented and tested with an experimental scenario. In this section, we illustrate and examine our algorithm along with the relative results and analysis. In this experiment, our system is used to plan avoiding collision among a group of 6 agents moving toward

their goals, in our two dimensional experiment space. They are not able to communicate with to one another during the experiment; however, they are all using the same strategy to find and correct their path into their goals. This scenario is formed by arranging agents and goals in space, randomly, with considering assigning the farthest possible goals in terms of distances, to agents. This type of distributing agents and goals on the experimental site, leads to increasing the possibility of facing agents to more into collision sites, and hence, using our collision avoidance strategy as much as it possible. Figure 4 shows the positions of agents and goals on our experimental space.

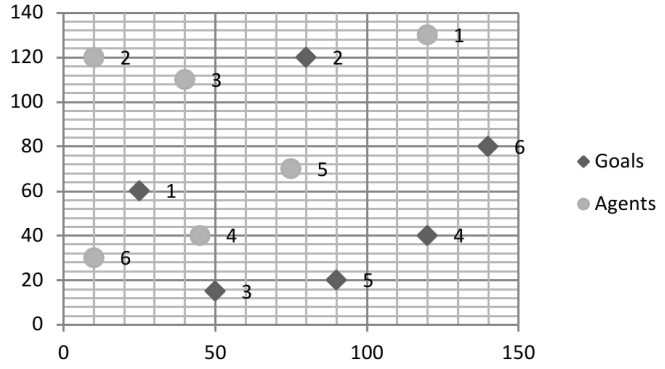


Fig 4. Agents and goals distribution on experimental site

Each agent is assigned a unique number in order to be recognized by other agents during the experiment. In addition, there is a one to one correspondence relationship between agents and goals. In other words, each agent in agent set is assigned the same number goal in goal set. During the cycle of experiment, we collected many key features such as, the times of start moving toward goals by agents, as well as the times of reaching goals, along with the total number of collision sites and virtual circles that each agent is face during the experiment, shown in the following Table 1.

TABLE 1.
RESULTS TABLE OF AGENTS MOVEMENT TOWARD THEIR GOALS

(a_i,g_i)	ΔTa_i	cs_i	vc_i	ΔTg_i
(1,1)	0	7	5	196
(2,2)	2	9	8	314
(3,3)	4	4	2	88
(4,4)	6	8	6	264
(5,5)	8	5	3	165
(6,6)	10	6	5	211

We also collected the times of entering and exiting collision sites for each agent, during the experiment, as shown in the Table 2.

Results depicted in Table 2, show significant differences for ΔT_g , for agents that encountered a larger number of collision sites. In other words, each collision site, based on its situation, and the total number of agents that participate in forming it, can potentially cause a significant delay for agents to reach their goals. This is because, agents that are located in collision sites, change their normal behaviour to choose their paths based on following the shortest possible routes, to a temporary paths which forms based on other agents participating to a same collision site, in order to handle, and hence, exiting from them. We observed no collisions among agents meaning they were able to reach their goals successfully, during the experiment.

TABLE 2.
ENTRANCE AND EXIT TIMES FOR COLLISION SITES

a_i	$\Delta T\,ent_{cs_i}$	$\Delta T\,ext_{cs_i}$
1	{5, 14, 32, 48, 89, 102, 188}	{13, 30, 45, 80, 101, 183, 192}
2	{12, 29, 54, 88, 109, 164, 195, 248, 299}	{26, 50, 87, 103, 162, 193, 245, 296, 312}
3	{22, 56, 64, 80}	{54, 63, 79, 82}
4	{11, 130, 52, 83, 113, 142, 206, 252}	{129, 50, 81, 112, 140, 202, 251, 260}
5	{15, 43, 64, 78, 99, 125, 142, 161}	{40, 62, 77, 98, 120, 139, 155, 161}
6	{22, 45, 84, 99, 128, 197}	{42, 80, 98, 125, 192, 202}

VI. CONCLUSIONS

In this paper, we proposed a solution to prevent collision among a group of moving agents toward their specific goals. Our demonstrated algorithm is able to analyse the information gathered by equipped vision sensors, in order to decide the best possible route, in terms of safety and collision avoidance during the time of attempting to reach to their goals. We assumed our agents are not able to communicate and hence do not share details of their environment among one another. Our solution, thus, is able to help our agents to decide and routing toward their goals independently.

Our approach is able to control collision among moving agents into their goals successfully, however, using it, causes agents to have a significant delay before reaching their goals. These delays, depend of the total number of collision sites that each agent involves during the time of pursuing goals can substantially increase the cost of time. Future works includes optimizing our solution, in terms of minimizing the cost of time needed for agents to handle and exit collision sites.

REFERENCES

[1] H. Asama K. Ozaki, H. Itakura, A. Matsumoto, Y. Ishida, I. Endo,"Collision avoidance among multiple mobile robots based on rules and communication," *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, vol., no., pp.1215-1220 vol.3, 3-5, 1991.

[2] C. M. Clark, S. M. Rock and J. C. Latombe, "Motion planning for mobile robots using dynamic networks", *Proc IEEE Int Conf on Robotics and Automation,* 2003.

[3] F. Expert, S. Viollet and F. Ruffier, "Outdoor field performances of insect-based visual motion sensors", *Journal of Field Robotics,* vol. 28, no. 4, pp. 529–541, 2011.

[4] J. Gaspar, N. Winters and J. Santos-Victor, "Vision-based Navigation and Environmental Representations with an Omni-directional Camera", *IEEE Transactions on Robotics and Automation,* Volume 16 Number 6, pp 890 -898, 2000.

[5] Y. Guo and L. Parker, "A distributed and optimal motion planning approach for multiple mobile robots", *Proc IEEE IntConf on Robotics and Automation,* pp. 2612–2619, 2002.

[6] P. Kamkarian and H. Hexmoor, "A Collision Control Strategy for Multiple Moving Robots", *In Advances in Intelligent Systems and Computing,* Volume 208, pp. 863-871, Springer, 2012.

[7] S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning”, *International Journal of Robotics Research,* Volume 30 Issue 7, pp. 846-894, 2011.

[8] R. Kumar and A. Menon. "Collision Avoidance in a Multi-Robot System by Emulating Human Behavior." *International Conference on Information Systems Analysis and Synthesis and World Multiconference on Systemics,* Cybernetics and Informatics, 2001.

[9] L. Marcolino and L. Chaimowicz, “Traffic control for a Swarm of Robots: Avoiding Group Conflicts”, *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems,* 2009.

[10] F. Mazzini, D. Kettler, J. Guerrero and S. Dubowsky, “Tactile Robotic Mapping of Unknown Surfaces”, *With Application to Oil Wells, IEEE Transactions on Instrumentation and Measurement,* vol. 60, pp. 420-429, 2011.

[11] P. Srivastava, S. Satish and P. Mitra, “A distributed fuzzy logic based n-body collision avoidance system”, *Proc of the 4th Int Symposium on Intelligent Robotic Systems,* Bangalore, pp. 166–172, 1998.

[12] L. Zhang, T. Zhang, H. Wu, A. Borst and K. K¨uhnlenz, “Visual Flight Control of a Quad rotor Using Bio inspired Motion Detector”, *International Journal of Navigation and Observation,* Volume 2012, 9 pages, Hindawi pub, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 69–76

Application of Ant-Colony Optimisation to Compute Diversified Entity Summarisation on Semantic Knowledge Graphs

Witold Kosiński[1,2], Tomasz Kuśmierczyk[3], Paweł Rembelski[1], Marcin Sydow [1,3]
[1] Department of Computer Science, Polish-Japanese Institue of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa
[2] Institute of Mechanics and Applied Computer Science, Kazimierz-Wielki University,
ul. Chodkiewicza 30, 85-072 Bydgoszcz
[3] Institutue of Computer Science, IPI PAN, Warszawa, ul. Ordona
Email: wkos@pjwstk.edu.pl, t.kusmierczyk@phd.ipipan.waw.pl, rembelski@pjwstk.edu.pl, msyd@poljap.edu.pl

***Abstract*—We present ant colony optimisation approach, enriched with a novel self-adaptation mechanism, applied to solve DIVERSUM Problem that consits of generating a small diversified entity summarisation in a knowledge graph. The recently proposed DIVERSUM problem is viewed in this paper in a novel way as a NP-hard combinatorial optimisation problem. The presented preliminary experimental results indicate superiority of this approach to the previously proposed solutions to the DIVERSUM problem.**

***Index Terms*—Ant Colony Optimisation, Diversified Entity Summarisation, Max Sum Dispersion, Semantic Knowledge Graphs**

I. Introduction

TRADITIONAL optimisation techniques and optimisation methods are related to different types of problems. Some of them deal with constraint handling by using penalty methods, however, they often get stuck in local optima. Moreover, they usually need knowledge of first/second order derivatives of objective functions and constraints.

Hence one looks for more sophisticated methods, especially in cases when: Search space is discrete, discontinuous, non-convex, etc. or objective functions and constraints are non-differentiable or computationally expensive.

If we look at the branch of computer science, called *computational intelligence* then we find here 3 main divisions: nature inspired algorithms, fuzzy logic systems, neural networks.

In the last quarter of the previous century a dozen or more different optimisation algorithms have been proposed that are nature-based methods. One can list some of them: genetic, or more general – evolutionary algorithms, that are based on the Darwin's theory of evolution, swarm optimisation techniques, that copy the swarm intelligence, and they include: ant colony systems, particle swarm optimisation; and simulated annealing methods, that is based on the process of steel production; generalised extremal optimisation, that is based on point-wise equilibrium phenomenon, [1], in which the evolution happens step-wise in contrast to the Darwin continuous evolution, and many more methods.

In this paper we present application of the Ant Colony Optimisation (ACO) approach (Section II), to solve the DIVERSUM problem, i.e. a recently [2] introduced hard optimisation problem of Diversified Entity Summarisation on Knowledge Graphs (Section III). The DIVERSUM problem consists in selecting a small set S of arcs incident to a node x in a knowledge graph so that S gives a good "summary" of the node x. In this paper we present a diversity-aware objective measure so that the DIVERSUM problem is viewed as a combinatorial optimisation problem that is NP-hard.

Then, we report preliminary experimental results (Section IV) run on real data that illustrate the applicability of the presented ACO optimisation technique on the DIVERSUM problem and show that the peformance is higher compared with some previous solutions to the problem.

Ant Colony Optimisation approach was described in [3] and subsequently in a large number of publications. A more recent examples are [4] and [5] where also its application to a hard computational problem such as TSP was presented.

The problem of Diversified Entity Summarisation on Knowledge Graphs has been recently described and studied in [2]. In this paper, we study this problem from the optimisation perspective as optimising a properly defined diversity-aware objective function (Section III-C).

A. Contributions

The contributions of this paper include:

- novel self-adaptation mechanism for one of the parameters of ant colony optimisation algorithm, presented in Section II-G
- formulating the DIVERSUM problem as a combinatorial optimisation one, via defining a novel objective function (Section III-C)
- application of ant colony optimisation approach to such defined optimisation problem (Section III)
- promising preliminary experimental results on real data (Section IV)

978-83-60810-53-8/$25.00

II. Ant Colony Optimisation Algorithm

Ant System (AS), has been proposed by M. Dorigo in 1992, is a nature-based heuristic which tackles a range class of the real-life optimisation problems by information cooperation (i.e. *pheromone trials* evolution) inside a set $A = \{a_1, a_2, \ldots, a_{\mathfrak{m}}\}$ of $\mathfrak{m} \geq 1$ abstract agents called *ants*.

AS algorithm is an iterative method in which the three main rules are repeated sequentially:

- *Neighbour Choosing Rule* (AS-NCR) – which defines the probability of adding a new element to the solution being constructed,
- *Solution Construction Rule* (AS-SCR) – which stands for a process of constructing a complete solution for an input problem (AS-NCR is used inside),
- *Pheromone Update Rule* (AS-PUR) – which determines how the ants exchanges local information to build a global description of a space being under optimisation process.

One extra rule is often added, namely:

- *Pheromone Evaporation Rule (AS-PER)* – which defines quantity of evaporation (loss) of global information written down in pheromone trials.

Ant System heuristic schema:

while (stop condition is false) **do**
  for every ant a_i **do**
    construct solution with AS-SCR (AS-NCR inside)
      for every ant a_i **do**
      update pheromone trails with AS-PUR
  evaporate pheromone trials with AS-PER

Let us list the main components of the Ant System.

A. Neighbour Choosing Rule

Neighbour Choosing Rule is based on the probability of adding to constructed solution a new element x_j, just after x_i is previously added, is equal to

$$p_{i,j} = \frac{(\tau_{i,j})^{\alpha}(\eta_{i,j})^{\beta}}{\sum_{k \in \Theta}(\tau_{i,k})^{\alpha}(\eta_{i,k})^{\beta}}$$

where Θ is a set of indexes of all actually reachable elements and:

- $\tau_{i,j}$ – is pheromone trail value connected with an action of choosing element x_j just after an element x_i is chosen,
- $\eta_{i,j}$ – is heuristic value connected with an action of choosing element x_j just after an element x_i is chosen (this coefficient is defined within an input problem),
- α, β – are parameters that control the relative weight of pheromone trail and heuristic value; they need to be adapted.

One of the results of our paper is to define and implement a self-adaptation algorithm for the parameter α; here we assume $\beta = 0$.

B. Pheromone Update Rule

In Pheromone Update Rule for ant a_k the following action is executed

$$\tau_{i,j} \leftarrow (1-\rho)\,\tau_{i,j} + \Delta_k$$

where Δ_k is a value of fixed function of solution quality, and the following Pheromone Evaporation Rule has been already incorporated

$$\tau_{i,j} \leftarrow (1-\rho)\,\tau_{i,j}$$

where ρ is a fixed evaporation coefficient.

In our paper we are generalise the Ant System in the form of a *Discrete Ant System*. In its heuristic we face with a collective work, performed by a discrete number of objects and with the use of a discrete number of functions (tools), that may possess only a discrete number of different values. In this way the set of information collection by individuals - ants in order to construct individual solutions by the use of global knowledge of $\mathfrak{m}$ ants from the set A, is finite, as well. Now we define the optimisation problem for a simple ant system.

By an *optimisation problem* we understand a given quadruple $(\Sigma, \mathcal{R}, \Delta, \|\cdot\|)$, where:

- $\Sigma = \{x_1, x_2, \ldots, x_{\mathfrak{n}}\}$ – is a finite set of $\mathfrak{n}$ indexed *objects* (symbols),
- $\mathcal{R} \subset \Sigma^*$ – is a finite set of $\mathfrak{r}$ indexed *solutions* (words),
- $\Delta : \Sigma^* \to \{0,1\}$ – is a *solution acceptance function* such, that

$$\Delta(\omega) = \begin{cases} 1, & \text{if } \exists\left(\omega' \in \Sigma^*, \omega'' \in \mathcal{R}\right)\left(\omega \circ \omega' = \omega''\right) \\ 0, & \text{in other case} \end{cases}$$

- $\|\cdot\| : \mathcal{R} \to \mathbb{R}_+ \cup \{0\}$ – is a *solution quality function*, where we deal here with the minimisation problem for the quality function.

Notice that this definition of the optimisation problem is suitable for a wide range of real-world computation problems including NP-hard combinatorial problems. For example in the traveling salesman problem (TSP) the set Σ contains all labels of vertices of the graph under consideration and $\mathcal{R}$ is the set of all permutations of Σ. In a discrete knapsack problem, on the other hand, Σ is the set of all objects to be put into knapsack, while $\mathcal{R}$ is the set of words-solutions representing filling methods of the knapsack, taking into account all combinations and constrains put on the number of objects.

The solution $\omega^* \in \mathcal{R}$ is an *optimal* one if

$$\forall\,(1 \leq i \leq \mathfrak{r})\,(\|\omega_i\| \geq \|\omega^*\|)$$

Let $\mathcal{R}^* \subset \mathcal{R}$ denote the set of all optimal solutions. Our task in the optimisation problem $(\Sigma, \mathcal{R}, \Delta, \|\cdot\|)$ is to find any optimal solution $\omega^* \in \mathcal{R}^*$.

C. Pheromone Trail: Basic Variant

The value of saturation of pheromone trail left by an ant is upper bounded by $\tau_{max} \in \mathbb{N}_+$. We assume that

$$\mathbb{H} = \{1, 2, \ldots, \tau_{max}\}$$

is a *set of possible values* of saturation of pheromone trail. Then $F \in \mathbb{H}^{\mathfrak{n}}$ is a column vector of size $\mathfrak{n}$, which defines the *saturation level* of pheromone trail, where $F[i]$ is a value connected with an ant's possibility of choosing an object x_i. Next

$$\mathcal{F} = \{F_1, F_2, \ldots, F_{\mathfrak{f}}\}$$

is a finite set of all possible indexed vectors of saturation level. Notice that

$$\mathfrak{f} = (\tau_{max})^{\mathfrak{n}}$$

In our DAS algorithm we introduce a new object the matrix $H \in \mathbb{H}^{\mathfrak{n}\times\mathfrak{n}}$ of size $\mathfrak{n} \times \mathfrak{n}$, which determines the *saturation level* of pheromone trail, where $H[i, j]$ is a value connected with an ant's moving action from an object x_i to an object x_j. Thus

$$\mathcal{H} = \{H_1, H_2, \ldots, H_{\mathfrak{h}}\}$$

is a finite set of all possible indexed matrices of saturation level. Similarly

$$\mathfrak{h} = (\tau_{max})^{\mathfrak{n}^2}$$

D. Pheromone Trail – Probabilistic

Now let $\widehat{\mathbb{H}}$ be a set of discrete probabilistic values over set $\mathbb{H}$ of all possible values of saturation of pheromone trail, namely

$$\widehat{\mathbb{H}} = \left\{ \frac{a}{b} : a \in \{1, 2, \ldots, \tau_{max}\}, b \in \{1, 2, \ldots, \mathfrak{n} \cdot \tau_{max}\} \right\}.$$

where $0 < \frac{a}{b} \leq 1$. Then, we define the following *reduction function* Ω for the column vectors of set $\mathcal{F}$ where $\mathcal{F} \times \mathcal{R} \times \mathbb{R}_+ \to \widehat{\mathbb{H}}^{\mathfrak{n}}$ and $\Omega(F, \omega, \alpha) = \widehat{F}$ is a stochastic column vector, that according to the rule

$$\widehat{F}[i] = \begin{cases} \dfrac{F[i]^{\alpha}}{\sum\limits_{\{j:\Delta(\omega x_j)=1\}} F[j]^{\alpha}} & \text{if} \quad \Delta(\omega x_i) = 1 \\ 0 & \text{in other case} \end{cases}, \tag{1}$$

for α being a coefficient which describes an individual behavior of a single ant.

E. Neighbour Choosing Rule

The action of Neighbour Choosing Rule is governed by the nondeterministic function, called here $NCR : \mathcal{F} \times \Sigma^* \to \Sigma$ which at given $F \in \mathcal{F}$ and actually constructed word $\sigma \in \Sigma^*$ assume values x_i with the probability given by Eq.(1). It can be formulated as follows:

Remark 1. For any column vector $F \in \mathcal{F}$ and any word $\sigma \in \Sigma^*$ and arbitrary symbol $x \in \Sigma$ the following takes place

$$Pr(NCR(F, \sigma) = x) \in (0, 1), \text{ if } \Delta(\omega x) = 1, \text{and} \tag{2}$$

$$Pr(NCR(F, \sigma) = x) = 0 \text{ in other cases}.$$

This is the basis of the next nondeterministic evolution mechanism in DAS, related to the solution construction rule *SCR*, where $SCR : \mathcal{F} \times \mathcal{H} \to \mathcal{R}$. This rule is the composition of a sequence of independent events generated by a multiple application of the mechanism NCR, unless an internal stop condition is meet. In the reality after the initiation of the algorithm and the formation of the first element of the word-solution, according to the column vector F and

$$\omega \leftarrow NCR(F, \varepsilon), \tag{3}$$

at some i-th stage of the iteration process we obtain a word $\omega = x_{l_1}, x_{l_2}, \ldots, x_{l_i}$, say, then as the result of the l_i-th column vector of the matrix H, denoted here by $H[l_i, \cdot]$, we determine the next element

$$x_{l_{i+1}} \leftarrow NCR(H[l_i, \cdot], \omega) \tag{4}$$

with the probability given by the property of the rule NCR. Then we add the next symbol xl_{i+1} at the end of the word ω, i.e.

$$\omega \leftarrow \omega x_{l_{i+1}}. \tag{5}$$

From this description we can formulate the following remark concerning the probabilistic nature of SCR.

Remark 2. For any column vector $F \in \mathcal{F}$, an arbitrary matrix $H \in \mathcal{H}$ and a word-solution $\omega \in \mathcal{R}$ of the form $\omega = x_{l_1}, x_{l_2}, \ldots, x_{l_r}$, the following is true:

$$Pr(SCR(F, H) = \omega) =$$

$$= Pr(NCR(F, \varepsilon) = x_{l_1}) \cdot \prod_{i=1}^{r-1} Pr(NCR(H[l_i, \cdot], x_{l_1}, x_{l_2}, \ldots, x_{l_i}), \tag{6}$$

which means, by Remark 1, that

$$Pr(SCR(F, H) = \omega) \in (0, 1). \tag{7}$$

The last statement has a fundamental meaning for our nondeterministic algorithm: each word-solution $\omega \in \mathcal{R}$, and by this each optimal word-solution from $\mathcal{R}^* \subset \mathcal{R}$ may be obtained with a positive probability, as a result of the application of the SCR rule in an arbitrary pheromone structure F and H. To these, rather static, mechanisms NCR and SCR we are adding a dynamic one that is related with the actualisation of the pheromone intensity on the trails, governed by the rule PUR. In this way we are introducing a dynamic exchange of information.

The PUR mechanism has two components: short range and long range, and is deterministic. They act on elements of the sets $\mathcal{F}$ and $\mathcal{H}$. We are adding (increasing) some amount of pheromone, by the application of two operators

$$\text{inc}_{\mathcal{F}} : \mathcal{F} \times \mathcal{R} \to \mathcal{F}, \& \ \text{inc}_{\mathcal{H}} : \mathcal{H} \times \mathcal{R} \to \mathcal{H}, \tag{8}$$

and we are decreasing the amount of pheromone by the application of two next operators

$$\text{dec}_{\mathcal{H}} : \mathcal{H} \times \mathcal{R} \to \mathcal{H}, \& \ \text{dec}_{\mathcal{H}} : \mathcal{H} \times \mathcal{R} \to \mathcal{H}. \tag{9}$$

If $\omega = x_{l_1}, x_{l_2}, ..., x_{l_r}$ is a word-solution, then the action of $\mathrm{inc}_{\mathcal{F}}$ on the vector F is

$$F[l_1] \leftarrow min(\tau_{max}, F[l_1]+1)$$

and the action of $\mathrm{dec}_{\mathcal{F}}$ is

$$F[l_1] \leftarrow max(1, F[l_1]-1),$$

on the vector element with the index l_1, being the index of the first element of the solution ω.

The application of two other operators $\mathrm{dec}_{\mathcal{H}}$ and $\mathrm{dec}_{\mathcal{H}}$ is governed by two formula:

$$H[l_i, l_{i+1}] \leftarrow min(\tau_{max}, H[l_i, l_{i+1}]+1), \&$$
$$H[l_i, l_{i+1}] \leftarrow (max(1, H[l_i, l_{i+1}-1]-1)$$

respectively, for all elements $H[l_i, l_{i+1}]$ of the matrix H, where $i = 1, 2, ..., r$ is the subsequent index of the symbols of the solution ω.

F. Theoretic model of DAS

Now we describe the theoretical model of DAS for the single ant, because of its simplicity. The extension for a set of ants is rather of technical nature, because of the sequential character of the algorithm in a loop.

By a state of an ant at time t we understand a quadruple

$$(F_{(t)}, H_{(t)}, \omega_{(t)}, \omega^*_{(t)})$$

where all objects, but the last one, have been already introduced and the index (t) means that it is their actual values at time t. The last object is the historical value of the best word-solution obtained till the moment t.

From the previous considerations we may form the following obvious proposition.

Proposition 1.

A. A state of the ant at time $t+1$ depends on the state of the ant at time t, only.

B. The set of all possible states of the ant is finite.

Now we may introduce an algebraic model of an DAS algorithm evolution for a single ant, where:

- $\widehat{U}_{(t)} \in [0,1]^{\mathfrak{s}}$ – is a column stochastic vector of size $\mathfrak{s}$ such, that $\widehat{U}_{(t)}[i]$ determines a value of probability of a chance that an ant state in moment t is s_i,
- $\widehat{T} \in [0,1]^{\mathfrak{s}\times\mathfrak{s}}$– is a column stochastic matrix of size $\mathfrak{s}\times\mathfrak{s}$ such, that $\widehat{T}[i,j]$ determines a value of probability of a chance that an ant changes its state from s_i to s_j.

Therefore

$$\widehat{U}_{(t)}\widehat{T} = \widehat{U}_{(t+1)} \tag{10}$$

gives a probabilistic evolution of an ant state between moments t and $t+1$. In general if $\widehat{U}_{(0)}$ is an initial distribution of probability of an ant start state, then

$$\widehat{U}_{(i)} = \widehat{U}_{(0)}\widehat{T}^i \tag{11}$$

describes ant state at moments $i = 1, 2, 3, \ldots$. Hence we formulate the next proposition.

Proposition 2.

Evolution process of a single ant in our theoretical model of DAS is a Markovian chain.

The method how to fill the matrix $\widehat{T}$ is described in [6], [7]. The question of the convergence has been solved positively as well, by the application of the convergence theorem formulated for the case of evolutionary algorithms in [8]. In fact we may formulate the result on the so-called pointwise convergence of the DAS, by enlarging the space Σ^* and adding a super-state composed of all states in which the last elements are undistinguishable by the solution quality function.

G. Self-adaptation of α parameter

For the purpose of numerical experiments presented in the article, we introduced a new mechanism for self-adaptation of α parameter occurring in the equation determining the probability of obtaining by a single ant a solution $\omega \in \mathcal{R}$ (see the equation at the beginning of Section II-E and Eq.1). The mechanism is based on the notion of radius variation of α parameter which is further denoted by γ. For practical reasons (arithmetic capabilities of computer's CPU) we aim to reduce γ possible values by $0 \leq \gamma < max$, where *max* is some fixed constant. Then, value of α parameter at time t is in range from 1.0 to $1.0+\gamma$ (strengthening the pheromone trace) or from $\frac{1.0}{1.0+\gamma}$ to 1.0 (reducing the pheromone trace).

Initially (time $t = 0$) condition $\gamma = 0.0$ is satisfied, thus α parameter is equal to 1.0 (neutral state). In each subsequent iteration $t > 0$ we increment γ radius by $\frac{1}{N}$, where N is the size of the ant nest. Next α^* is value of α coefficient which was previously used while currently best result ω^* was found. Further α self-adaptation mechanism is in accordance with the following rules:

- if $|\alpha - \alpha^*| \leq \frac{1}{N}$ holds, then α value is reset randomly with uniform probability in the range from 1.0 to $1.0+\gamma$ or form $\frac{1.0}{1.0+\gamma}$ to 1.0,
- if $|\alpha - \alpha^*| > \frac{1}{N}$ holds, then α value is reset to $\frac{\alpha+\alpha^*}{2}$ (bisection scheme).

Finally, if solution ω found at time t is better than solution so far the best ω^*, γ is again set at 0.0.

The above mechanism of self-adaptation of α parameter can be limited in the space of discrete values. Thus, all of the properties of Discrete Ant Algorithm introduced before, including most significant pointwise convergence, are sustained.

III. Application to Entity Summarisation

In this section, we demonstrate the application of the described optimisation technique on an interesting NP-hard optimisation problem concerning entity summarisation on semantic knowledge graphs that was recently proposed in [2].

A. Semantic Knowledge Graphs

A semantic knowledge graph (henceforth denoted as KG) is a quite novel format for represented semantic data. Knowledge graphs can be automatically or semi-automatically constructed in a process of "knowledge harvesting" from large corpora of text documents e.g. from the WWW, with use of advanced open-domain information extraction technology. There are

existing large datasets in such formats, e.g. DBpedia [9] or YAGO [10]

Basicly, KG constists of: fact graph and ontology.

The fact graph is a directed multi-graph where

- each node represents some entity (e.g. musician, actor, politician etc.) from some domain (e.g. art, movies, politics, etc.), e.g. "Fryderyk Chopin"
- each directed arc represents some "fact" concerning the entities being its ends, e.g. "Fryderyk Chopin is composer". Such facts are commonly represented as so called RDF-triples, that consist of subject, predicate and object, that is often denoted in the relational form: `predicate(subject, object)` (e.g. `is(Fryderyk Chopin, composer)`.

The ontology represents type hierarchy of the entities, i.e. each entity is connected to type node(s) in the ontology by special arcs like "type", etc.

B. Entity Summarisation

[2] studies a problem of entity summarisation in KG, i.e. given an entity (to be summarised) x, knowledge graph G and (small) $k \in N^+$ (a limit on number of facts in the summary) to select a set S of facts concerning the entity x that make a concise "summary" of x. In graph terms the problem consists in selecting a small set of "representative" arcs incident with x in the graph.

Since a node in a large KG (like DBpedia or YAGO) can easily have degree of 100 or higher, and the typical value of k is around 10, it naturally leads to a hard optimisation problem of how to select "the best" facts to the summary.

C. Novel Diversity-aware quality measure of entity summary

A good summary should select the most *important* facts concerning the entity. Furthermore, as observed and experimentally confirmed in [2], a desired summary would be *diversified* i.e. contain facts that concern various *aspects* of the entity being summarised.

This would be stated as a combinatorial optimisation problem via a properly defined bi-criteria objective function obj that takes into account 1) importance of facts, 2) mutual "dissimilarity" between facts.

More precisely, the problem can be defined as: out of the (given) set D of all facts concerning the entity, select a subset S of (up to) k facts so that the following measure is maximised:

$$obj(S) = (k-1)\sum_{d\in S} imp(d) + 2\delta \sum_{d_1 \neq d_2 \in S} diss(d_1, d_2)$$

where $imp : D \to Q^+$ is a "importance" weight of a fact and $diss : D^2 \to Q^+$ is represents pairwise "dissimilarity" of two facts and δ is a parameter to balance between the two criteria of "importance" and "dissimilarity" (to be tuned experimentally). We assume that imp is an increasing function of importance and $diss$ is an increasing function of dissimilarity of facts.

E.g. given entity "Albert Einstein" and two facts concerning this entity: d_1= `hasWonPrize(Albert Einstein, Nobel Prize)` and d_2 = `hasWonPrize(Albert Einstein, Mateucci Medal)` it seems reasonable that d_1 is more important than d_2. In addition, given a fact d_3=`hasChild(Albert Einstein, Evelyn Einstein)` it seems reasonable that $diss(d_1, d_2)$ is lower than $diss(d_1, d_3)$, etc.[1].

The imp and $diss$ functions can be computed based on structural and statistical properties of the underlying knowledge graph. In this paper we assume that the values of the functions are computed externally and we do not focus on this issue.[2]

The factors of $(k-1)$ and 2 are used in the formula to reflect the fact that there are $k(k-1)/2$ possible pairs of a k-element set.

The problem of optimising the obj function defined above is NP-hard (i.e. it can be reduced from MaxSumDispersion NP-hard optimisation problem). If the $diss$ function was a metric, one would adapt the existing 2-approximation algorithm for the MaxSumDispersion problem to this problem.

In the next section, we experimentally apply the ACO optimisation method described in Section II to solve hard optimisation problem defined here. This is additionally justified by the fact that $diss$ function is not necessarily a metric what excludes the applicability of some known approximation algorithms for the MaxSumDispersion problem.

TABLE I
ENTITIES SELECTED FOR EVALUATION.

Entity	Number of facts	Number of predicates
Albert Einstein	32	10
John Wayne	135	12
Denzel Washington	34	7
Robert Mitchum	61	5

IV. EXPERIMENTS

To study practical properties of the described approach we used YAGO2 semantic knowledge database [10] as the underlying knowledge graph. We run the experiments on a sample of entities representing some known people. Some of them, that represent actors, were earlier used for evaluation in [2] and some (e.g. "Albert Einstein") are new. Table I presents selected entities with number of facts assigned. Also number of unique predicates per entity is shown in third column. One can see that the number of facts varies much between entities but for each of them is much higher than 12. This implies that applying a brute force approach of considering all possible 12-element subsets of incident arcs (facts) would be of prohibitive time complexity.

For each fact of an entity we calculated its importance weight and for each pair of facts we calculated dissimilarity. Weights and dissimilarities were later used in $obj(S)$.

A weight of a fact is given by normalised weight of an object (ending of an edge). We applied undirected random

[1]All the mentioned examples are real and taken from the YAGO dataset

[2]The detailed description of how to compute imp and $diss$ functions and other concepts concerning diversified entity summarisation measure mentioned above are out of scope of this publication and deserves a separate publiction that is currently under preparation.

walk (with probability of getting back to the entity set to 0.15) over graph of facts. From this we obtained global weights of entities. In the next step we took facts (with associated endings = objects) associated to a studied entity and scaled weights in a way to fit in range $[0,1]$. At the end, the most important fact assigned to the entity has weight 1.0 and the least important $\varepsilon(\approx 0)$.

A dissimilarity of two facts associated with an entity was calculated using information about types of the entity. Using information about neighbourhood in hierarchy of types we measure how often both facts occur in similar entities. By similar we understand these of close types. Applied method returns dissimilarities always in range $[0,1]$.

A. Preliminary Results

Figure 1 presents sample results ($k=7$) of our method for $\delta = 0.0$ (diversity oblivious variant) and $\delta = 10.0$ (diversity aware variant) in comparison to the results of diversity aware algorithm from [2]. Although we show full results only for single entity ("Denzel Washington") outcomes for other entities have similar properties. Facts shown on different images were placed to resemble their similarity. One should note that since the time of writing [2] the YAGO2 database has been substantially updated and that could slightly influence final results. However we can see that in Diversum image and in image for $\delta = 10.0$ some facts are equal e.g. `directed(D.Wash., The Great Debaters)`, `influences(D.Wash., Noah Sife)`. Correspondence between other facts can also be found e.g. a fact `actedin(D.Wash., DJ Vu)` was replaced with fact `actedin(D.Wash., The preacher's wife)`, information about birth date and children were replaced with place of birth, wife and gender. Similar correspondence can be found between facts `hasWonPrize(D.Wash., Academy Award for Best Actor)` and `hasWonPrize(D.Wash., Tony Award)`.

Additional remarks can be stated after comparison of images for $\delta = 0.0$ and $\delta = 10.0$. One can see that for $\delta = 0.0$ facts connected to acting career dominate summarisation. Information about influence, wife and birth place was replaced with further information about movies that Denzel Washington acted in.

To measure overall quality of obtained summarisations we used Wikipedia info-boxes. Our method is based on this used in [2]. Let $S(e)$ (where $|S(e)| = k$) denote set of facts in summarisation of an entity e and $W(e)$ denote set of facts that can be found in info-box on corresponding Wikipedia pages. Let $f_1 \rhd f_2$ mean that f_1 is equal or more specific fact than f_2 and f_2 can be inferred from f_1. For example: the fact that someone is an actor can be inferred from the fact that he played in some specific movie. $Recall'$ is then defined as:

$$Recall'(e) = \frac{|\{f : f \in W(e) \wedge \exists_{f' \in S(e)} f' \rhd f\}|}{|W(e)|}$$

Tables II and III present comparison of our method to results of Diversum and Precis [2]. $Recall'$ measure for $obj(S)$ with $\delta = 0/1/10$ and for Diversum and Precis is shown. Table II contains results for $k=7$ and Table III for $k=12$. One can see that our approach performs much better than Precis which seems to be the weakest one. The second worst approach is $obj(S)$ with $\delta = 0$ what shows how important diversity is. Diversum algorithm is much better and its results are comparable to our method with $\delta = 1$. When δ is set to 10 and diversity is high quality is the best according to the measure. Overall results are promising but need to be verified in further experiments on more entities.

Authors of $Recall'$ assumed that info-boxes have size of about 7-12. This is not necessary true in our case e.g. for "Albert Einstein" there is 54 facts. Therefore we suggest

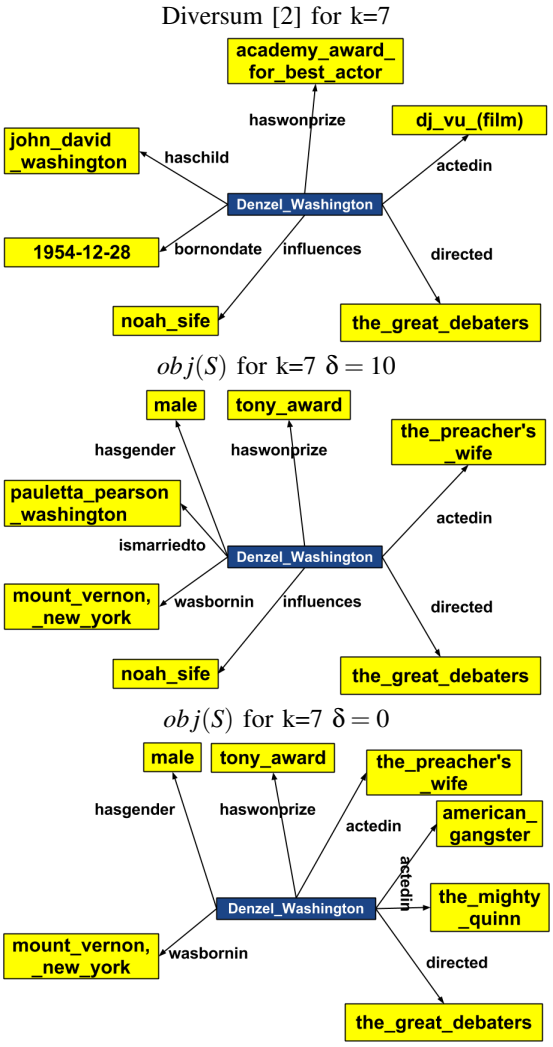


Fig. 1. Comparison of results for "Denzel Washington".

TABLE II
COMPARISON OF $Recall'$ FOR $k=7$.

Entity	δ=0	δ=1	δ=10	Div.	Prec.
J. Wayne	0.31	0.38	0.31	0.29	0.06
D. Wash.	0.50	0.50	0.67	0.60	0.20
R. Mitch.	0.30	0.50	0.50	0.50	0.25
mean:	0.37	0.46	0.49	0.46	0.17

TABLE III
COMPARISON OF *Recall'* FOR $k = 12$.

Entity	δ=0	δ=1	δ=10	Div.	Prec.
J. Wayne	0.46	0.54	0.54	0.42	0.59
D. Wash.	0.50	0.67	0.67	0.60	0.20
R. Mitch.	0.30	0.50	0.50	0.50	0.25
mean:	0.42	0.57	0.57	0.51	0.35

TABLE IV
DIFFERENT MEASURES FOR "D. WASHINGTON" $k = 7$.

δ	*Recall'*	*Recall*	*Precision*	F_1
0	0.50	0.43	0.71	0.54
1	0.50	0.43	0.57	0.49
10	0.67	0.57	0.57	0.57

modified measures:

$$Recall(e) = \frac{|\{f : f \in W(e) \wedge \exists_{f' \in S(e)} f' \rhd f\}|}{|S(e)|}$$

$$Precision(e) = \frac{|\{f' : f' \in S(e) \wedge \exists_{f \in W(e)} f' \rhd f\}|}{|S(e)|}$$

Precision says how many facts out of returned k infer facts from info-box. *Recall* measures how many facts in info-box can be derived from summarisation or in different words: what fraction of facts we covered out of k possible to cover. As long as single fact from summarisation can infer one or zero facts in info-box, what is a specific property of considered data, *Recall* is limited to 1.0.

One should note that Wikipedia info-boxes contain some types of facts that are not included in YAGO2. Therefore we decided to update sets $S(e)$ in a way to keep only this facts that can be connected to facts in knowledge base according to relation $\rhd$.

Sample comparison of different measures for entity "Denzel Washington" can be found in Table IV. For this entity there is only 6 facts from Wikipedia info-box included into $S(e)$ therefore $Recall < Recall'$. Recall that parameter δ controls diversity of results. For $\delta = 0.0$ only weights are taken into account and dissimilarities are omitted. For $\delta = 1.0$ we expect results balanced in weight and diversity and for $\delta = 10.0$ we expect highly diversified results. One can note that *Precision* is very high for $\delta = 0$. The reason is that from every fact "acted in movie" we can infer that "D. Washington" is an actor. From practical point of view it would be enough to include single fact of such type. This situation points out vulnerability of this measure.

Tables V and VI presents average values of four entities from Table I of measures for $k = 7$ and $k = 12$. An analysis shows that both *Recall* and *Recall'* increase when δ increases. It means that for higher values of δ more facts from info-boxes is covered. *Precision* is the highest for $\delta = 0$. The reason of this behaviour was explained earlier (in context of "Denzel Washington" entity). The highest values of F_1 were obtained for $\delta = 10$ and at the end this value seem to be the best option.

TABLE V
AVERAGE (FOUR ENTITIES) MEASURES FOR $k = 7$.

δ	*Recall'*	*Recall*	*Precision*	F_1
0	0.30	0.50	0.75	0.59
1	0.36	0.61	0.68	0.64
10	0.39	0.61	0.68	0.64

TABLE VI
AVERAGE (FOUR ENTITIES) MEASURES FOR $k = 12$.

δ	*Recall'*	*Recall*	*Precision*	F_1
0	0.34	0.36	0.77	0.46
1	0.46	0.46	0.73	0.55
10	0.46	0.46	0.73	0.55

To study deeper the influence of a parameter δ in $obj(S)$ we performed three experiments for different values (0.0, 1.0, 10.0) of this parameter. Tables VII and VIII compare selected facts for different values of parameter δ. Each column says what fraction of facts is common for results with two different values of δ e.g. second column ("1 vs. 0") says how many facts is retained when $\delta = 0.0$ is changed to $\delta = 1.0$. The Table VII presents results for $k = 7$ and Table VIII for $k = 12$. Additionally Figure 2 visually compares second columns from tables.

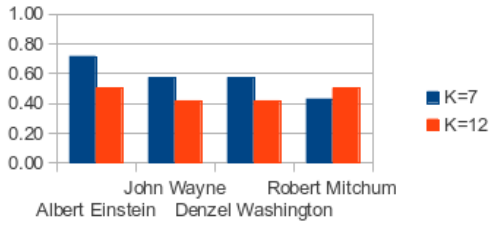


Fig. 2. Fraction of facts shared for $\delta = 0.0$ and $\delta = 10.0$.

TABLE VII
COMMON FACTS FOR DIFFERENT VALUES OF δ (K=7).

Entity / δ	1 vs. 0	10 vs. 0	10 vs. 1
A. Einstein	0.71	0.71	1.00
J. Wayne	0.57	0.57	0.57
D. Wash.	0.57	0.57	0.71
R. Mitchum	0.43	0.43	0.57
mean:	0.57	0.57	0.71
std:	0.12	0.12	0.20

TABLE VIII
COMMON FACTS FOR DIFFERENT VALUES OF δ (K=12).

Entity / δ	1 vs. 0	10 vs. 0	10 vs. 1
A. Einstein	0.50	0.50	0.92
J. Wayne	0.50	0.42	0.75
D. Wash.	0.50	0.42	0.92
R. Mitchum	0.50	0.50	1.00
mean:	0.50	0.46	0.90
std:	0.00	0.05	0.10

An analysis of Tables VII and VIII and Figure 2 leads to several remarks. At first, when δ is changed from 0.0 to 1.0 or 10.0 about half of facts is retained. It holds both for $k = 7$ and $k = 12$. However facts selected for $\delta = 1.0$ and $\delta = 10.0$ differs much e.g. about 30% facts differs for $k = 7$ and about 10% for $k = 12$. The difference in behaviour can be explained by small differences between top facts.

What is more, one can see that there are entities that are resistant to changes of δ. Figure 3 presents comparison of facts shared between situations $\delta = 1.0$ vs. $\delta = 10.0$ for $k = 7$ and $k = 12$. One can see that for "Albert Einstein" all facts are retained. By checking Table VII one can also see that for this entity adding information on dissimilarity ($\delta \neq 0$) changed results much less than for the others.

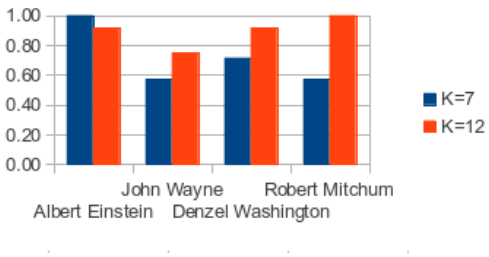


Fig. 3. Fraction of facts shared for $\delta = 1.0$ and $\delta = 10.0$.

V. Conclusions

Results obtained at current level of advance of our project are very promising. In all cases our results were not worse than these described in [2]. Most importantly, the preliminary results indicate appropriate tuning of the δ parameter in the newly proposed solution quality measure makes it possible to beat the performance of the previously proposed diversity-aware algorithm (DIVERSUM) presented in [2]. Anyways, further experiments and evaluations need to be done to confirm this, including user-based evaluations.

Also the issue of tuning the δ parameter should be separately studied as well as theoretical properties of the proposed measure.

In future work it would be also interesting to experimentaly compare the performance of the presented optimisation technique, especially the influence of the novel self-adaptation mechanism, with some other existing sub-optimal approaches, e.g. approximation algorithms for the Max Sum Dispersion problem and other NP-hard optimisation problems.

Acknowledgements

The work was supported by the Polish National Science Centre (NCN) grant N N516 481940 (DIVERSUM project) and partially by grant N N519 5788038 and by research fellowship within "Information technologies: research and their interdisciplinary applications" agreement POKL.04.01.01-00-051/10-00.

References

[1] N. Eldredge and S. J. Gould, "Punctuated equilibria: An alternative to phyletic gradualism," in *Models in Paleobiology*, T. J. M. Schopf, Ed. San Francisco: Freeman Cooper, 1972, pp. 82–115.

[2] M. Sydow, M. Pikuła, and R. Schenkel, "The notion of diversity in graphical entity summarisation on semantic knowledge graphs," *Journal of Intelligent Information Systems*, pp. 1–41, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10844-013-0239-6

[3] M. Dorigo, "Optimization, learning and natural algorithms (in Italian)," Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy, 1992.

[4] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *Computational Intelligence Magazine, IEEE*, vol. 1, no. 4, pp. 28–39, 2006.

[5] M. Dorigo and L. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *Evolutionary Computation, IEEE Transactions on*, vol. 1, no. 1, pp. 53–66, 1997.

[6] P. Rembelski, "Theoretical model of sas ant colony optimisation algorithm (*In Polish*)," pp. 37–42, 2010, proc. of XII International PhD Workshop OWD 2010.

[7] W. Kosiński and S. Kotowski, "Limit properties of evolutionary algorithms," *In-Tech*, pp. 1–28, 2008.

[8] J. Socała, W. Kosiński, and S. Kotowski, "On asymptotic behaviour of a simple genetic algorithm (*In Polish*)," *Matematyka Stosowana*, vol. 6/47, pp. 70–86, 2005.

[9] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," *The Semantic Web*, vol. 4825, pp. 722–735, 2007. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-76298-0-52

[10] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *Proceedings of the 16th international conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 697–706. [Online]. Available: http://doi.acm.org/10.1145/1242572.1242667

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 77–82

Semantic Tagging of Heterogeneous Data: Labeling Fire&Rescue Incidents with Threats

Adam Krasuski* and Andrzej Janusz†
*Chair of Computer Science, The Main School of Fire Service
Słowackiego 52/54, 01-629 Warsaw, Poland
†Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
krasuski@inf.sgsp.edu.pl, janusza@mimuw.edu.pl

***Abstract*—In the article we present a comparison of the classification algorithms focused on labeling Fire&Rescue incidents with threats appearing at the emergency scene. Each of the incidents is reported in a database and characterized by a set of quantitative attributes and by natural language descriptions of the cause, the scene and the course of actions undergone by firefighters. The training set for our experiments was manually labeled by the Fire Service commanders after deeper analysis of the emergency description. We also introduce a modified version of Explicit Semantic Analysis method and demonstrate how it can be employed for automatic labeling of the incident reports. The task we are trying to accomplish belongs to the multi-label classification problems. Its practical purpose is to support the commanders at a emergency scene and improve the analytics on the data collected by Polish State Fire Service.**

***Keywords*-Domain Knowledge, Multi-label Classification, Explicit Semantic Analysis, Fire Services**

I. Introduction

THE MAIN goal of Fire Services activity at the fire ground is elimination or neutralisation of arisen threats. Therefore, the core of the Fire&Rescue (F&R) action is to adequately recognize possible dangers for the involved people and properties. A specific emergency generates specific threats. It implies that if we posses a description of the emergency we could predict dangers, or more precisely threats, related to the actual emergency.

The recognition of the threats at the fire ground is only a part of activities that should be performed at the beginning of a F&R action. No less important is the recognition of threatened individuals or objects. Together, those two tasks play a pivotal role in planning of the further actions at the emergency scene.

The task of recognition and categorization of threats is formalized in the tactic of German Fire Service [1]. After arriving at a fire ground or an emergency scene German commanders have to evaluate and recognise the appearing threats. In order to do this systematically and not to miss any of the threats, they have to fulfill the Threats Matrix (in German – Gefahrenmatrix) [1]. The Threats Matrix helps to identify the threats emerging at the scene and the threatened objects. The columns of the matrix represent threats, and the rows represent objects which can be threatened. The Table I depicts the Threats Matrix.

TABLE I
The Threats Matrix used by German commanders. Legend: A1 – Fear, A2 – Toxic smoke, A3 – Radiation, A4 – Fire spreading, C – Chemical substances, E1 – Collapse, E2 – Electricity, E3 – Disease or injury, E4 – Explosion

Threat/object	A1	A2	A3	A4	C	E1	E2	E3	E4
People (ME)									
Animals (T)									
Environment (U)	–					–	–	–	
Property (S)	–	–						–	
Rescuers (MA)									
Equipment (G)	–	–						–	

In German language, column names are chosen so that they can be easily remembered. In order to help to memorize all threats by commanders, German threats' names were taken to form the following pattern: AAAA-C-EEEE *Angstreaktion, Atemgifte, Atomare Strahlung, Ausbreitung, Chemische Stoffe, Einsturz, Elektrizität, Erkrankung, Explosion*. The sign '–' in the table indicates, that this threat in general does not apply to this object. At the background of filled Threats Matrix, German commanders define the Threat Focus and organize their actions accordingly.

If we could create a computer system which can recognize threats at the fire ground, we would effectively support the commanders in doing their duty. Moreover, our previous research [2], [3], [4] show that analytics performed in abstract spaces, such as Threats Matrix allows to reduce significantly the number of dimensions without loosing the information about complexity of the real phenomena. Unfortunately, the Polish Fire Services do not use the method of filling the Threats Matrix at the incident scene. Therefore Polish Incident Data Reporting System called EWID lacks of this information. In the previous work [4] we labeled incidents with threats manually. The reports from EWID database were analysed and labeled by extramural students of The Main School of Fire Service with commanding experience. We selected only commanders having at least seven years experience in commanding. They were involved as *experts – practitioners* in labeling real action reports from the EWID system.

We created a special system to support manual labeling the reports. The labeling process consists of two main phases:

978-83-60810-53-8/$25.00

tutorial phase and *labeling phase*. The tutorial phase was focused on introducing the Threats Matrix and the layout of EWID incident reports to the experts. It was divided into three consecutive parts. In the first part, experts were introduced to the format and purpose of the Threats Matrix. In the second part, some examples of filled Threats Matrices were presented and discussed with the experts. In the third part, experts received an exemplary EWID report together with a Threats Matrix describing this report. The labeling phase consisted of many evaluation stages. At every stage the experts were provided with a single EWID report. On the ground of the information about the incident described in the report, they were asked to evaluate threats which appeared during the reported incident and to complete its Threats Matrix. Every expert was asked to label at least 100 EWID reports. Every report description was labeled by only one expert. In total, we collected 406 labeled incident descriptions.

The presented method has very serious shortcomings – it is not scalable. If we need more labeled reports we need more commanders. Up to now approximately 7 million reports have been collected In the EWID database. Moreover, every day 1 500 new reports are submitted into the system. It is obvious that such a number of incidents is not manually manageable by people.

This article is devoted to prediction methods from Machine Learning which can be used to label the incidents automatically. We use the 406 labeled incidents as a data set to train and evaluate our classification algorithms. We analyse different data representation and different multi-label classification methods in order to find the best one.

The remaining of the paper is structured as follows. In Section II we describe our data set which was used as an input in our experiments. In Section III we present our method of the analysis focused on determining the best classification algorithm and data representation. Section IV contains the results of the conducted experiments. The article is concluded with the interpretation and a summary of the research results, as well as a discussion on perspectives for the future research.

II. Description of the Data

Our data set consists of 291 683 F&R reports. They contain information about the incidents responded by Fire Service, from the years 1992 to 2011. The data concerns the incidents which happened in Warsaw City and its surroundings. In this data set 136 856 reports represent fires, 123 139 local threats and 31 688 where false alarms.

Each of the reports consists of an attribute section and a natural language description part. The attribute section contains 506 attributes describing all types of incidents. However, depending on category of the incident, the number of attributes that take values different than zero varies from 120 to 180 for a report. Most of the attributes are boolean (True/False) type but there are also numerical values (i.e. fire area, amount of water used).

The natural language description (NL) part is an extension to the attribute part. It was designed to store information, which can not be represented in a form of a set of attributes. Unfortunately there is no clear regulation what should be written in the NL part. Therefore, in this part a full spectrum of information, from detailed information including time coordinates, to the very general and brief descriptions can be found. The simple statistics reveal that NL part contains approximately three sentences that describe the situation at the fire ground, actions undertaken and weather conditions. Figure II depicts the idea of a report representation in EWID database.

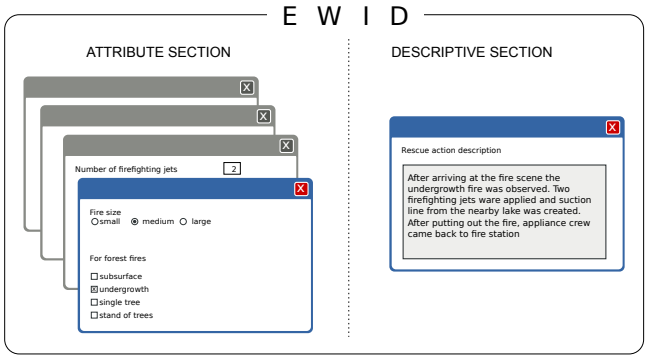


Fig. 1. Representation of a report in EWID database.

In factual aspects, the data stored in the EWID contain information about persons, objects involved in the incident and methods used to eliminate the arisen threats.

In our experiments we used a subset of this data set. For the process of labeling (assigning threats) the incidents by domain experts, we selected only the reports representing fires of residential buildings. This subset of the data consisted of 31 556 reports. From this set 406 reports were labeled by the experts. We used these reports in our experiments described in Section III.

III. Method

The labeling methodology was briefly presented in Section I and is broadly discussed in [4]. In this section we present several approaches to automated labeling of the reports.

In this research we pay a special attention to two aspects of the task: finding an appropriate classification algorithm and selecting a good input data representation. The first approach can be divided into two groups: classifiers which operate on incident features and classifiers which operate on features of the threats. The set of possible representations for the second task consists of: structured part only (SP), NL part only (NLP), structured part plus bag-of-words of descriptions of object (SP-OD), structured part bag-of-words and NL part transformed to the LSA space (SP-OD-LSA). In the next subsections we describe the utilized methods in detail. All the performance evaluation experiments were conducted on a training set of 285 incidents and a test set consisting of 122 incidents.

A. Classification on Structured Part Only

As was mentioned in section II the structured of part EWID database is represented by 506 attributes. However, for our subset of 406 incidents many of them have zero values for each of the incidents. Therefore, we removed those attributes from our subset. We also removed semantically irrelevant attributes, such as ID of a fire station. As a result we obtained an information system with 208 attributes, from which 24 were numeric and 184 were of a boolean type. The prediction targets were sets of combinations of threats and threatened objects (the risks). The set of the possible risks was created as a Cartesian product of threats and objects from the Threats Matrix. Such a representation constituted an input for our classifiers.

Our first experiment was focused on determining the best classifier for a given representation. In this experiment we used the whole set (406 cases) and 5 folds cross-validation technique to evaluate the efficiency of different classifiers. We tested: Naive-Bayes (NB), Classification Tree, Support Vectors Machine (SVM), Clark Niblett induction algorithm (CN2) and Random Forrest. We used Matthew correlation coefficient (MCC)[1] to evaluate the efficiency of the selected classifiers. Table II depicts the comparison of the results.

TABLE II
COMPARISON OF THE CLASSIFIERS. AUC – AREA UNDER CURVE, MCC – MATTHEW CORRELATION COEFFICIENT, SVM – SUPPORT VECTORS MACHINE, CN2 – CLARK NIBLETT INDUCTION ALGORITHM.

Method	AUC	F1-score	MCC
Naive Bayes	0.76	0.61	0.33
Classification Tree	0.71	0.53	0.26
SVM	0.75	0.46	0.27
CN2 rules	0.76	0.50	0.31
Random Forest	0.71	0.45	0.20

According to the criteria (MCC measure) presented in Table II the Naive-Bayes classifier obtained the best results and was selected as a representative for the rest of the experiments. Next, we used training and test methodology to compare the current methods with other approaches. The classifier was trained and tested separately for each of the decisions classes. Then we calculated some performance measures i.e. precision, recall, F1-score in two way: for each of the incidents and for each of the decision classes.

B. Classification on Structured and Object Description Parts

We repeated the experiment described in Section III-A extending the incident representation by the *object description* attribute. The attribute is an extension to the *object type* attribute and contains information such as: storey or room of the building where fire occurred, trash localization (inside, outside in the case of a trash fire), etc. This attribute stores the information in natural language form. In order to use this part as a feature vector we transformed it into the bag-of-words representation. We created Term-Document-Matrix (TDM) to transform the NL attribute into a feature vector. However, in order to reduce the number of dimensions, we firstly lemmatized the descriptions using the Morfologik software [5]. As a result of lemmatization we obtained 1 029 unigrams in bag-of-words representation. Then we repeated the experiments described in Section III-A and calculated the performance measures.

[1]http://en.wikipedia.org/wiki/Matthews_correlation_coefficient

C. Classification on Structured Object Description and NL Part

In this experiment we extended the incidents representation by the entire NL part. The difference between this approach and the one presented in Section III-B is that, we are not limited to the object description attribute only. We used the whole description of the incidents stored in the NL part.

As in the case of object description we lemmatized the textual data and created Term Document Matrix. The TDM revealed that the number of unique words equals 1 277. It our opinion the direct representation of NL part throughout term vector is to exhaustive due to the fact that TDM is a sparse matrix. In order to reduce the number of dimensions and increase the separation of the incidents we use the Latent Semantic Analysis method [6]. After conversion of the TDM into LSA space we obtain 85 dimensions. We calculate the number of dimensions finding the first position in the descending sequence of singular values where their sum, divided by the sum of all values, meets or exceeds the 0.5 share value. Next in order to obtain categorical arguments we discretized the attributes of incidents represented in the LSA space. After discretization, each of the LSA attributes had tree values: -1, 0 or 1. Then, we repeated the experiments described in Section III-A.

D. Classification Based on Features of the Threats

In this experiment we change our approach to the labeling task. Instead of training a classifier to recognize which attribute values should the incident have in order to be labeled with a given threat, we try to learn the features of the possible risks. In other words, we learn which features of the incidents most adequately represent the given risk.

The risks are defined as a Cartesian product of threats and objects from the Threats Matrix and are represented by a concatenation of a threat and an abbreviation of an object name (i.e. E1_ME). We decided to utilize the Explicit Semantic Analysis method [7] to devise the new representation of the risks in order to facilitate the labeling.

Explicit Semantic Analysis (ESA) proposed in [7] is a method for automatic tagging of textual data with predefined concepts. It utilizes natural language definitions of concepts from an external knowledge base, such as an encyclopedia or an ontology, which are matched against documents to find the best associations. The definitions of concepts are regarded as a regular collection of texts, with each description treated as a separate document. The model structure imposed by ESA can be interpreted as a one layer neural network [8] with L input nodes corresponding to terms and K output nodes

corresponding to concepts. The associations between terms and concepts have numerical weights. Figure 2 depicts the idea of ESA in a form of a neural network.

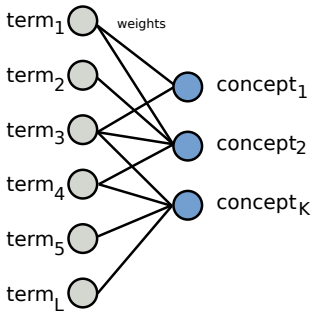


Fig. 2. Representation of the ESA as a neural network.

The implementations of ESA described in [7], [9] use external resources (e.g., an ontology, Wikipedia) which contain the definitions of concepts. In our case there are no external sources of knowledge which can serve as definitions of the risks, therefore we modified the primary idea of ESA and we created *self-defined ESA*. We aggregated the NL parts of incident descriptions from the training set by the assigned risks. In particular, all descriptions of incidents which were labeled with the same risk were concatenated into one document. We repeated this operation for consecutive threats obtaining as many documents as there were threats. Next, we created a Term-Document-Matrix where columns represent the risks and rows represent the terms. The intersection of a column and a row represents an association between a term and a risk. Then, for each incident description in the test set we iterate through the terms, obtain and sum the values of associations to the consecutive risks. The risks for which the sum of the associations is higher than zero constitute a bag-of-risks for a given incident.

For the sake of clarity, we formalize this approach. Let $\mathcal{T}$ be a set consisting of M incidents from the training set, $\mathcal{T} = \{I_1, \ldots, I_i, \ldots, I_M\}$ and $\mathcal{T}'$ is a set consisting of J incidents from test set $\mathcal{T}' = \{I'_1, \ldots, I'_i, \ldots, I'_J\}$. Let $\mathcal{R} = \{r_1, \ldots, r_k, \ldots, r_K\}$ be a set of risks at a fire ground defined as $\mathcal{R} \subseteq \mathcal{H} \times \mathcal{O}$, where $\mathcal{H}$ is a set of the threats from the Threats Matrix and $\mathcal{O}$ is a set of the objects from the Threat Matrix (see Table I). Moreover, let us assume that there were identified L unigrams (e.g. words, stems) $W = \langle w_1, \ldots, w_i, \ldots, w_L \rangle$ from the descriptions of incidents in the training set (NL part of the EWID database). Any incident I_i in the set $\mathcal{T}$ can be represented by a vector of features $I_i = \langle E_i, U_i \rangle$, where $E_i = \langle e_1, \ldots, e_j, \ldots, e_L \rangle \in \mathbb{R}^L$ is a bag-of-words representation of the incident I_i and $U_i = \langle u_1, \ldots, u_k, \ldots, u_K \rangle \in \{0,1\}^K$ represents risks assigned by the experts to the I_i. Each coordinate e_i expresses a value of some relatedness measure for i-th term in vocabulary (w_i), relative to the given E_i. The coordinate $u_k = 1$ if the risk r_k was assigned to the incident I_i by the experts. Respectively, any incident I'_i from the test set is represented by a vector of features $I'_i = \langle E'_i, U'_i \rangle$. However each coordinate of U'_i equals 0 because the incidents from the test set were not labeled by the experts.

We can now define the description of a risk r_k. The description D_{r_k} for risk the r_k is a sum of all bag-of-words representations of incidents in $\mathcal{T}$ where $u_k = 1$

$$D_{r_k} = \sum_{i:u_k=1} E_i \tag{1}$$

Next, all the D_{r_k} from set $\mathcal{D}$ (set of all designations) were converted to the Term-Document-Matrix (TDM), where columns are labeled by descriptions, rows by terms from vocabulary W and the coordinates d_i represent relatedness measure for i-th term in vocabulary (w_i), relative to the given D_{r_k}. The measure used by us to calculate d_i is the *tf-idf* (term frequency-inverse document frequency) index (see [10]) defined as:

$$d_i = tf_{i,k} \times idf_i = \frac{n_{k,i}}{\sum_{j=1}^{L} n_{k,j}} \times \log \frac{|D|}{|\{D_{r_k} : w_i \in D_{r_k}\}|}, \tag{2}$$

where $n_{k,i}$ is the number of occurrences of the term w_i in the description D_{r_k}, $|D|$ is the cardinality of $\mathcal{D}$ which equals K, and $|\{D_{r_k} : w_i \in D_{r_k}\}|$ is the number of the descriptions where the term w_i appears.

Next, the TDM representation of the risks descriptions can be used as an inverted index that maps terms into lists of risks. Each row of TDM expresses an association of the given term from W to K risk descriptions. The inverted index is utilized as a semantic interpreter to assign the risk into incidents from the set $\mathcal{T}'$. Given an incident description E'_i, it iterates over terms from the description, retrieves the corresponding entries and merges them into a weighted vector of risks that represent the given incident.

Let $E'_i = \langle e'_1, \ldots, e'_j, \ldots, e'_L \rangle$ be a bag-of-words representation of the description of incident I'_i from the set $\mathcal{T}'$. Let $inv_{i,k}$ be an inverted index entry for e'_i. It quantifies the strength of association of the term w_i to a risk r_k. For convenience, all the weights $inv_{i,k}$ can be arranged in a sparse matrix structure with L rows and K columns, denoted by INV, such that $INV[i,k] = inv_{i,k}$ for any pair (i,k).

Next we can create a new representation U_i^{INV} of incident I'_i from the test set as a sum of values from the TDM of terms which appear in E'_i:

$$U_k = \sum_{i:e'_i \neq 0} e'_i \times inv_{i,k} = W_i * INV[\cdot, k]. \tag{3}$$

In the above equation $*$ is the standard dot product and $INV[\cdot, k]$ indicates k-th column of the sparse matrix INV. This new representation will be called a *bag-of-risks* of an incident I_i.

The new representation of incident I'_i can be now definded as $I''_i = \langle E'_i, U''_i \rangle$ where U''_i is created as follows $U''_i = \langle u''_1, \ldots, u''_k, \ldots, u''_K \rangle \in \{0,1\}$ $u_k^{INV} > 0 \Rightarrow u''_k = 1$.

For practical reasons it may also be useful to represent the incidents only by the most relevant risks. In such a case, the association weights can be used to rank the risks and to select only the top risks from the ranked list. Therefore, we changed the rule $u_k^{INV} > 0 \Rightarrow u_k'' = 1$ into $u_k^{INV} > var \Rightarrow u_k'' = 1$ where var is some threshold.

IV. Results of the Experiments

In Table III we summarized the results obtained in the experiments. We calculated the performance measures separately for each of the consecutive incidents from the test set, and then we calculated the average for each of the methods. The number of assigned risks for the ESA was set to five with the highest score.

TABLE III
THE PERFORMANCE COMPARISON FOR THE CLASSIFICATIONS METHODS.

Method	Precision	Recall	F1-score
SP	0.68	0.64	0.61
SP-OD	0.45	0.50	0.43
SP-OD-LSA	0.43	0.51	0.43
ESA NLP	0.48	0.70	0.54

The second summarization (see Table IV) compares the performance of different classification methods, according to the risks from Threats Matrix. In this table we also presented the number of the incidents in the training ($\#\mathcal{T}$) and test ($\#\mathcal{T}'$) sets, which were labeled by a given threat.

Figure 3 outlines the comparison of versatility of the methods. We compare the classification methods according to the number of different risks which were at least once properly assigned to an incident.

The practical usefulness and the importance of the results presented in the Tables and Figure 3 is more broadly discussed in Section V.

V. Discussion of the Results

The obtained results revealed that for the maximization of F1-score for a given document we should choose the method which is based on the attribute section only and the classification algorithm (SP). This method achieved very good performance – F1-score reach the value 0.61 (see Table III). However the value of recall is lower than value obtained by ESA NLP approach. That means the best scoring method does not detect the full spectrum of risks.

The intuition is confirmed by the results from Table IV. We observed that if we calculate the F1-score according to the risks, the SP method is classified as the second best. Table IV also reveals the reason for this situation. The SP approach achieves a very good performance for the risks which are assigned very often to the incidents. As an example may serve the risks: A1_ME (86% of incidents labeled in the training set and 88% in the test set), A2_MA (85% and 89%, respectively), A2_ME (88% and 84% of incidents). For these risks the SP method achieves scores 0.86, 0.81 and 0.83, respectively. However, for the risks which were rarely assigned to the incidents the SP methods fails to achieve good performance.

TABLE IV
THE PERFORMANCE COMPARISON (F1-SCORE) OF THE CLASSIFICATIONS METHODS RELATIVE TO THE RISKS. $\#\mathcal{T}$ – NUMBER OF INCIDENTS IN TRAINING SET LABELED BY THE GIVEN RISK, $\#\mathcal{T}'$ – NUMBER OF INCIDENTS IN TEST SET LABELED BY THE GIVEN RISK. THE RISKS ARE DEFINED AS A CARTESIAN PRODUCT OF THREATS AND OBJECTS FROM THE THREATS MATRIX AND ARE REPRESENTED BY A CONCATENATION OF AN ABBREVIATION OF A THREAT AND AN OBJECT NAME (I.E. E1_ME: CALLAPSE_PEOPLE).

Risk	$\#\mathcal{T}$	$\#\mathcal{T}'$	SP	SP-OD	SP-OD-LSA	ESA NL
A1_MA	99	46	0.38	0.49	0.54	0.45
A1_ME	245	107	0.86	0.71	0.69	0.82
A1_T	27	5	–	0.10	0.06	0.07
A2_MA	242	108	0.81	0.64	0.65	0.84
A2_ME	251	103	0.83	0.64	0.70	0.84
A2_S	8	6	0.29	–	0.11	0.22
A2_T	36	8	0.05	0.14	0.06	0.14
A2_U	57	28	0.39	0.37	0.37	0.30
A4_G	3	4	–	–	–	0.08
A4_MA	15	7	0.30	0.32	0.14	0.22
A4_ME	18	9	0.27	0.27	0.15	0.17
A4_S	20	13	–	0.38	0.05	–
A4_T	1	1	–	–	–	0.13
E1_MA	6	3	–	–	0.50	0.11
E1_ME	3	1	–	–	–	–
E2_MA	29	13	0.11	–	0.11	0.31
E2_ME	14	6	–	–	0.13	0.24
E2_S	9	1	–	–	–	0.15
E3_G	3	1	–	–	–	0.12
E3_MA	9	13	–	–	0.10	0.50
E3_ME	2	7	–	–	–	0.12
E4_MA	3	4	–	–	–	–
E4_ME	2	1	–	–	–	–
E4_S	5	2	–	–	–	–
Average F1-score			0.179	0.169	0.181	0.243

Figure 3 compares the versatility of the methods. It depicts the spectrum of risks used by different methods. It illustrates that for the method SP and SP-OD only 10 out of 24 risks could be successfully assigned. The extension of the information by the NL part of the EWID database increases the spectrum of the utilized risks. The SP-OD-LSA method successfully assigned 15 out of 24 risks. However, the most versatile method is ESA NL which is able to properly assign 19 out of 24 risks. We may conclude that the attribute section lacks information related to very rare risks. Only the extension by the NL part allows labeling the incidents with these risks.

The conducted experiments proved that there is a potential in ESA method even for short texts and even in a situation when there are no descriptions available for the concepts in a form of external knowledge base (compare the experiments with long text and an external ontology [7], [9], [11]). However, it should be stated that the descriptions stored in the NL part of EWID database are very specific. In the future work the method should be tested against some more general texts like blogs or news.

The future work should also concentrate on methods for improving ESA. Results of our preliminary experiments suggest that a properly adjusted weights in the inverted index

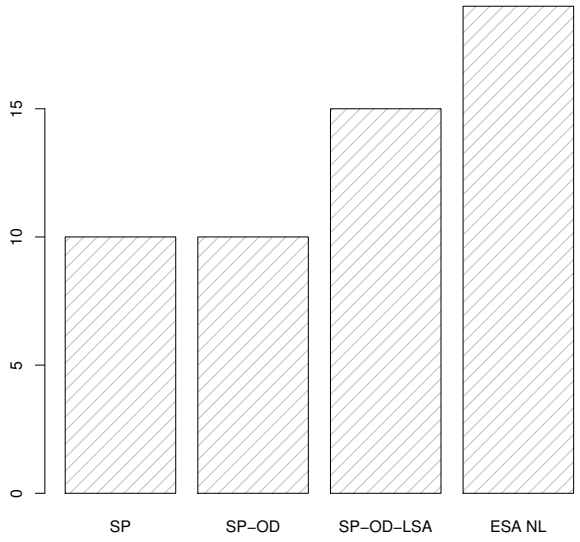


Fig. 3. The comparison of versatility of the methods. Y-axis represents the number of different risks which were at least once properly assigned to an incident.

used by ESA can increase the average F1-score by more than 100%. Therefore, in our research we will focus on finding an appropriate algorithm for updating ESA on for this set.

The experiments also reveal that different risks have different best scoring methods. Therefore, we also consider utilization an ensemble approach in order to assemble a multi-classifier algorithm [12].

ACKNOWLEDGMENTS

The research was supported by the Polish National Centre for Research and Development (NCBiR) - Grant No. O ROB/0010/03/001 in the frame of Defence and Security Programmes and Projects: "Modern engineering tools for decision support for commanders of the State Fire Service of Poland during Fire&Rescue operations in the buildings" and also by the grants N N516 077837 from the Ministry of Science and Higher Education of the Republic of Poland, and 2011/01/B/ST6/03867 from the Polish National Science Centre.

REFERENCES

[1] A. Graeger, U. Cimolino, H. de Vries, and J. Sümersen, *Einsatz- und Abschnittsleitung: Das Einsatz-Führungs-System (EFS)*. Ecomed Sicherheit, 2009.

[2] A. Krasuski, K. Kreński, and S. Łazowy, "A Method for Estimating the Efficiency of Commanding in the State Fire Service of Poland," *Fire Technology*, vol. 48, no. 4, pp. 795–805, 2012.

[3] A. Krasuski, D. Ślęzak, K. Kreński, and S. Łazowy, "Granular Knowledge Discovery Framework," *New Trends in Databases and Information Systems*, pp. 109–118, 2013.

[4] A. Krasuski and P. Wasilewski, "The Detection of Outlying Fire Service's Reports. The FCA Driven Analytics," in *Processings of the 11-th International Conference on Formal Concept Analysis*, 2013, pp. 35–50.

[5] "Morfologik – About the project," http://morfologik.blogspot.com/2006/05/about-project.html.

[6] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.

[7] E. Gabrilovich and S. Markovitch, "Computing semantic relatedness using wikipedia-based explicit semantic analysis," in *Proc. of The 20th Int. Joint Conf. on Artificial Intelligence*, Hyderabad, India, 2007, pp. 1606–1611. [Online]. Available: http://www.cs.technion.ac.il/~shaulm/papers/pdf/Gabrilovich-Markovitch-ijcai2007.pdf

[8] T. M. Mitchell, *Machine Learning*, ser. McGraw Hill series in computer science. McGraw-Hill, 1997.

[9] A. Janusz, W. Świeboda, A. Krasuski, and H. Nguyen, "Interactive Document Indexing Method Based on Explicit Semantic Analysis," *Rough Sets and Current Trends in Computing*, pp. 156–165, 2012.

[10] C. Manning, P. Raghavan, and H. Schütze, "Introduction to information retrieval. 2008," *Online edition*, 2007.

[11] M. Szczuka and A. Janusz, "Semantic clustering of scientific articles using explicit semantic analysis," in *Transactions on Rough Sets XVI*. Springer, 2013, pp. 83–102.

[12] K. Kurach, K. Pawłowski, Ł. Romaszko, M. Tatjewski, A. Janusz, and H. S. Nguyen, "An ensemble approach to multi-label classification of textual data," in *Advanced Data Mining and Applications*. Springer, 2012, pp. 306–317.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 83–89

Combining One-Class Support Vector Machines for Microarray Classification

Bartosz Krawczyk
Department of Systems and Computer Networks,
Wrocław University of Technology
Wrocław, Poland
bartosz.krawczyk@pwr.wroc.pl

***Abstract*—The advance of high-throughput techniques, such as gene microarrays and protein chips have a major impact on contemporary biology and medicine. Due to the high-dimensionality and complexity of the data, it is impossible to analyze it manually. Therefore machine learning techniques play an important role in dealing with such data. In this paper we propose to use a one-class approach to classifying microarrays. Unlike canonical classifiers, these models rely only on objects coming from single class distributions. They distinguish observations coming from the given class from any other possible states of the object, that were unseen during the classification step. While having less information to dichotomize between classes, one-class models can easily learn the specific properties of a given dataset and are robust to difficulties embedded in the nature of the data. We show, that using one-class support vector machines can give as good results as canonical multi-class classifiers, while allowing to deal with imbalanced distribution and unexpected noise in the data. To cope with high dimensionality of the feature space, we propose to form an ensemble, based on Random Subspace and prune it with the usage of diversity measure. Experimental investigations, carried on public datasets, prove the usefulness of the proposed approach.**

***Index Terms*—machine learning, one-class classification, multiple classifier systems, classifier ensembles, bioinformatics, microarray analysis, high dimensionality.**

I. Introduction

CONTEMPORARY high-throughput technologies produce massive volumes of biomedical data. Transcriptional research and profiling, with the usage of microarray technologies are powerful tools to gain a deep insight into the pathogenesis of complex diseases that plague modern society, such as cancer. Recent works on cancer profiling showed without a doubt, that gene expression patters can be used for high-quality cancer subtype recognition [1] - leukemias [2], melanoma [3], breast cancer [4] or prostate cancer [5] to name a few.

Identifying cancer properties, based on their distinct expression profiles may provide necessary information for a breakthrough, that is required for patient-tailored therapy. Currently there are no distinct rules on how individuals respond to chemotherapy and existing chemotherapies have in most cases severe side-effects with varying medical efficiency.

Due to massive amounts of data generated by microarray experiments and their high complexity and dimensionality, one requires a decision support system to extract the meaningful information from them. Machine learning is widely used for this task [6], with two distinct areas - unsupervised [7] and supervised learning [8]. In this paper we will focus on the latter one.

Supervised machine learning is a promising approach for analyzing microarray results in context of predicting patients outcome. Support Vector Machines are among the most popular classifiers used for this task [9]. Multiple Classifier Systems [10], or classifier ensembles, have gained an significant attention of the bioinformatics community in recent years. Random Forest [11] and Rotation Forest [12] ensembles have displayed an excellent classification accuracy for small-sample, high dimensionality microarray datasets, outperforming single-model approaches.

Another important issue is the problem of curse of dimensionality. Microarray data suffer from a relatively small number of objects, in comparison to the feature space dimensionality, often reaching several thousands. This causes difficulties for machine learning algorithms, reducing their performance and increasing their computational complexity. Among this data flood a major number of parameters possess small discriminative power and is irrelevant to the classification process, which makes feature selection a crucial step in microarray analysis [13].

Although there are many applications of machine learning-based decision support systems in bioinformatics, there are still many unresolved problems, such as:

- How to integrate heterogeneous data sources to achieve better insight into the mechanism behind complex diseases?
- How to organize, store, analyze and visualize high-dimensionality data obtained from the biomedical data flood?
- How to deal with the problem of high-dimensionality, small sample size, which strongly affects the classification performance and may lead to overfitting, poor generalization and unstable predictors?
- How to cope with difficulties embedded in the nature of microarray data, such as noise or class imbalance, as canonical machine learning classifiers cannot cope with them easily?

In this paper the last two issues are addressed.

978-83-60810-53-8/$25.00

We propose to analyze microarray data with the usage of one-class classifiers, instead of commonly applied binary ones. Up to author's knowledge this is the first work on applying one-class ensembles and one-class classification in general, to the microarray classification.

To cope with the high dimensionality problem we apply an ensemble approach, based on Random Subspaces [14]. By decomposing the feature space we at the same time reduce the overall computational complexity of the classification model and assure initial diversity among the pool of individual classifiers in the committee. A diversity-based pruning method is applied to discard redundant classifiers and to chose mutually complementary one-class predictors. Experiments, based on a set of publicly available microarray datasets, show that the proposed approach maintains a good classification accuracy, while displaying an improved robustness to atypical data distribution and prevalent noise.

II. One-class classification

The aim of one-class classification (OCC) is to recognize one specific class from the more broad set of classes (e.g., selecting horses from all animals). The given class is known as target class ω_t, while the remaining are denoted as outliers ω_O. During the learning only examples target class (known also as positive examples) are being presented to learner, while it is assumed that during the exploitation phase new, unseen objects from other classes may appear.

OCC problems are common in the real world where positive examples are widely available but negative ones are hard, expensive or even impossible to gather [15]. Let us consider an engine. It is a quite easy and cheap to collect data about its normal work. Collecting observations about failures it is expensive and sometimes impossible, because in this case we would have to spoil the engine.

Such approach is very useful as well for many practical cases especially when the target class is "stable" and outlier one is "unstable". To explain this motivation let us consider a computer security problem as spam filtering or intrusion detection (IDS/IPS) [16].

Among several types of classifiers dedicated to OCC, the most popular is one concentrating on estimation of a closed boundary for given data, assuming that such a boundary will describe sufficiently the target class [17]. The main aim of those methods is to find the optimal size of the volume enclosing given training points. Too small size could lead to overfitting the model, while too big size might lead to extensive acceptance of outliers into the target class. Those methods rely strongly on the distance between objects [18]. Boundary methods require smaller number of objects to properly estimate the decision criterion, which makes them a perfect tool for applications suffering from a small sample size,such as microarrays classification. The well-known boundary methods are one-class support vector machine (OCSVM) [19] and support vector data description (SVDD) [20]. In this work we will use the former one.

A. One-class support vector machine

One-class SVM classifier (OCSVM) [19] can deal with datasets containing only patterns from one target class. OCSVM classification aims at discriminating one class of target samples from all other ones. It consists of learning the minimum volume contour that encloses most of the data in a given dataset. Its original application is the outlier detection finding data that differ from most of the data within a dataset.

Let $\chi = \{x_1, x_2, ..., x_m\}$ be a given dataset in $\mathbb{R}^d$. Each x_j is a feature vector describing an object. OCSVM use the training data to learn a function $f_\chi : \mathbb{R}^d \mapsto \mathbb{R}$ such that most of the data in χ belong to the set $\mathcal{R}_\chi = \{x \in \mathbb{R}^d; f_\chi(x \geqslant 0)\}$ while the volume of $\mathcal{R}_\chi$ is minimal. This problem is known as $MinimalVolumeSet$ (MVS) estimation. Membership of x to $\mathcal{R}_\chi$ indicates whether this estimated volume is overall similar to χ or not. Therefore when considering a M-class recognition problem we have to learn M membership functions f_{χ_i} - one for each class.

OCSVM uses the following approach to estimate the MVS. A kernel function $k(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$. In our research we use a Gaussian Radial Basis Function(RBF) kernel :

$$k(x, x') = exp[- \parallel x - x' \parallel^2 /2\sigma^2], \tag{1}$$

where x' is the object after mapping to a hypersphere, $\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^d$. The kernel induces a new, artificial feature space $\mathcal{H}$ by the usage of mapping $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ dened by $\phi(x) \triangleq k(x, \cdot)$. It has been shown that $\mathcal{H}$ reproduces kernel Hilbert spaces of given functions, with dot product denoted as $\langle \cdot, \cdot \rangle_\mathcal{H}$. The reproducing kernel property implies that:

$$\langle \phi(x), \phi(x') \rangle_\mathcal{H} = \langle k(x, \cdot), k(x', \cdot) \rangle_\mathcal{H} = k(x, x'), \tag{2}$$

which makes the evaluation of $k(x, x')$ a linear operation in $\mathcal{H}$, while it is a nonlinear operation in $\mathbb{R}^d$.

Considering the RBF:

$$\parallel \phi(x) \parallel^2_\mathcal{H} \triangleq \langle \phi(x), \phi(x) \rangle_\mathcal{H} = k(x, x) = 1. \tag{3}$$

From this one may assume that all the data mapped into $\mathcal{H}$ are located on the hypersphere with radius equal to one, centered onto the origin of $\mathcal{H}$, which is denoted $S_{(o,R=1)}$. The OCSVM determines in $\mathcal{H}$ the hyperplane $\mathcal{W}$ that separates most of the data from the $S_{(o,R=1)}$, while at the same time maximizing the distance from it. This practically implements the solution to the MVS estimation problem.

Let:

$$\mathcal{W} = \{h(\cdot) \in \mathcal{H}; \langle h(\cdot), w(\cdot) \rangle_\chi - \rho = 0\}, \tag{4}$$

where parameters $w(\cdot)$ and ρ are the results of the following optimization problem

$$\min_{w,\xi,\rho} \frac{1}{2} \parallel w(\cdot) \parallel^2_\mathcal{H} + \frac{1}{vm} \sum_{j=1}^{m} \xi_j - \rho, \tag{5}$$

subject to (for $j = 1, ..., m$)

$$\langle w(\cdot), k(x_j, \cdot) \rangle_\chi \geqslant \rho - \xi_j, \tag{6}$$

where $\xi_j \geqslant 0$, v is a control parameter for the fraction of the data that are allowed to be located on the wrong side of the $\mathcal{W}$ (outliers which do not belong to the $\mathcal{R}_\chi$) and ξ_j are slack variables.

It can be shown that a solution to Eq. (5,6) can be expressed by the following:

$$w(\cdot) = \sum_{j=1}^{m} \alpha_j k(x_j, \cdot), \tag{7}$$

where α_j comes from the dual optimization problem

$$\min_{\alpha} \frac{1}{2} \sum_{j,j'=1}^{m} \alpha_j \alpha_{j'} k(x_j, x_{j'}), \tag{8}$$

subject to $0 \leq \alpha_j \leq \frac{1}{vm}, \sum_j \alpha_j = 1$.

The OCSVM decision function $f_\chi(x)$ is given as follows:

$$f_\chi(x) = \sum_{j}^{m} \alpha_j k(x_j, x) - \rho, \tag{9}$$

where the value of ρ is calculated from knowing that $f_\chi(x_j) = 0$ for those $x_j \in \chi$ that verify both $\alpha_j \neq 0$ and $\alpha_j \neq \frac{1}{vm}$. Objects from χ that satisfies those conditions are located onto a decision boundary.

III. Proposed approach

In this paper we propose to use a one-class classification approach to microarray analysis. Let us list the main features and advantages of the proposed approach:

1) We utilize one of the classes as the target concept ω_T and the remaining one as outliers. In case of imbalanced dataset the minority class is considered as the target concept, while in case of balanced distributions we chose the more numerous class. This is motivated by the fact, that while sacrificing the additional information about the second class, we gain a classifier that is able to adjust itself to the specificity of the given class and is more robust to difficulties that may be encountered, such as class imbalance or in-class noise.
2) The high dimensionality of the feature space is difficult to handle for one-class boundary classifiers. It significantly increases their complexity, the training and execution times and lead to a much more difficult optimization task (and hence to a degradation of the recognition quality). To cope with this problem we propose to use a Random Subspace ensemble to decompose the feature space into smaller competence areas and build an ensemble of simpler one-class models.
3) As Random Subspace may lead to creation of similar classifiers, or classifiers with low discriminative power, a pruning procedure is beneficial, as it may discard irrelevant predictors. We use a diversity-based method, which uses a criterion optimized for OCC task.

A. Dealing with the high dimensionality problem

The one-class boundary compute a distance between the object x and the estimated boundary, which encloses the target class ω_T. This allows to apply fusion methods, that are based on the discrete output (returned class label) of the individual classifiers - such as the voting methods. However, to apply more sophisticated fusion methods, which assume the continuous outputs of each of the individuals, the support of an object x for a given class is required.

We propose to use the following heuristic support function produced on the basis of a distance:

$$F(x, \omega_T) = \frac{1}{c_1} exp(-d(x|\omega_T)/c_2), \tag{10}$$

which models a Gaussian distribution around the classifier, where $d(x|\omega_T)$ is a distance (Euclidean distance is used) from the evaluated object to the support vectors describing the target concept, c_1 is the normalization constant and c_2 is the scale parameter. Parameters c_1 and c_2 should be fitted to the target class distribution.

Estimating this mapping for high dimension is very complex and requires a significant computational power and time. To cope with this difficulty we propose to use a Random Subspace method to partition the dataset into many subspaces of smaller dimensionality. Each base classifier is trained on a new subset, which is highly smaller than the original feature space size. This boosts the training time, while applying ensemble principles makes sure that despite using weaker predictors, we still get a satisfying accuracy [21].

B. Pruning the ensemble

As Random Subspace may produce classifiers of different level of individual quality and diversity, a classifier section step is most beneficial to forming an one-class ensemble. Multiple Classifier Systems, in order to work properly, must consist of predictors of at the same time high accuracy and diversity. Only mutually complementary classifiers may lead to an improvement over using a single-model approach. Diversity is one of the most popular measures for this task. It may be applied to one-class classifiers, but after modifications, that take into consideration the nature of the OCC problem [22]. For this application an one-class entropy measure [23] is used.

Let's assume that the highest ensemble diversity for a given object $x_j \in X$ is displayed by $[R/2]$ of the ensemble votes with the same value (ω_T or ω_O) and remaining R - $[R/2]$ with the other value. If all votes returned identical response the ensemble cannot be considered as a diverse one. Let us denote by $r(x_j)$ the number of one-class classifiers that correctly recognize the object x_j. Assuming there are N objects in the training set, one may use entropy to measure the diversity using the presented concept:

$$E_{oc}(\Pi^r) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{(R - [R/2])} \min\{r(x_j), R - r(x_j)\}. \tag{11}$$

where Π^r is the considered pool of classifiers.

TABLE I
STATISTICS OF THE DATASETS USED IN THE EXPERIMENTS.

dataset	samples (class 1 / class 2)	features
Breast Cancer	78 (34 / 44)	24481
Breast Cancer - noise	78 (34 / 44)	24481
Central Nervous System	60 (21 / 39)	7129
Colon Tumor	62 (22 / 40)	6500
Lung Cancer	181 (31 / 150)	12533

This is a non-pairwise (global) diversity measure, which take values from [0,1]. 0 corresponds to identical ensemble and 1 corresponds to the highest possible diversity.

C. Fusion method

As a fusion method we use a one-class mean vote, which combines binary output labels of one-class classifiers. It can be written as:

$$y_{mv}(x) = \frac{1}{L}\sum_{k}[(P_k(x|\omega_T) \geq \theta_k)], \qquad (12)$$

where $[(\cdot)]$ is the *Iverson brackets* and θ_k is threshold for the target class. When a threshold equal to 0.5 is applied this rule transforms into a majority vote for binary problems.

IV. EXPERIMENTAL INVESTIGATIONS

In this section we evaluate the proposed one-class ensemble on the basis of datasets available at [1], whose details are given in Table I. Four different datasets were used and additional, fifth one, was generated. It was based on the Breast Cancer dataset. To test the performance of classifiers in difficult scenarios we have affected 25% of objects with Gaussian noise, thus creating in-class outliers in the data.

As base classifier we have used an OCSVM with RBF kernel [24].

To put the obtained results into context we have tested the performance of multi-class classifiers used for this task - single SVM (trained with RBF kernel and SMO procedure), Random Forest (consisting of 100 decision trees) and Rotation Forest (consisting of 100 decision trees). Additionally we show the performance of a single OCSVM and the proposed ensemble without the pruning step.

Results are based on leave-one-out cross-validation (LOOCV).

All experiments were carried out in the R environment [25], with classification algorithms taken from the dedicated packages, thus ensuring that the results achieved the best possible efficiency and that the performance was not decreased by a bad implementation. The Friedman ranking test [26] was done for comparison over multiple benchmark datasets.

Firstly the parameters for the proposed pruned one-class ensemble are examined. We test the correlations between the accuracy and size of the subspaces / number of classifiers in the pool. For analyzing the optimal number of the classifiers, a subspace size equal to 0.2 was used. Then, when the size was selected, the subspace size parameter was investigated.

[1] http://datam.i2r.a-star.edu.sg/datasets/krbd/

One should note that these results are prior to the pruning phase - which firther improves the accuracy while reducing the number of classifiers in the pool. Results are presented in Fig 1 - 5.

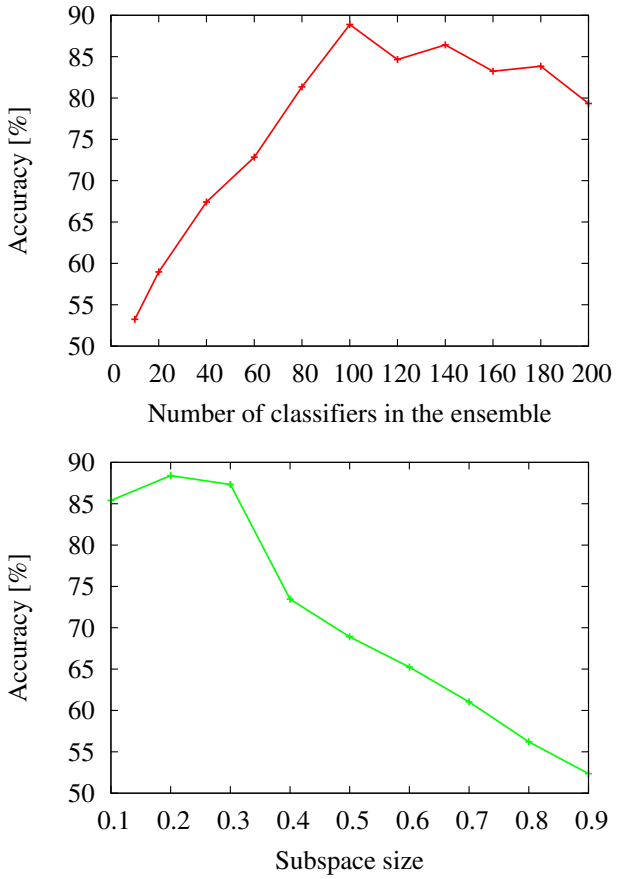


Fig. 1. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Breast Cancer dataset.

The established optimal settings are then used for the second stage of the experimental investigation - comparison with other classification methods. Results with respect to sensitivity and specificity, are given in Tab. II.

Analyzing the results of parameter settings for one-class models shows us, that there are some common properties regardless the analyzed dataset. The optimal size of the ensemble was around 100-120 classifiers, built on a small subspaces (consisting of 10% - 20% of features). This allowed to maintain high diversity of the ensemble and allowed for a pruning procedure to select valuable classifiers with mutually complementary areas of competence. Additionally smaller size of the subspaces allowed for training less complex OCSVMs, which in turn prevented them for too overfitted decision boundary.

From the results one may clearly see, that in case of standard microarray datasets the proposed approach returns both specificity and sensitivity similar to those of the state-of-

TABLE II
RECOGNITION SENSITIVITY [%] AND SPECIFICITY [%] FOR EXAMINED METHODS. $RandF$ STANDS FOR RANDOM FOREST, $RotF$ FOR ROTATION FOREST, $OCCE$ FOR AN ONE-CLASS ENSEMBLE WITHOUT PRUNING AND $POCCE$ FOR THE PROPOSED PRUNED ONE-CLASS ENSEMBLE. AVERAGE RANK OF TESTED CLASSIFIERS, ACCORDING TO FRIEDMAN RANKING TEST, ARE GIVEN AT THE BOTTOM.

Dataset	SVM		RandF		RotF		OCSVM		OCCE		POCCE	
	Sens [%]	Spec[%]	Sens [%]	Spec[%]	Sens [%]	Spec[%]	Sens [%]	Spec[%]	Sens [%]	Spec[%]	Sens [%]	Spec[%]
Breast Cancer	90.23	91.46	92.32	93.65	92.32	93.65	87.85	90.07	88.11	90.86	92.89	92.70
Breast Cancer - noise	74.46	83.59	77.36	84.90	80.05	85.72	75.20	82.98	80.95	85.20	89.09	90.05
Central Nervous System	85.60	94.36	88.20	95.90	88.20	95.90	82.95	90.11	84.07	92.01	87.84	93.96
Colon Tumor	78.90	91.25	81.35	94.03	82.70	93.90	80.15	92.36	83.85	93.05	84.05	93.83
Lung Cancer	61.72	93.05	65.89	95.11	67.00	94.85	69.22	92.08	70.98	93.90	74.61	94.78
Avg. score	4.85		2.90		2.25		5.21		4.11		1.68	

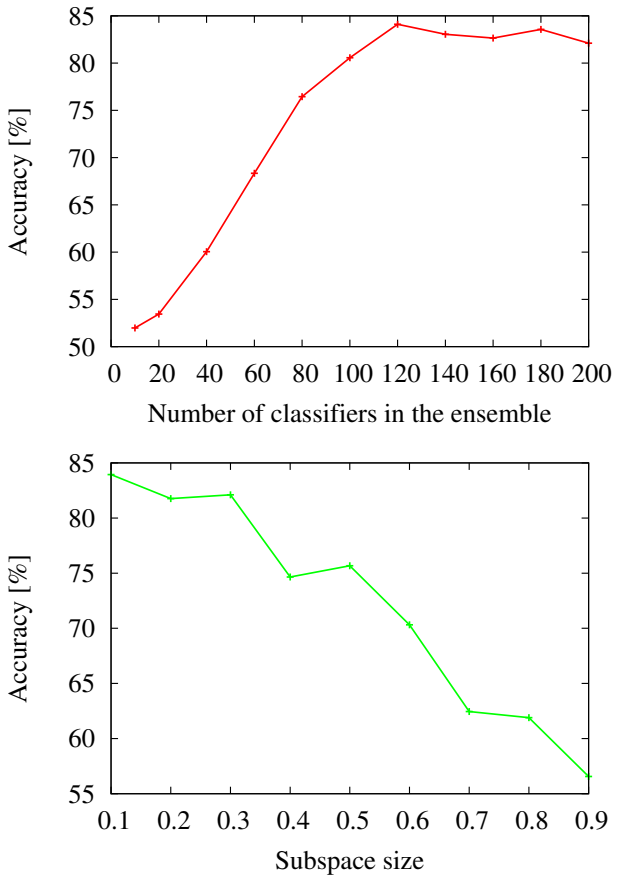


Fig. 2. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Breast Cancer - noise dataset.

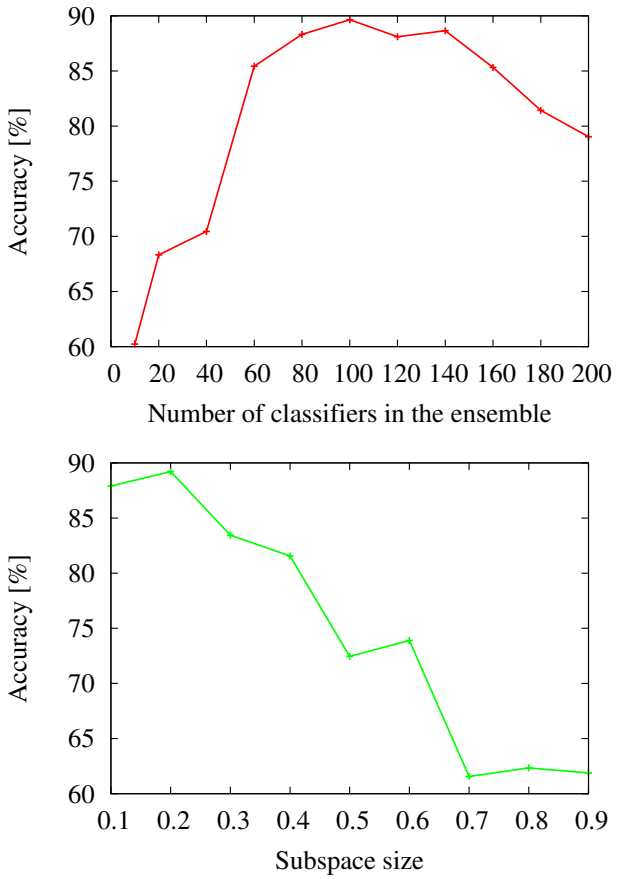


Fig. 3. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Central Nervous System dataset.

the-art multi-class models. However in case of noisy (dataset no. 2) and imbalanced (datasets no. 4 and no. 5) our proposed approach is able to outperform significantly the standard classifiers. This happens due to the nature of OCC models - as they are able to learn the distinct properties of the target class, they are able to cope with in-class difficulties.

V. CONCLUSIONS

In this paper a novel approach for microarray analysis, based on an ensemble of one-class support vector machines, was presented. To deal with the problem of high dimensionality, which may cause difficulties for one-class model, a Random Subspace method was applied. This, combined with a diversity-based pruning step, allowed for an effective classifier, returning similar performance as state-of-the-art multi-class methods. The strong points of the proposed method were revealed when dealing with noisy and imbalanced data. In such a case the proposed combined one-class classifier displayed superior quality over its competitors.

The proposed approach may be an attractive tool for bioinformatics decision support systems, in which we deal with uncertain, noisy data or data coming from uneven distributions.

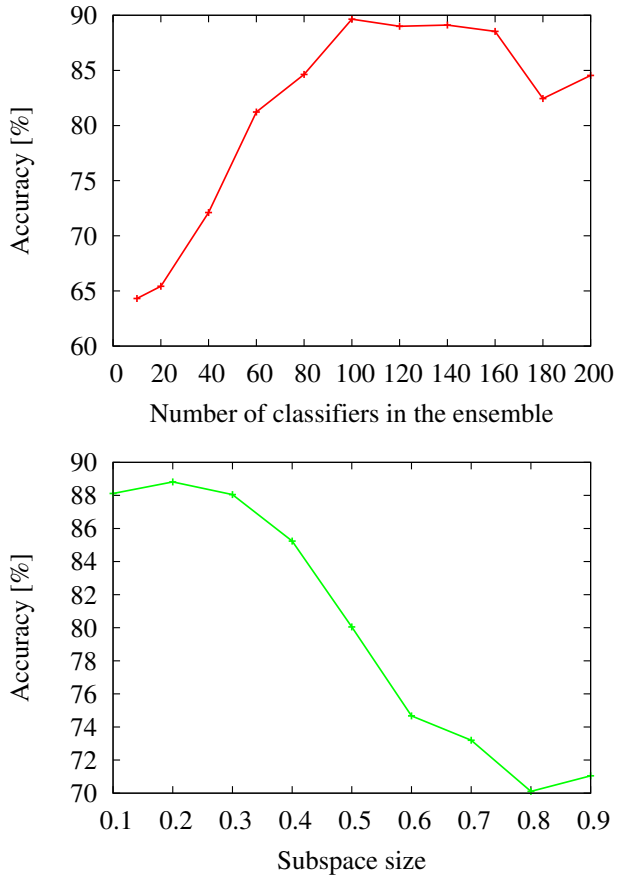


Fig. 4. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Colon Tumor dataset.

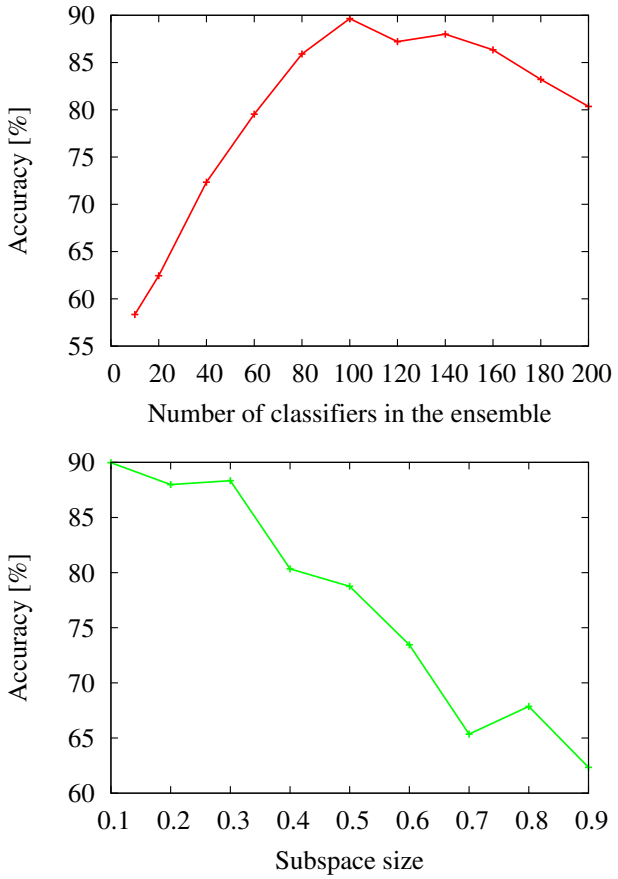


Fig. 5. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Lung Cancer dataset.

VI. Acknowledgment

The work was supported by the fellowship co-financed by European Union within European Social Fund.

References

[1] A. V. Tinker, A. Boussioutas, and D. D. L. Bowtell, "The challenges of gene expression microarrays for the study of human cancer," *Cancer Cell*, vol. 9, no. 5, pp. 333–339, 2006.

[2] V. S. Silveira, C. A. Scrideli, D. A. Moreno, J. A. Yunes, R. G. P. Queiroz, S. C. Toledo, M. L. M. Lee, A. S. Petrilli, S. R. Brandalise, and L. G. Tone, "Gene expression pattern contributing to prognostic factors in childhood acute lymphoblastic leukemia," *Leukemia and Lymphoma*, vol. 54, no. 2, pp. 310–314, 2013.

[3] T. Schatton, G. F. Murphy, N. Y. Frank, K. Yamaura, A. M. Waaga-Gasser, M. Gasser, Q. Zhan, S. Jordan, L. M. Duncan, C. Weishaupt, R. C. Fuhlbrigge, T. S. Kupper, M. H. Sayegh, and M. H. Frank, "Identification of cells initiating human melanomas," *Nature*, vol. 451, no. 7176, pp. 345–349, 2008.

[4] G. Finak, N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, and M. Park, "Stromal gene expression predicts clinical outcome in breast cancer," *Nature medicine*, vol. 14, no. 5, pp. 518–527, 2008.

[5] C. C. Lynch, A. Hikosaka, H. B. Acuff, M. D. Martin, N. Kawai, R. K. Singh, T. C. Vargo-Gogola, J. L. Begtrup, T. E. Peterson, B. Fingleton, T. Shirai, L. M. Matrisian, and M. Futakuchi, "Mmp-7 promotes prostate cancer-induced osteolysis via the solubilization of rankl," *Cancer Cell*, vol. 7, no. 5, pp. 485–496, 2005.

[6] P. Larranaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J. A. Lozano, R. Armananzas, G. Santaf, A. Perez, and V. Robles, "Machine learning in bioinformatics," *Briefings in Bioinformatics*, vol. 7, no. 1, pp. 86–112, 2006.

[7] Y. . Wang, Z. . Yu, and V. Anh, "Fuzzy c-means method with empirical mode decomposition for clustering microarray data," *International Journal of Data Mining and Bioinformatics*, vol. 7, no. 2, pp. 103–117, 2013.

[8] M. Ringner, C. Peterson, and J. Khan, "Analyzing array data using supervised methods," *Pharmacogenomics*, vol. 3, no. 3, pp. 403–415, 2002, cited By (since 1996):43. [Online]. Available: www.scopus.com

[9] D. Bariamis, D. Maroulis, and D. K. Iakovidis, "Unsupervised svm-based gridding for dna microarray images," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 418–425, 2010.

[10] M. Woźniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, 2013, article in Press.

[11] K. Moorthy and M. S. Mohamad, "Random forest for gene selection and microarray data classification," ser. Communications in Computer and Information Science, vol. 295 CCIS, 2012, pp. 174–183.

[12] K. Liu and D. Huang, "Cancer classification using rotation forest," *Computers in biology and medicine*, vol. 38, no. 5, pp. 601–610, 2008.

[13] I. Inza, P. Larraaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains," *Artificial Intelligence in Medicine*, vol. 31, no. 2, pp. 91–103, 2004.

[14] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 832–844, August 1998.

[15] B. Cyganek, "Image segmentation with a hybrid ensemble of one-class support vector machines," ser. Lecture Notes in Computer Science

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6076 LNAI, no. PART 1, pp. 254–261.

[16] K. Noto, C. Brodley, and D. Slonim, "Frac: A feature-modeling approach for semi-supervised and unsupervised anomaly detection," *Data Mining and Knowledge Discovery*, vol. 25, no. 1, pp. 109–133, 2012.

[17] D. M. J. Tax, P. Juszczak, E. Pekalska, and R. P. W. Duin, "Outlier detection using ball descriptions with adjustable metric," in *Proceedings of the 2006 joint IAPR international conference on Structural, Syntactic, and Statistical Pattern Recognition*, ser. SSPR'06/SPR'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 587–595.

[18] B. Krawczyk and M. Woźniak, "Experiments on distance measures for combining one-class classifiers," in *2012 Federated Conference on Computer Science and Information Systems, FedCSIS 2012*, 2012, pp. 89–92.

[19] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, ser. Adaptive computation and machine learning. MIT Press, 2002.

[20] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[21] T. Wilk and M. Woźniak, "Soft computing methods applied to combination of one-class classifiers," *Neurocomput.*, vol. 75, pp. 185–193, Jan. 2012.

[22] B. Krawczyk and M. Woźniak, "Accuracy and diversity in classifier selection for one-class classification ensembles," in *2013 IEEE Symposium Series on Computational Intelligence, Symposium on Computational Intelligence and Ensemble Learning CIEL*, 2013, pp. 46–51.

[23] B. Krawczyk, "Diversity in ensembles for one-class classification," in *New Trends in Databases and Information Systems*, ser. Advances in Intelligent Systems and Computing, M. Pechenizkiy and M. Wojciechowski, Eds. Springer Berlin Heidelberg, 2012, vol. 185, pp. 119–129.

[24] B. Cyganek, "One-class support vector ensembles for image segmentation and classification," *Journal of Mathematical Imaging and Vision*, vol. 42, no. 2-3, pp. 103–117, 2012.

[25] R. D. C. Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008.

[26] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 91–98

Flow–level Spam Modelling using separate data sources

Marcin Luckner
*Faculty of Mathematics and Information Science
Warsaw University of Technology
pl. Politechniki 1, 00-661 Warszawa, Poland
Email: mluckner@mini.pw.edu.pl

Robert Filasiak
†Orange Labs Poland
ul. Obrzeżna 7, 02-691 Warszawa, Poland
Email: Robert.Filasiak@orange.com

***Abstract*—Spam detection based on flow–level statistics is a new approach in anti–spam techniques. The approach reduces number of collected data but still can obtain relative good results in a spam detection task. The main problems in the approach are selection of flow–level features that describe spam and detection of discrimination rules. In this work, flow–level model of spam is presented. The model describes spam subclasses and brings information about major features of a spam detection task. The model is the base for decision trees that detect spam. The analysis of detectors, which was learned from data collected from different mail servers, results in the universal spam description consists of the most significant features. Flows described by selected features and collected on Broadband Remote Access Server were analysed by an ensemble of created classifiers. The ensemble detected major sources of spam among senders IP addresses.**

***Index Terms*—Spam detection, Flow analysis, Anomaly detection**

I. Introduction

RAPID development of the Internet and associated services induced growth of the desired bandwidth for their execution. Customers expect their Internet Service Providers (IS) to provide a flexible, fully secured access to the Internet. Requirements related to privacy and confidentiality increasingly important. The political environment inside European Union is discussing adjustment of law regulations to market needs. ISPs have to consider Quality of Service (QoS), security, Service Level Agreement (SLA) among others committed to privacy guarantee. This is one of the reasons for development of methods used for monitoring and analysis of traffic in the ISP's core network.

Two approaches are interesting from the perspective of multi–gigabit stream: packet header analysis and flow analysis. Both techniques do not use information contained in payload, which is very important from data reduction and privacy point of view. Additionally, the hash function can be executed on IP addresses to guarantee higher data protection. In some cases, data does not need to be stored. A good example is the statistical detection of Distributed Denial of Service (DDoS) attacks [1] and solutions developed on Field Programmable Gate Array (FPGA) cards [2]. Moreover, such limited data were successful used in an Internet traffic classification task [3].

The packet header analysis is focused processing of headers. The flow analysis is focused on sets of headers determined by a source, a destination IP, source and destination port, timestamps, etc. depending on parameters used to define a flow. The flow analysis enables more compact data reduction. IP Flow Information Export (IPFIX) [4] and NetFlow [5] are well known standards. The equivalent given by Juniper is named J-flow.

Regardless of what was an object of analysis (headers or flows), methods are based on a language that describes analysed events. Usually components of the language are values from packet headers, statistical values such as the average time, the maximal, or the minimal size of packets. Wide lists of defined primitives for flow analysis and honeypot detection are given in [6] and [7] respectively.

In all cases, the following schema is used:

1) select observed parameters $[(m_1, \ldots, m_n) \subset \{M\}]$ (metric),
2) capture values for the parameters $[m_1 \leftarrow .09, m_2 \leftarrow 11, \ldots, m_n \leftarrow false]$ (measurement),
3) calculate features $[f_1 \leftarrow m_1 * m_4, f_2 \leftarrow m_3 + m_5, \ldots]$(features),
4) determine logical relationships between features [if $(f_1 > f_2)$ then action$_1$ else action$_2$](decision).

One of the most important issues in the presented schema is a selection of features. The features that describe spam can be used to create its detection rules. In this paper, flow–level parameters $\{m\}$ selected by Žádník [8] as a subset of the set $\{M\}$ defined in [6] are used create a primary model of spam. In collected spam records, well–separate subclasses are detected (Section II-A). The comparison of subclasses defines important discriminants (Section II-B) that can be used to determine separation rules between spam and the rest of the flows (Section II-C).

Created model was trained and tested on separate data but collected from the same mail server. Therefore, a new data set was created and the most significant features were calculated once again on new data (Section III-A). Both sets of features were compared (Section III-B) and the comparison resulted in a universal set of features (Section III-C).

The final model was checked with Broadband Remote Access Server (BRAS) data (Section IV-A). Decision trees created on the base of both learning sets were used to detect

978-83-60810-53-8/$25.00

spam among BRAS data (Section IV-B). The classification resulted in detection of main sources of spam (Section IV-C).

Conclusions from all tests are presented in Section V.

II. Spam model

The spam model is based on flows collected by Žádník and Michlovský [8]. The flows are defined by by NetFlow protocol that contains:

- source IP address,
- destination IP address,
- IP protocol,
- source port,
- destination port,
- IP type of service.

The NetFlow version 9 allows the user to collect additional features. The features collected in the flows are a subset of features presented in [6].

The authors collected data from the SMTP server hosting mailboxes the Liberouter project group. The data set contains over 58 thousand records described by 64 features and divided into several classes. Among classes, two describe spam. The first class *dnsbl* contains flows from IP address mentioned on DNS black lists. The second class *y_spam* consists of flows that have been successfully received and marked as spam by SpamAssassin.

In the case of the *dnsbl* class flows were labelled because of a source IP address. All flows send from the denied addresses are labelled as spam. For the second class *y_spam*, the labelling process is more complex. The flows are labelled in the off–line mode. The IP addresses and the time of arrival for the flow are compared with the SpamAssassin logs. If a mail with the same source IP address and the destination IP address was marked as spam then all flows with a similar time of arrival are labelled as spam.

The described division of spam is a consequence of used methodology and cannot be used as a framework of the model without any doubt. In the following section, the statistical methodology that creates spam subclasses is presented.

A. Detection of subclasses

Spam subclasses are created in two steps. Firstly, a clustering method is used to detect inner clusters. Next, a decision tree is created to find discrimination rules.

1) Clustering: The predefined spam subclasses *dnsbl* and *y_spam* are a consequence of used methodology. It should not be assumed that this division has a statistical base. Therefore, a new division is created using k–means clustering [9].

In the analysis, all members of *dnsbl* and *y_spam* are treated as a single class *spam*. The class consists of almost 54 thousand records.

A number of spam subclasses is unknown. Therefore, various values of k coefficient from the range $k \in [2, 25]$ are tested. The results of subsequent tests are compared in v–fold cross–validation process. In such test, random samples are drawn v times. Summary indices of the accuracy of the prediction are computed over the v replications. In the described case, the value v is fixed as 10.

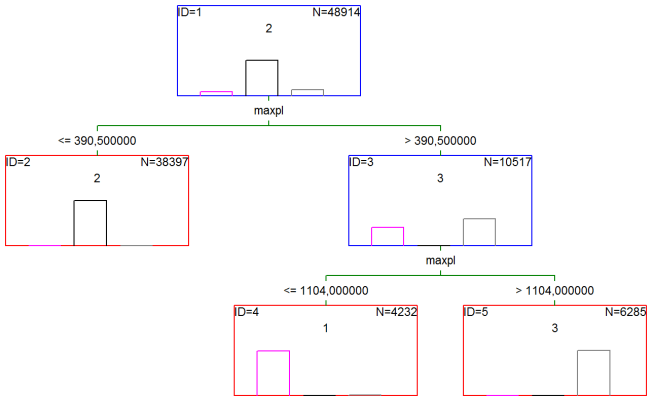


Fig. 1. The classification tree that divides spam into subclasses

A classification error calculated among cross–validation probes was smallest for division into four classes. In such case, the obtained error was 8 percent. However, one of the created classes has a relatively small cardinal number (632 records, which is about 1.3 percent of spam).

Significant differences between cardinal numbers of classes determine *a priori* probability used in a classification task. Members of classes with small *a priori* probability are classified as members of numerous classes to reduce the risk of misclassification.

To avoid future problems the number of classes was reduced to three. The cross–validation classification error increased over 8 percent, which is still an acceptable level. The smallest class was eliminated and the new distribution is more reasonable. The biggest class 2 contains 78.5 percent of record when classes 3 and 1 contain 12.9 and 8.6 percent respectively.

2) Separation rules: The second step of modelling creates discrimination rules between subclasses created during the clustering process. This task is done using a C&RT tree [10]. Such tree is not a very advanced classifier but creates clear decision rules.

The classification accuracy was over 99 percent. That proves a good division of spam into subclasses. Detailed information about classification errors is given in Table I.

TABLE I
THE MISCLASSIFICATION MATRIX FOR SUBCLASSES OF SPAM

	Observed 1	Observed 2	Observed 3
Predicted 1	4206	2	24
Predicted 2	1	38396	0
Predicted 3	6		6279

However, not the accuracy of classifier but the simplicity of created rules is the point to stress in this case. The classifier uses a single feature **maxpl**, which is the maximal length of a packet. The structure of classification tree is given in Figure 1.

Reasons for selection of **maxpl** as the most important discrimination factor are given in the next section.

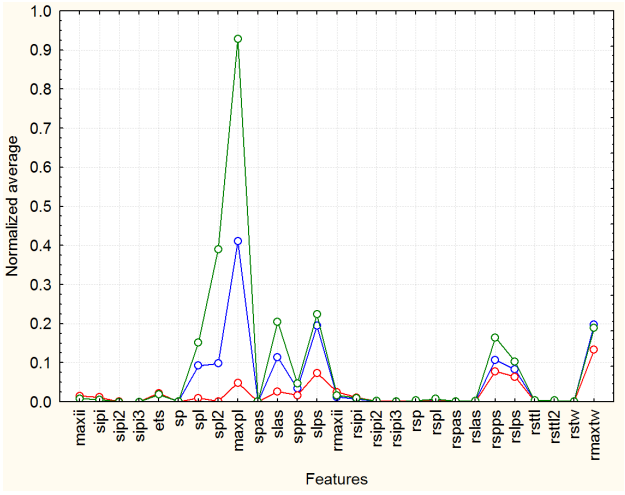


Fig. 2. Normalised average values of the most important spam features calculated separately for subclasses

B. Description of spam subclasses

A classification tree selects features because of the Gini coefficient [11] that is a measure of statistical dispersion. The feature significance is computed by summing over all nodes in the tree the drop in node impurity. The results are expressing relative to the largest sum found over all predictors where the largest sum gets 100 points. Details are given in [10].

The maximal length of package as the feature with major statistical dispersion should differ significantly among classes. It is proved by an analysis of normalised average of features for all three subclasses (presented in Figure 2). The differences between normalised average values calculated for separate subclasses are most significant for this feature. On the same basis, a set of important discriminants may be defined as:

spl Average package length,
spl2 Variation of package length,
maxpl Maximal package length,
slas Average length of package having the ACK flag,
slps Average length of package having the PUSH flag,
rspps Number of packages having the PUSH flag in response,
rslps Average length of package having the PUSH flag in response,
rmaxtw Maximal size of the response TCP window.

All mentioned features (except **rmaxtw**) take highest values for the subclass number 3 and the lowest for the subclass number 2. The subclass number 1 is somewhere between.

The presented analysis is limited to the normalised averages. More information is given by a probability density function. The function that was calculated for selected features (separately for all subclasses) is presented in Figure 3.

The presented normal distribution functions are determined by the following formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (1)$$

TABLE II
THE DISTRIBUTION OF *dnsbl* AND *y_spam* GROUPS AMONG SPAM SUBCLASSES

	dnsbl [%]	*y_spam* [%]
Subclass 1	2.9	97.1
Subclass 2	98.3	1.7
Subclass 3	0.0	100.0

where μ is the mean and σ is the standard deviation. Both values are estimated on the base of values of a feature in the given subclass.

The distribution for the feature **maxpl** presented in 3(b) shows once again that the maximal length of packets is a good discriminant of subclasses. The feature gives three well–separate distributions.

Finally, an analysis of distribution of groups *dnsbl* and *y_spam* among calculated subclasses is presented. The analysis shows that the subclass 2, which has the lowest values of analysed features, consists of spam sent by IP addresses from DNS black lists. In this case, the analysis confirms division of original data between spam detected by blacklists and a spam analyser. However, spam in the *y_spam* group, which contains spam detected by SpamAssassin, is not a solid group and its members split between subclasses 2 and 3. Details are given in Table II.

C. Spam subclasses in classification task

In section II-A, the decision tree was presented as a source of discrimination rules. The tree, which separates spam subclasses, brings the description of the most significant spam discriminants.

The features are used to create a similar tree that separate spam from the rest of network traffic. The tree separates known spam subclasses (1–3) from flows without spam. This new class, labelled as 0, consists of records accepted by SpamAssassin as well as outgoing traffic. The class consists of about 4 thousand of records, which is significant less than number of spam records (54 thousand).

The created tree is given in Figure 4. The tree uses the **rslps** feature that describes the average length of package having the PUSH flag in response.

TABLE III
THE CLASSIFICATION MATRIX FOR THE CLASSIFICATION TREE THAT SEPARATES SPAM SUBCLASSES FROM THE REST OF TRAFFIC

	Class 0	Class 1	Class 2	Class 3
Predicted 0	2920	2	0	282
Predicted 1	242	4169	2	0
Predicted 2	86	1	38396	0
Predicted 3	924	41	0	6021

The classes 1 to 3 are covered with a false positive error. A part of valid traffic is recognised as spam. The error is not equally distributed among classes. The error is minimal for the class 2 where it is only 0.2 percent. For classes 1 and 3

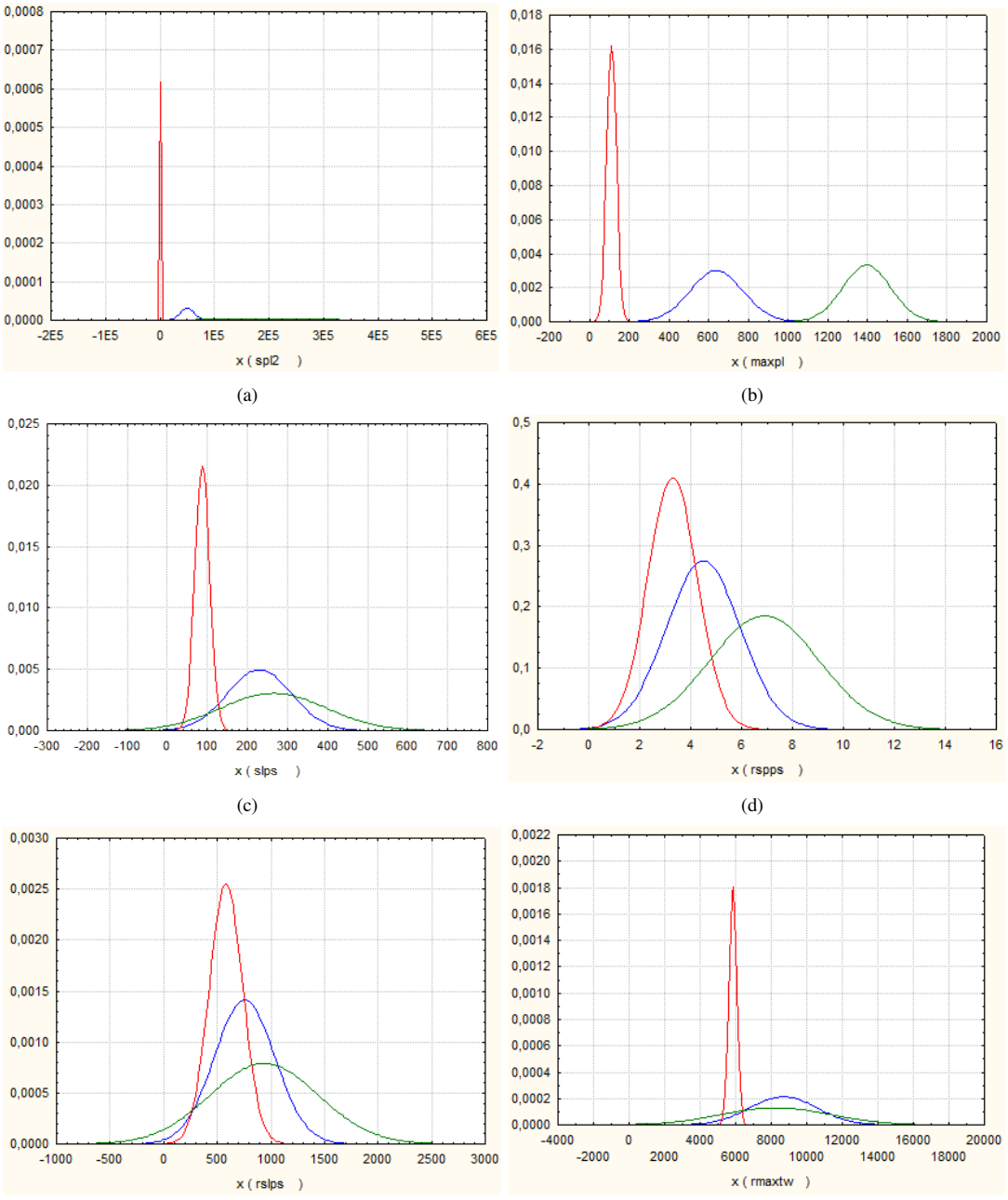


Fig. 3. Probability density calculated for the most characteristic spam features

it is 5.5 and 13.8 percent respectively. Details are given in Table III.

The total accuracy of the tree is 97 percent. When spam is treated as a single class, the accuracy is a slightly lower and the classification tree creates seven rules instead of three.

Similar results, for the same data set but described by different features, are presented in [12] (over 95 percent) where the Principal Component Analysis (PCA) was used to reduce number of features and in [8] (about 96 percent) where 64 features were used.

Although the accurate rate is higher than in cited works, a high error (about 30 percent) in classification of class 0 can be observed. This class is corresponding to the flows without spam. Obviously, this is the most important class since the classifier has to avoid misclassifying non–spam flows with spam flows. The high false positive ratio makes the binary classifier useless, but the classifier that recognises spam subclasses can be still used as a filter.

In the described example, there are three different cases. The first case concerns separation of valid traffic (class 0) and

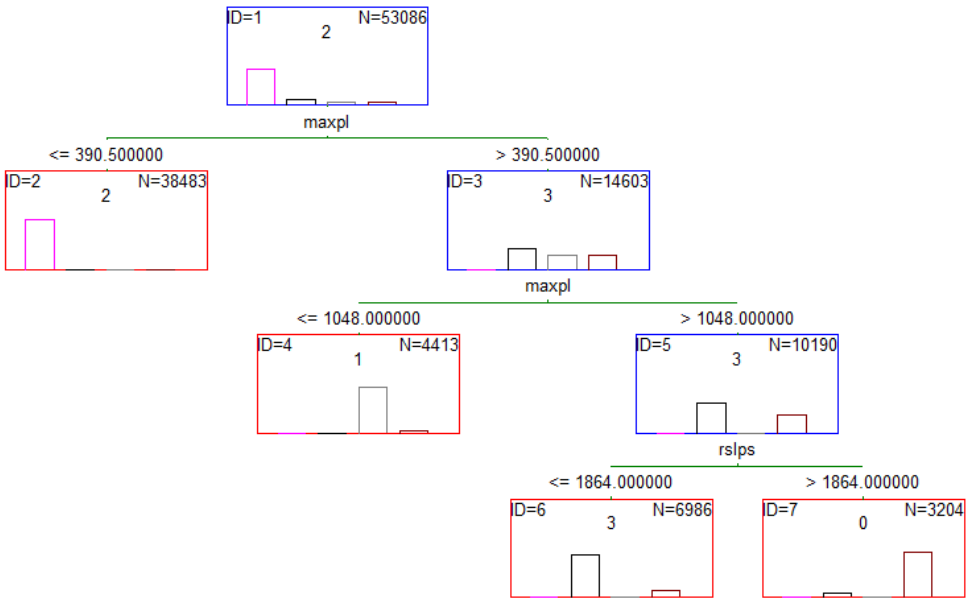


Fig. 4. The classification tree that separates spam subclasses from the rest of traffic

spam from blocked DNS addresses (class 2). In this case, the tree is an adequate tool to separate classes. The second case concerns the class 1. Nearly all members of this class are recognised correctly. However, some flows without spam are also classified as members. This class requires future analysis of false positive. The last case is focused on the class 3. Misclassifications are noticed between this class and the class 0. A more powerful classifier is needed to separate these classes.

III. Model verification

The results of analysis presented in Section II were verified with separate data. Significance of features determined on Žádník's data was validated by comparison with features calculated on Warsaw University of Technology mail server called Alpha.

A. Alpha

The described spam model was created on the base of NetFlow records collected by Žádník. For validation, a new set of NetFlow records was collected at Warsaw University of Technology. The set originates from the mail server called Alpha and consists of NetFlow records described by the same collection of features as Žádník's set. Data was collected through one working week. Over 42 thousand NetFlow records were collected. Among them 589 were labelled as spam.

This ratio between spam and non–spam data is completely different from the data discussed before. Here, the number of non–spam data is lower than the spam data. Therefore, the model was created on a dataset that is not very similar to the dataset collected from the Alpha.

Using the same method as for the Žádník's set (section II-C) a classifier based on a classification tree was created. The classifier separates spam from the rest of flows. The accuracy for both classifiers is very similar (about 97 percent). However, the disparate distribution of classes in the Alpha set results in much lower observed error (about 3 percent) in classification of class 0.

The created classification trees have different structures and splits are based on different features. Therefore, a set of features chose on the base of analysis of the Žádník's set should be verified.

B. Verification of features

Two decision trees that separate spam from the rest of traffic were created. The first one was trained on data from the Žádník's set, the second one on data collected from the Alpha server.

In the training process, the Gini coefficients were calculated. Therefore, the same approach that was used before to evaluate features describing spam can be applied once again to evaluate importance of features. The importance was calculated for all features proposed in [8].

An importance ranking on a 0–100 scale for each feature was created separately for each set. Figure 5 presents calculated significance of features.

The significant correlation between sets cannot be detected. There are several possible reasons for the difference between features selected by different classifiers.

The first reason may lie in the method. When a classification tree is created, the algorithm focuses on the most important features. Third–rate features may be different in various solutions.

The second reason lies in difference between the contents of discussed sets. As an example the feature **spas**, which describes the number of packages having ACK the flag set

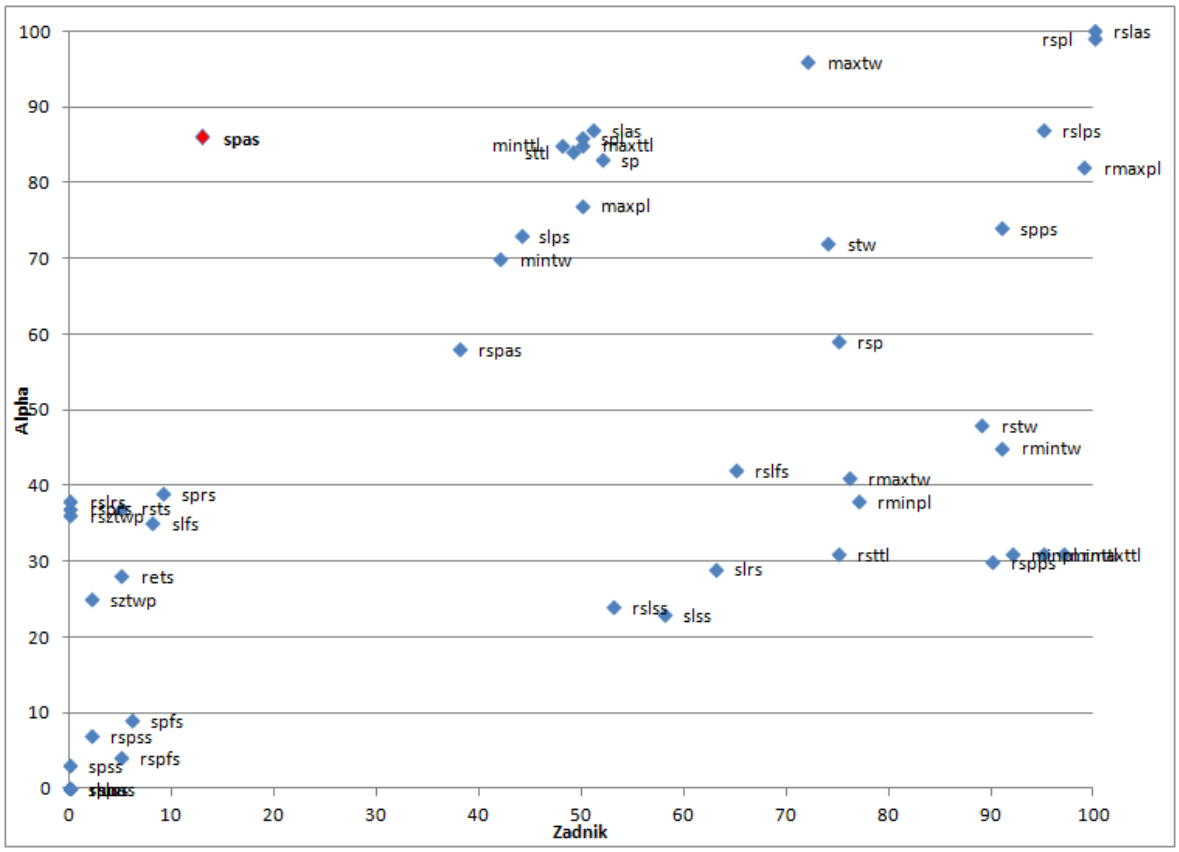


Fig. 5. Significance of features for both data sets: Žádník and Alpha

will be discussed. The feature is quite important for the classifier created on the Alpha set whereas of little importance for classifier created on the Žádník's set. The observation is highlighted in Figure 5.

The ACK flag is used to acknowledge the successful receipt of packets. Therefore, it should be turned on in almost all packages. The only exception is the first package sent in an exchange. In a transaction, the total number of packages with the ACK flag should be the total number of packages sent by a sender minus one. In practice, packages can be resend or missed and exceptions to the rule can be observed.

For discussed sets, the relation between the total number of packages (**sp**) and the total number of packages with the ACK flag (**spas**) in a flow was estimated using the method of least squares [13]. Results are presented in Figure 6.

For data collected on the Alpha server, (Figure 6(b)) the relation is similar to the ideal one:

$$\textbf{spas} = \ \textbf{sp} - 0.6. \tag{2}$$

Meanwhile, in the Žádník's set (Figure 6(a)), the relation is farther from the ideal:

$$\textbf{spas} = 0.9 \ \times \ \textbf{sp} - 9.4. \tag{3}$$

Such differences may influence features evaluation.

C. Universal features

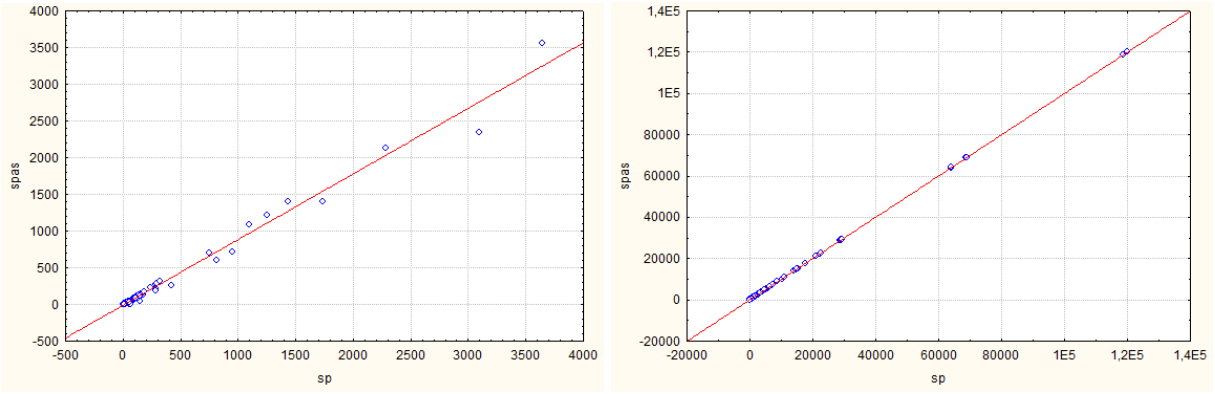
Despite the differences in importance of features evaluated on the base of data from different sources there is a small number of features such as **rslas** or **rspl** significant for both classifiers. However, their number is not enough to create a universal set of features. Therefore, the following method was used to create such set.

Each feature f calculated for a flow from source to designation has its equivalent rf calculated for a response. It is assumed, in the described method, that if a feature is added to the universal set then a response equivalent will be also added and vice versa.

The significance of features is evaluated on the base of Gini coefficients calculated during the decision tree creation. The evaluation functions g_A and g_Z are calculated from the Alpha and the Žádník's sets respectively. Each feature should be evaluated on the base of both evaluations. The presented assumptions result in the following evaluation function

$$g(f) = \frac{\max(g_A(f), g_A(rf)) + \max(g_Z(f), g_Z(rf))}{2}. \tag{4}$$

Because the range of evaluation functions g_A and g_Z is $[0, 100]$ it is reasonable to assume that significant features should at least achieve the level of 50 points.



(a) (b)

Fig. 6. The relation between the total number of packages and the total number of packages with the ACK flag calculated for the Žádník 6(a) and the Alpha 6(b) sets

TABLE IV
THE MOST SIGNIFICANT FEATURES DESCRIBING SPAM

Rank	Name	Description
100,0	slas	Average length of package having the ACK flag
99,5	spl	Average package length
91,0	slps	Average length of package having the PUSH flag
90,5	maxpl	Maximum package length
84,0	maxtw	Maximum TCP window size
82,5	spps	Count of packages having the PUSH flag
73,0	stw	Average TCP window size
68,0	mintw	Minimum TCP window size
67,5	maxttl	Maximum TTL
67,0	sp	Packages count
66,5	minttl	Minimum TTL
66,5	sttl	Average TTL

Among features, twelve have the evaluation result greater than 50 points. The most significant features are collected in Table IV. Information about direction of traffic is skipped. It is assumed that mentioned features should be calculated for both directions. That gives 24 features.

IV. Model application

The model was applied on data collected from a Broadband Remote Access Server (BRAS). On the base of learning data, an ensemble of classifiers was created and used to detect main sources of spam.

A. BRAS

Data was collected from a Broadband Remote Access Server (BRAS). Firstly, a probe that contains full headers was created. In eight seconds, almost 50 million PCAP packages were captured. From the packages payload was removed. The total size of remaining headers was 4.44 GB. Among all packages, 29.5 thousand were transferred by STMP protocol. That produced 407 NetFlow records. The ratio of collected records to the size of created file forced a different approach.

The second set was calculated in a reduced form. Because of huge size of collected PCAP headers NetFlows record were captured instead. Each record was described by 12 features including minimum, maximum, and average size of package, and binary information about flags occurrence. 176 thousand records were captured. The total size of collected NetFlow records was 18 MB.

Collected records was analysed by a spam detector to get out information about sources of spam.

B. Detector

The classifiers that detect spam were created on the base of two learning sets: Alpha and Žádník. Both sets were divided into learning and testing sets. The cardinal number of the training set was similar to the cardinal number of the learning set. Moreover, the proportion of spam to the rest data was similar in the learning and the testing sets, although the proportion of spam in the Žádník's set and the Alpha set are different.

It was mentioned before that a single tree should not be used as a detector. Instead, detector was projected as an ensemble of trees. Each tree was trained on the learning set and validated on the training set.

Each classifier from the ensemble recognises two classes: spam and background traffic. Before, the subclasses of spam were considered to determine the spam model. However, for the final user a determination of the spam subclass is not such important as a detection of spam. Therefore, the binary classification is performed.

The classes are labelled by 1 and -1 respectively. The classifier C_i that returns a decision y_i is described by the s_i coefficient that is the accuracy of discrimination between spam and the background. The final classification decision is the sign of a weighted sum given by the formula:

$$y = \operatorname{sgn}\sum_{i=1}^{n} \frac{s_i}{\sum_{j=1}^{n} s_j} y_i, \tag{5}$$

where n is the number of classifiers in the ensemble.

The detector can return 0 what means that the decision is uncertain.

C. Detection of spammers

The created detector was used to detect spam among flows collected from the BRAS. Over 60 trees trained on various subsets of the universal features set were used to create the detector defined by (5). In the result, 934 records from 176 thousand records were labelled as spam.

In the next step, sources of spam were localised. Records from the captured data came from 64088 unique IP addresses. Among them, 359 have sent at least one record labelled as spam. Most of them (211 addresses) sent just one record labelled as spam, but the record holder sent 221 spam records.

It is hard to assume that each sender from this set is a spammer. Therefore, a spammer was defined as a source of at least 10 spam records. This limitation results in seven main spam senders. Together they sent 46 percent of records labelled as spam. The main spammers can be easy blocked, which results in a significant reduction of spam.

V. Conclusions

In this work, the approach to detect main sources of spam in collected network traffic is presented.

Firstly, flow–level model of spam is created. The model describes spam subclasses and brings information about major features for a spam detection task. The presented model was verified on separate data. The verification resulted in a universal set of features.

The universal set consists of features that should be collected from a network in the form of NetFlow records. Among features are length of packages, information about window size, information about flags etc.

Selected features from the universal set were collected on Broadband Remote Access Server. Next, the detector, which was an ensemble of decision trees learned on various datasets, was created. The detector showed main source of spam among senders of collected flows. An elimination of detected spammers will reduce a number of spam by over 45 percent.

It should be noticed that the spam traffic properties will be changing over time and model will need to be retrained. The traffic properties can be also different in case of intensive spam attack. However, a regular collection of learning records from a network should resolve the first problem. Additionally, special learning sets that simulate intensive attacks can be used to improve model.

Gradually modifications of the model can be easy done by addition of new classifiers to the detector represented by equation (5) (although the total number of classifiers should be limited to avoid the drift learning problem).

Moreover, the detector based on decision trees can be implemented as a network probe. A software solution can be implemented as an nProbe [14] plug–in, but a hardware solution is also possible if the decision algorithm will be implemented on FPGA card.

References

[1] L. Limwiwatkul and A. Rungsawang, "Distributed denial of service detection using tcp/ip header and traffic measurement analysis," vol. 1, pp. 605 – 610 vol.1, oct. 2004.

[2] P. Kobiersky, J. Korenek, and L. Polcak, "Packet header analysis and field extraction for multigigabit networks," pp. 96 –101, april 2009.

[3] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," *SIGMETRICS Perform. Eval. Rev.*, vol. 33, no. 1, pp. 50–60, Jun. 2005. [Online]. Available: http://doi.acm.org/10.1145/1071690.1064220

[4] B. Claise, *Specification of the IP flow information export (IPFIX) protocol for the exchange of IP traffic flow information*, 2008.

[5] ——, *Cisco systems NetFlow services export version 9*, 2004.

[6] A. Moore, M. Crogan, A. W. Moore, Q. Mary, D. Zuev, D. Zuev, and M. L. Crogan, "Discriminators for use in flow-based classification," Tech. Rep., 2005.

[7] C. Leita and M. Dacier, "Sgnet: A worldwide deployable framework to support the analysis of malware threat models," pp. 99 –109, may 2008.

[8] M. Žádník and Z. Michlovský, "Is spam visible in flow-level statistics?" Tech. Rep., 2009.

[9] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979. [Online]. Available: http://dx.doi.org/10.2307/2346830

[10] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.

[11] G. Behera, "Privacy preserving c4.5 using gini index," in *Emerging Trends and Applications in Computer Science (NCETACS), 2011 2nd National Conference on*, march 2011, pp. 1 –4.

[12] M. Grzenda, "Towards the reduction of data used for the classification of network flows," in *Proceedings of the 7th international conference on Hybrid Artificial Intelligent Systems - Volume Part II*, ser. HAIS'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 68–77.

[13] D. Ruppert, S. J. Sheather, and M. P. Wand, "An effective bandwidth selector for local least squares regression (Corr: 96V91 p1380)," *Journal of the American Statistical Association*, vol. 90, pp. 1257–1270, 1995.

[14] L. Deri, "nprobe: an open source netflow probe for gigabit networks," in *In Proc. of Terena TNC 2003*, 2003.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 99–105

RBF ensemble based on reduction of DAG structure

Marcin Luckner
*Faculty of Mathematics and Information Science
Warsaw University of Technology
pl. Politechniki 1, 00-661 Warszawa, Poland
Email: mluckner@mini.pw.edu.pl

Karol Szyszko
†Faculty of Mathematics and Information Science
Warsaw University of Technology
pl. Politechniki 1, 00-661 Warszawa, Poland
Email: szyszkok@student.mini.pw.edu.pl

***Abstract*—Binary classifiers are grouped into an ensemble to solve multi–class problems. One of proposed ensemble structure is a directed acyclic graph. In this structure, a classifier is created for each pair of classes. The number of classifiers can be reduced if groups of classes will be separated instead of individual classes. The proposed method is based on the similarity of classes defined as a distance between classes. For near classes the structure of DAG stays immutable. For the distant classes more than one is separated with a single classifier. In this paper, the proposed method is tested in variants based on various metrics. For the tests, several datasets from UCI repository was used and the results were compared with published works. The tests proved that grouping of radial basis functions into such ensemble reduces the classification cost and the recognition accuracy is not reduced significantly.**

***Index Terms*—Classification, Radial Basis Function, Directed Acyclic Graph, Support Vector Machines**

I. Introduction

RADIAL Basis Functions (RBF) can be used both as multi–class and binary classifiers. Binary RBF classifiers are useful, still improved linear models in a nonlinear subspace [1], [2] and can be used to create an ensemble of classifiers to solve multi–class problems. Such approach was confirmed by research on Support Vector Machines [3]. In fact, there is equivalence between a decision rule created by an SVM with radial kernels and radial networks [4]. Therefore, an ensemble of classifiers can be created for binary RBF classifiers as well as in the case of SVM equivalents.

Although one–step solutions were proposed [5], [6] to solve multi–class tasks by an SVM, they are not efficient [7]. An alternative solution is creation of an ensemble. Main approach to create an ensemble are discussed in [7], [8].

Proposed approaches are One–Against–All (one class is compared against the rest) [9], One–Against–One (a classifier is created for each pair of classes) [10], Error-Correcting Output Codes (class binarisation in order to enhance generalisation ability) [11], and several methods based on graph structures [12], [13].

Among several graph ensemble fusion methods can be used to solve multi–class problems using binary RBF classifiers [14] one of the most popular strategy is grouping SVM classifiers into a directed acyclic graphs (DAG) [13].

In the case of an n-classes problem a tree implementation requires $n-1$ classifiers and the average decision process uses $n-1/2$ classifiers. A DAG ensemble needs $n(n-1)/2$ classifiers to solve the same problem. However, only $n-1$ classifiers is used in the classification process and obtained results are usually better [15], [16].

In the work [17], the method for reduction of number of classifiers in a DAG structure was presented. The proposed method is based on a class similarity. Similar classes are grouped and separated from diametrical different classes as a whole group. In this case, a number of used classifiers is reduced. The method was projected for linear classifiers and verified on a single recognition task. The similarity calculated in the work was a derivative of the Euclidean metric. The algorithm was tested on a single recognition task.

This work is based on the algorithm proposed in [17]. However, several improvements with respect to the previously published works have been done. The algorithm has been tested on a wider set of classification problems. Researches on new problems resulted in modifications of the algorithm. The most important modification was done in limitation rules presented in Section II-C.

The algorithm was a basis for several models based on various definitions of similarity. The following metrics are used in modelling: the Euclidean, the Chebyshev, the Manhattan, the Minkowski, and the Pearson.

Models are verified on various datasets from UCI repository. Therefore, results can be compared with others works and the results of the modified algorithm were compared with results of RBF classifiers grouped into a DAG ensemble.

II. Reduced DAG ensemble

A. DAG ensemble

A directed acyclic graph G, which is described by the set of vertices $V(G)$ and the set of edges $E(G)$, can be declared as an ensemble of binary classifiers.

The vertices are grouped in layers. The first layer contains a single vertex – the root. Each subsequent layer has one more vertex. The last layer consists of n vertices where n is the number of recognised classes.

Each vertex from any layer except the last one has connections with two vertices from the next layer. Vertices on the last layer are leaves. Each leaf from $L(G)$ is connected with a final classification decision (one of the recognised classes). The rest of vertices $V(G)\backslash L(G)$ contains binary classifiers. The decision of the classifier defines which connected vertex

978-83-60810-53-8/$25.00

will be activated next. If a leaf is activated a final decision is determined by a class connected with the leaf.

A vertex is also identified with a group of classes that can be achieved from the vertex. Each vertex can be declared as a root of sub–DAG. Such root represents a group of classes collected on the last layer of the sub–DAG. If $v \in V(G)$ is the root of the sub–DAG G_v then $L(G_v) \subseteq L(G)$ determines classes identified with the vertex.

Sub-DAGs can be also used to determine classifiers. A vertex $v \in V(G) \backslash L(G)$ has two successors v_i and v_j. The successors are identified with classes connected to $L(G_{v_i})$ and $L(G_{v_j})$ respectively. Therefore, the binary classifier in the vertex v divides dataspace between two groups of separable classes from sets $L(G_{v_i})$ and $L(G_{v_j})$.

The aim of the proposed method is to eliminate some of vertices from layers. Then an activated vertex can lie on a layer further than the next one. The example of reduced structure is presented in Figure 1. The elimination of vertices shortens the average classification time. However, not each vertex can be eliminated without a significant reduction of the accuracy ratio. Therefore, the selection of eliminated vertices will be based on similarity between classes.

B. Similarity

A similarity between classes is estimated on the base of a distance. The distance between classes $d(C_X, C_Y)$ depends on the distance between elements of those classes $d(x, y)$ and can be defined as the distance between nearest, furthest elements or as the average distance between all pair of elements. However, mentioned above distances are very time-consuming. Instead, the distance may be approximated as the distance between centroids (the centres of gravity for the classes)

$$d(C_X, C_Y) = d\left(\frac{1}{n_{C_X}} \sum_{x \in C_X} x, \frac{1}{n_{C_Y}} \sum_{y \in C_Y} y\right). \quad (1)$$

The equation (1) can be also used to calculate a distance between groups of classes. If a group is an union of classes $C_X = \bigcup_{i=1}^{k} C_i$ then all members of classes C_i, where $i = 1 \ldots k$, are treated as members of C_X. The distance between such groups can be calculated as (1).

The distance between an individual elements of the data space $d(x, y)$ depends on the selected metric. Usually it is the Euclidean metric

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \quad (2)$$

However, if the recognised elements are described by some specific features it is sometime better to select a different measure.

Two potential candidates are Manhattan and Chebyshev metrics. The Manhattan distance

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i| \quad (3)$$

should be calculated when individual features are independent, but their sum may be treated as a rational measure of the similarity. In the Chebyshev distance

$$d(x, y) = \max_{i \in \{1, \ldots, n\}} |x_i - y_i| \quad (4)$$

the similarity will depend on the maximal difference among features.

In the tests, two more metrics were used.

The Minkowski metric is a parameterised metric that becomes the Euclidean metric for $k = 2$.

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^k\right)^{\frac{1}{k}}. \quad (5)$$

In this work $k = 3$ was used.

The Pearson metric bases on the Pearson product–moment correlation coefficient r_{xy} and presents inverse correlation between data vectors

$$d(x, y) = 1 - r_{xy}. \quad (6)$$

C. Creation of structure

The graph for the n–classes classification task has n layers where the last layer contains leaves labelled as recognised classes. The algorithm starts with the set $V(G)$ of n leaves. Therefore initially $V(G) = L(G)$.

In each step a pair of the vertices v_i and v_j is selected. The selected vertices are roots of DAGs with nearest groups of classes on the last layers $L(G_i)$ and $L(G_j)$.

$$(v_i, v_j) = \arg \min_{v_i, v_j \in V(G)} d(L(G_i), L(G_j)). \quad (7)$$

In this and the following equations, the notation $L(G_i)$ means all classes from the last layer (leaves) of the DAG with root in the vertex v_i. Therefore the distance $d(L(G_i), L(G_j))$ can be calculated as a distance between group of classes (1).

The new vertex v_k is added to the graph G. The vertices v_i and v_j are now its successors. The new graph G is given by the set of vertices

$$V(G) = \{v_k\} \cup V(G) \quad (8)$$

and the set of edges

$$E(G) = \{(v_k, v_i), (v_k, v_j))\} \cup E(G). \quad (9)$$

It is easy to observe that groups $L(G_i)$ and $L(G_j))$, determined by vertices v_i and v_j, may be still the nearest groups in the graph G. Moreover, the created group $L(G_k)$ is the union of groups $L(G_i)$ and $L(G_j)$ and consequently the union may be selected as the nearest with one of groups determined by the successors. Such situations should be avoided. Therefore, the set of vertices that are taken into consideration in the equation (7) should be limited by the following rules.

Under the second rule, that eliminates problem with the same pair selected again, the two vertices v_i and v_j can be joined if and only if the union of classes represented by them is not a subset represented by any existed vertex

$$\forall(v_k \in V(G))L(G_i) \cup L(G_j) \nsubseteq L(G_k). \tag{10}$$

The rule (10) is modification of the rule presented in [18]. The previous rule assumed that $L(G_i) \cup L(G_j) \neq L(G_k)$, however then the algorithm was not convergent for some datasets.

Under the first rule, that eliminates problem of joining successors with a predecessor, the two vertices v_i and v_j can be joined if and only if the set of classes represented by one of them is not a subset of the other

$$C(G_i) \nsubseteq C(G_j) \wedge C(G_j) \nsubseteq C(G_i). \tag{11}$$

Both conditions (11) and (10) can be used to create a limited set of allowed pairs of vertices

$$\begin{aligned} S_P &= \{(v_i, v_j) : v_i, v_j \in V(G) \\ &\wedge \ L(G_i) \nsubseteq L(G_j) \wedge L(G_j) \nsubseteq L(G_i) \\ &\wedge \ \forall(v_k \in V(G))L(G_k) \neq L(G_i) \cup L(G_j)\}. \end{aligned} \tag{12}$$

Moreover, the common part of classes is ignored when the distance is calculated and the final form of the formula (7) is

$$(v_i, v_j) = \arg \min_{(v_i,v_j)\in S_P} d(L(G_i)\backslash L(G_i \cap G_j), L_j \backslash L(G_i \cap G_j)), \tag{13}$$

where

$$L(G_i \cap G_j) = L(G_i) \cap L(G_j). \tag{14}$$

In each step of the algorithm, the two allowed vertices v_i, v_j are joined. The algorithm stops when no join can be made

$$\forall(v_i \in S_G)\exists(v_j \in S_G)L(G_i) \subseteq L(G_j) \vee L(G_j) \subseteq L(G_i). \tag{15}$$

D. Classification

The classification process starts in the root of DAG ensemble and finishes in a vertex from the last layer.

In the DAG ensemble, each classifier rejects one from the recognised classes. Therefore, if the vertex v_i is a root of a DAG with the last layer consists of vertices $L(G_i)$ then the vertex v_j, which is next in the classification path, is connected with classes defined by the following rule

$$L(G_j) = L(G_i) \setminus \{v_k\} \wedge v_k \in L(G_i). \tag{16}$$

In the reduced ensemble, the number of classification steps can be reduced, because a classifier can reject more than one class in one step and the rule (16) is replaced by

$$L(G_j) = L(G_i) \setminus V_k \wedge V_k \subset L(G_i). \tag{17}$$

When a vertex $v \in V(G)\backslash L(G)$ has two successors v_i and v_j, identified with classes connected to vertices $L(G_i)$ and $L(G_j)$ respectively, then the binary classifier in the vertex v divides dataspaces between two groups of separate classes $L(G_i) \setminus (L(G_i) \cap L(G_j)))$ and $L(G_j) \setminus (L(G_i) \cap L(G_j)$.

E. Time reduction

Three main elements that describe costs of the created classifiers are the construction time, the learning time, and the classification time.

The first element is a construction of the ensemble. In the case of a DAG, this element can be omitted. However, it is an important part of the presented method (described in Section II-C. If the cost of the ensemble creation is significant then potential reduction of the learning time can be balanced by a cost of the new element.

At the first glance, the learning time should be shorter in a reduced structure. However, a reduction of classifiers in the ensemble does not necessarily have to result in a reduction of the learning time. All classifiers in a DAG structure present the One–Against–One approach. Meanwhile, at least several classifiers in the reduced structure split dataspace between a class and a group of classes. It was proved that One–Against–One approach can results in lower costs than One–Against–All approach [8]. Therefore, the learning time can be longer for the reduced structure.

Considering pessimistic assumptions on the building time and the learning time, we should remember that the most important cost in the classifier evaluation is the classification time. The reduced structure has definitely lower classification costs that DAG.

In the n–classes recognition time, a DAG structure needs $n-1$ binary classifiers to assign the analysed case to one of recognised classes. If C is a set of analysed cases and C_i is a subset of cases assigned to the class labelled with i then the classification time for the set C is calculated as

$$\sum_{i=1}^{n} |C_i| * (n-1) * t = |C|(n-1) * t, \tag{18}$$

where t is the binary classification time, which should be equal for all binary classifiers from the ensemble.

In the case of the reduced structure, the same set C is defined as the union of cases $\hat{C}_i$ assigned to the class labelled with i by the new classifier. The number of binary classifiers used in the classification depends on the final classification decision and can be calculated as $d_i - 1$, where d_i describes number of vertices on the path from the root to the leaf assigned to the class labelled with i. Therefore, $d_i \leq n$.

The classification time for the set C is calculated as

$$\sum_{i=1}^{n} |\hat{C}_i| * (d_i - 1) * t. \tag{19}$$

From $d_i \leq n$ and $\sum_{i=1}^{n} \hat{C}_i = C = \sum_{i=1}^{n} C_i$ we know that

$$\sum_{i=1}^{n} |\hat{C}_i| * (d_i - 1) * t \leq \sum_{i=1}^{n} |\hat{C}_i| * (n-1) * t \leq |C|(n-1) * t. \tag{20}$$

Therefore, the classification time for the reduced structure is lower if at least one case from an analysed set belongs to the class with a reduced classification path.

III. Results and Discussion

In this work, four sets from UCI repository were used. Letter Image Recognition Data (Letter), Optical Recognition of Handwritten Digits (Optdigits), Glass Identification Database (Glass), and Wine recognition data (Wine).

TABLE I
DATASETS.

	Number of			
Dataset	Training data	Testing data	Classes	Attributes
Wine	125	53	3	16
Glass	150	64	6	16
Letter	15000	5000	26	16
Optdigits	3823	1797	10	64

In the case of Optdigits dataset and Letters relations between the training sets and the testing sets proposed in the reference works [15] and [16] had been used. However, in the case of datasets Wine and Glass solutions proposed in the reference works had been tested by the cross validation method. Therefore, the validation performance was measured by training 70 percent of the training set and testing the other 30 percent of the training set in these cases. Details on the sets are given in Table I.

The computational experiments for this section were done on an Intel Core i5–2500 with 8 GB of RAM.

All the problems were tested using RBF kernels. The accuracy rate was estimated using different kernel parameters γ and cost parameters C where $C = \{2^0, 2^1, \ldots, 2^{12}\}$ and $\gamma = \{2^{-12}, 2^{-11}, \ldots, 2^4\}$. The selection of parameters values is the most time–consuming part of the process. 221 tests must be done to check all pairs of parameters for one metric. On the testing computer, the test series lasted from few seconds to three hours depending on a dataset. Therefore, some proposals should be made to reduce the computation time.

Table II presents the accuracy ratio obtained by using different metrics. Results are very similar and any of metrics cannot be chosen as the best one. However, it can be observed that the Chebyshev and the Manhattan metrics gave worse results. The Minkowski metric, which had been gotten with $k = 3$, resulted in the accuracy similar as the Euclidean metric. In fact, the calculations can be limited to Euclidean and Pearson metrics to cover all of the best results.

The accuracy ratio obtained by the proposed method was compared with RBF classifiers ordered in the DAG structures, presented in works [15] and [16]. The detailed results including used parameters are presented in Table III.

The results of three compared approaches cannot be compared directly because works [15] and [16] covers different data set. Therefore, the proposed method should be compared with each work separately. The average accuracy ratio calculated among datasets discussed in the works is always minimal better for the proposed method.

The results of the proposed method were compared with two more state-of-the art algorithms: One–Against–All and One–Against–One methods. The average accuracy calculated for all data sets was 92.22 percent for the One–Against–One method, 91.92 percent for the One–Against–All method and 92.35 for the proposed method. Details on results and configuration are give in Table IV. The averages are very similar and any method cannot be pointed as the best one.

The most important aspect in the comparison is reduction of created vertices in the DAG structure. The number of vertices used to solve the Optdigits and Letter problems was reduced to 42 and 66 percent of vertices from the DAG structures. The proposed method reduced the number of RBF classifiers that has to be trained. Additionally, the average classification time will be shorter, because of reduction of decision paths. Two examples for Optdigits are shown in Figure 1.

Each graph has leaves marked with circles and vertices with classifiers marked with diamonds. Inside the diamonds two groups of classes are noticed. The classifier inside the vertex divides the data space between members of both groups.

The reduction of the average classification time is clearly visible in the case of the class zero. The class can be selected after three classification steps instead of nine as in the case of non–reduced DAG structure. In the proposed method, only two classes from the last layer are classified in nine steps.

The created graphs have the same number of vertices. However, the structures are different. First, numbers of vertices on layers are different. Moreover, different nearest classes were selected in the first step of the algorithm. For the Chebyshev metric, nearest classes are 1 and 8, whereas for the Euclidean metric the nearest classes are 3 and 9. The differences in the structure results in differences in the accuracy presented in Table II.

The ensembles created in the Letter problem is too complex to present a comparison between several graphs. However, the graph created using the Euclidean metric is presented in Figure 2. The letter 'L' can be recognised in 9 steps. 20 steps are needed to recognise such letters as 'K', 'G', 'Q', 'R', 'B', 'S', and 'Z'. Meanwhile the DAG structure needs 25 classifiers to recognise any letter. Therefore, both average and pessimistic costs are much lower in the proposed solution.

In the case of Glass problem, the reduction ratio is smaller. This is connected with the smaller number of recognised classes. In the Wine problem, a number of vertices cannot be reduced. However the algorithm determines the order of classifiers in the DAG. As it was shown in Table III, the determined order allows the ensemble to obtain the best result.

The method results in reduction of the classification time. In Table V, the learning time and the classification time obtained by DAG classifiers are compared with the time obtained by the proposed method. Additionally, the graph structure building time is presented in the case of the method from this work.

The proposed method reduces the classification time for complex tasks. Although, the classification time for various methods given in second has only an approximate character, the difference between methods bases on strong theoretical bases (Section II-E).

TABLE II
THE ACCURACY RATIO OBTAINED BY USING DIFFERENT METRICS. THE BEST RESULTS ARE EMPHASISED.

Dataset	Chebyshev	Euclidean	Manhattan	Pearson	Minkowski
Wine	98.86	**99.44**	**99.44**	**99.44**	**99.44**
Glass	73.79	**74.22**	73.29	73.81	**74.22**
Letter	97.42	97.63	97.63	**97.70**	97.35
Optdigits	97.5	**98.05**	97.94	**98.05**	**98.05**

TABLE III
THE ACCURACY RATIO OBTAINED BY THE PROPOSED METHOD IS COMPARED WITH RBF CLASSIFIERS ORDERED IN THE DAG STRUCTURES, PRESENTED IN WORKS [15] AND [16]. FOR EACH METHOD, USED PARAMETERS γ AND C ARE GIVEN. DAG VERTICES IS A NUMBER OF VERTICES CREATED BY DAG, AND VERTICES IS A NUMBER OF VERTICES CREATED BY THE PROPOSED METHOD. THE BEST RESULTS ARE EMPHASISED.

		Work [15]			Work [16]			Proposed method			
Dataset	DAG vertices	C	γ	accuracy	C	γ	accuracy	C	γ	accuracy	Vertices
Wine	3	2^2	2^3	**99.44**	2^8	2^{-9}	98.88	2^0	2^{-2}	**99.44**	3
Glass	21	2^{12}	2^1	73.49	2^{12}	2^{-3}	73.83	2^2	2^3	**74.22**	19
Letter	325	-	-	-	2^4	2^2	**97.98**	2^6	2^2	97.70	216
Optdigits	45	2^2	2^3	**98.44**	-	-	-	2^3	2^{-5}	98.05	19

TABLE IV
THE ACCURACY RATIO OBTAINED BY THE PROPOSED METHOD IS COMPARED WITH ONE AGAINST ONE AND ONE AGAINST ALL METHODS

	One–Against–One			One–Against–All			Proposed method		
Dataset	C	γ	accuracy	C	γ	accuracy	C	γ	accuracy
Wine	2^1	2^2	**99.44**	2^1	2^2	98.89	2^0	2^{-2}	**99.44**
Glass	2^{12}	2^1	73.01	2^{12}	2^{-3}	72.20	2^{12}	2^2	**74.22**
Letter	2^4	2^2	**97.98**	2^3	2^2	97.88	2^6	2^2	97.70
Optdigits	2^2	2^3	98.44	2^1	2^3	**98.72**	2^3	2^{-5}	98.05

TABLE V
THE LEARNING TIME AND THE CLASSIFICATION TIME FOR THE REFERENCE WORKS [15] AND [16]. THE BUILDING TIME, THE LEARNING TIME, AND THE CLASSIFICATION TIME FOR THE PROPOSED METHOD.

	Works [15], [16]		Proposed method		
Dataset	Learning [s]	Classification [s]	Building [s]	Learning [s]	Classification [s]
Wine	0.01	0.00	0.01	0.03	0.00
Glass	2.85	0.00	0.03	0.15	0.01
Letter	298.62	92.80	0.05	214.00	37.02
Optdigits	15.47	1.81	0.00	1.75	1.06

IV. CONCLUSION

In this work, the algorithm based on primary works [17], [18] was modified and used to create several models of RBF ensembles. The structures of created ensemble are reduced directed acyclic graphs. The method reduces a number of created classifiers. The classifiers, which discriminate distant classes, are replaced by the classifiers, which separate groups of classes. A theoretical estimation shows that the new structure should reduce the classification time in comparison to the DAG structure.

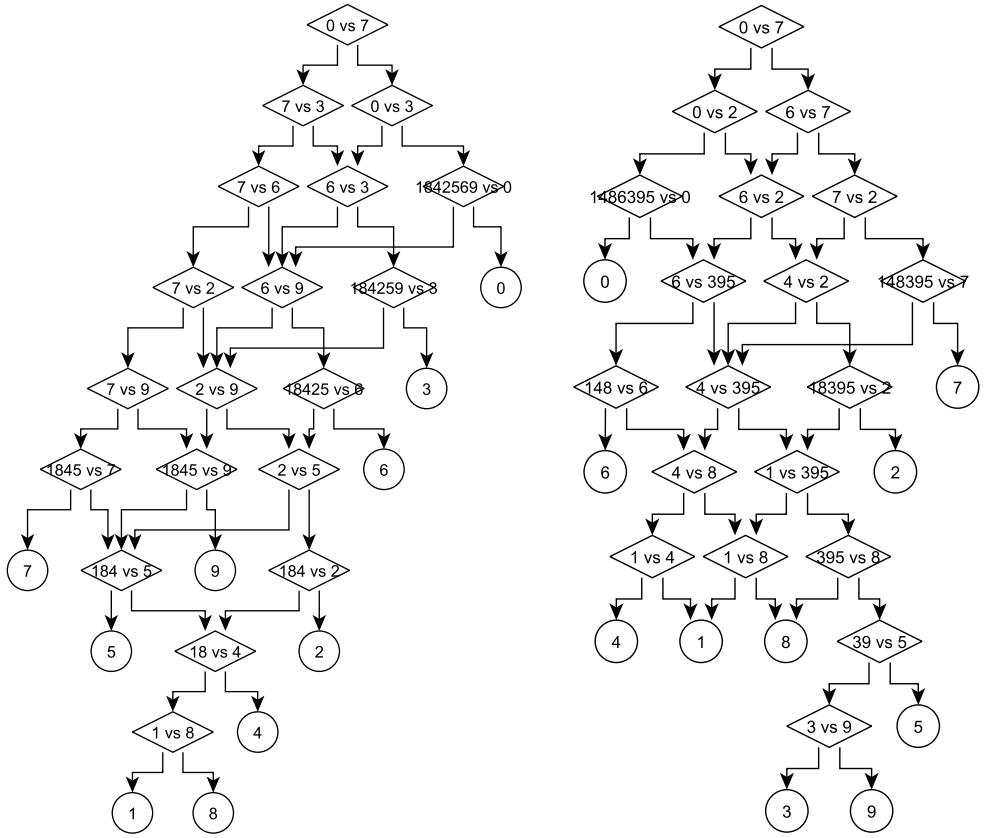
The algorithm was tested on four sets from UCI repository. The obtained results were compared with published results. The accuracy ratio obtained by proposed method is similar to presented in works [15] and [16]. Also results obtained by two state-of-the art algorithms One–Against–All and One–Against–One are nearly identical. The proposed algorithm should be also compared with other methods such as ECOC and Decision Template tested in [19], [20], but its main advantages is not the highest accuracy, but the significant reduction of the number of classifiers in the DAG structure.

The number of created classifiers was reduced to 42 and 66 percent of classifier from the DAG structure in some complex recognition tasks. The reduction of classifiers results in the reduction of the classification time.

REFERENCES

[1] F. Fernández-Navarro, C. Hervás-Martínez, P. Gutiérrez, M. Cruz-Ramírez, and M. Carbonero-Ruz, "Evolutionary q-gaussian radial basis functions for binary-classification," in *Hybrid Artificial Intelligence Systems*, ser. Lecture Notes in Computer Science, E. Corchado, M. GraÃąa Romay, and A. Manhaes Savio, Eds. Springer Berlin Heidelberg, 2010, vol. 6077, pp. 280–287. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13803-4_35

[2] P. Chudzian, "Radial basis function kernel optimization for pattern classification," in *Computer Recognition Systems 4*, ser. Advances in Intelligent and Soft Computing, R. Burduk, M. KurzyÅĎski, M. WoÅşniak, and A. Å£oÅÆnierek, Eds. Springer Berlin Heidelberg, 2011, vol. 95, pp. 99–108. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-20320-6_11



(a) (b)

Fig. 1. Created reduced DAGs for the Optdigits recognition task. The ensembles were created using the Chebyshev 1(a) and the Euclidean 1(b) metrics.

[3] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[4] F. Girosi, "An equivalence between sparse approximation and support vector machines," *Neural Computation*, vol. 10, no. 6, pp. 1455–1480, 1998.

[5] J. Weston and C. Watkins, "Multi-class support vector machines," 1998.

[6] K. Crammer and Y. Singer, "On the learnability and design of output codes for multiclass problems," in *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 2000, pp. 35–46.

[7] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Y. Bang, "Constructing support vector machine ensemble," *Pattern Recognition*, vol. 36, no. 12, pp. 2757–2767, December 2003.

[8] S. Abe, *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[9] K. P. Bennett, *Combining support vector and mathematical programming methods for classification*. Cambridge, MA, USA: MIT Press, 1999, pp. 307–326.

[10] U. H.-G. Kressel, *Pairwise classification and support vector machines*. Cambridge, MA, USA: MIT Press, 1999, pp. 255–268.

[11] M. Bagheri, Q. Gao, and S. Escalera, "Rough set subspace error-correcting output codes," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, 2012, pp. 822–827.

[12] M. A. Kumar and M. Gopal, "A comparison study on multiple binary-class svm methods for unilabel text categorization," *Pattern Recogn. Lett.*, vol. 31, pp. 1437–1444, August 2010.

[13] J. Platt, N. Cristianini, and J. ShaweTaylor, "Large margin dags for multiclass classification," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K. R. Mueller, Eds., 2000, pp. 547–553.

[14] M. Abdel Hady and F. Schwenker, "Decision templates based rbf network for tree-structured multiple classifier fusion," in *Multiple Classifier Systems*, ser. Lecture Notes in Computer Science, J. Benediktsson, J. Kittler, and F. Roli, Eds. Springer Berlin Heidelberg, 2009, vol. 5519, pp. 92–101. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02326-2_10

[15] Debnath, R., Takahide, N., Takahashi, and H., "A decision based one-against-one method for multi-class support vector machine," *Pattern Analysis and Applications*, vol. 7, no. 2, pp. 164–175, July 2004.

[16] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, no. 2, pp. 415–425, 2002.

[17] M. Luckner, "Reducing number of classifiers in dagsvm based on class similarity," in *Image Analysis and Processing âĂ¿ ICIAP 2011*, ser. Lecture Notes in Computer Science, G. Maino and G. Foresti, Eds., vol. 6978. Springer Berlin Heidelberg, 2011, pp. 514–523. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-24085-0_53

[18] ——, "Multiclass svm classification using graphs calibrated by similarity between classes," in *Knowledge-Based and Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, A. KÃśnig, A. Dengel, K. Hinkelmann, K. Kise, R. Howlett, and L. Jain, Eds. Springer Berlin Heidelberg, 2011, vol. 6884, pp. 435–444. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-23866-6_46

[19] T. Wilk and M. Wozniak, "Soft computing methods applied to combination of one-class classifiers," *Neurocomputing*, vol. 75, no. 1, pp. 185–193, 2012.

[20] M. Galar, A. Fernández, E. B. Tartas, H. B. Sola, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.

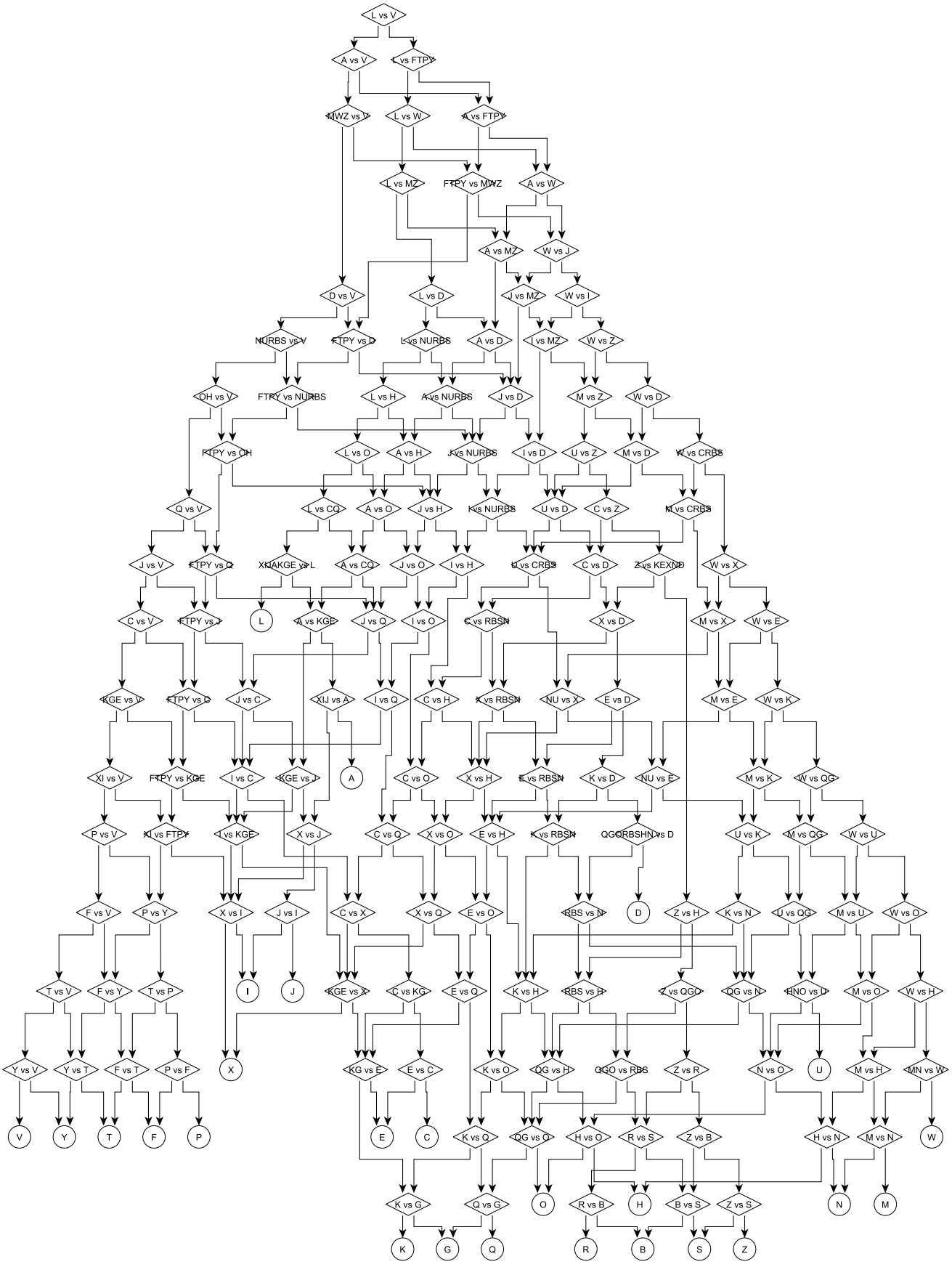


Fig. 2. Created reduced DAG for the Letters recognition task. The ensemble was created using the Euclidean metrics.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 107–110

Recommender System for Ground-Level Ozone Predictions in Kuwait

Mahmood A. Mahmood[1,*], Eiman Tamah Al-Shammari[2], Nashwa El-Bendary[3,*],
Aboul Ella Hassanien[4,*], Hesham A. Hefny[1]

[1]*ISSR, Computer Sciences and Information Dept., Cairo University, Cairo - Egypt*
mahmood.moneim@egyptscience.net, h.hefny@ieee.org
[2]*Faculty of Computing Science and Engineering, Kuwait University, Kuwait*
eiman.tamah@gmail.com
[3]*Arab Academy for Science,Technology, and Maritime Transport, Cairo - Egypt*
nashwa.elbendary@ieee.org
[4]*Faculty of Computers and Information, Cairo University, Cairo - Egypt*
aboitegypt@gmail.com
[*]*Scientific Research Group in Egypt, (SRGE), http://www.egyptscience.net*

***Abstract*—This article presents a recommender system based on rough mereology for predicting Ozone concentration in Kuwait through testing the data gathered from Al-Jahra station. The proposed recommender system consists of three phases; namely pre-processing, classification, and recommendation phases. To evaluate the performance of the presented recommender system, fifteen parameters were used. Those parameters were developed and validated between Jan. 2006 and Sept. 2010. The obtained results demonstrated the effectiveness and the reliability of the proposed recommender system.**

***Index Terms*—recommender system, rough mereology, air pollution, ground-level Ozone**

I. Introduction

INDUSTRIALIZATION, technical growth and over population in urban areas of Kuwait have resulted in increased air pollution [1], [2]. This has deteriorated the quality of fresh air. Toxic air pollutants in close proximity to populate areas can have adverse health effects. In this respect, surface Ozone can become a serious problem in the urban areas of Kuwait - if it frequently occurs in sufficient concentration to threaten human health and environment [3].

Ground-level Ozone (O3) is formed by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOC) in the presence of heat and sunlight. It is difficult to exactly define the formation and destruction mechanism of Ozone. This is because Ozone is an extremely reactive pollutant and can be scavenged by its precursors [2], [4]. As a result, the area of air pollution forecasting through empirical methods has gained importance with the availability of sufficient data. Earlier forecasting models were based on simple empirical data correlations, but the availability of a large amount of information has resulted in development of complex air pollution simulations for forecasting [2], [5]. Management of public warning strategies for Ozone levels in densely populated areas require accurate forecasts of ambient levels. Although Ozone prediction models exist or have been proposed at several cities [6]–[9], [16], [17], they have not been assessed in realistic conditions.

Recently, there has been an increase in air pollution in urban areas of Kuwait. This is the result of rapid industrialization, technical growth and over population. In the past, it has been observed that toxic air pollutants in close proximity of populated areas can have adverse health effects. Ozone is one such pollutant that can become a serious problem in urban areas of Kuwait if it occurs in sufficient concentration. Thus, accurate forecasting of surface Ozone is required as it can help with successful implementation of public warning strategies during episodic days in Kuwait.

This article presents a recommender system based on rough mereology for predicting Ozone concentration in Kuwait through testing the data gathered from Al-Jahra station. The proposed system firstly maps the Ozone dataset into a normalized dataset of ground-level Ozone predictions. Then, rough mereology and rough inclusion techniques are applied for clustering and classifying the normalized Ozone dataset into sets of granules with different radius. Voting by objects approach is subsequently applied in order to select the optimized granules. Finally, normalized rating matrix is acquired, then the predicted ground-level Ozone is recommended. To evaluate the performance of the presented recommender system, fifteen parameters were used. Those parameters were developed and validated between Jan. 2006 and Sept. 2010. The initial three years of data are used to develop the predicting models and the remaining data is used for testing and verifying these models.

The rest of this article is organized as follows. Section II presents the basic concept of rough mereology. Section III describes the different phases of the proposed recommender system; namely pre-processing, classification, and recommendation phases. Section IV introduces experimental results via firstly discussing the tested dataset in addition to the details of

978-83-60810-53-8/$25.00

the applied Air Quality Index (AQI), then presenting statistical analysis of the obtained experimental results. Finally, Section V presents and discusses conclusions.

II. Rough Mereology: An Overview

Rough mereology proposed by Lesniewski in [10] as the theory of concept. The relation of mereology is a part of relation, e.g. x mereology y means x is a part of y. According to Polkowski [11], the mereology relation described in equation (1), where $\pi(u, w)$ is a partial relation (proper part) and $ing(u, w)$ is ingredient relation means an improper part.

$$ing(u, w) \Leftrightarrow \pi(u, w) \ or \ u = w \tag{1}$$

$\mu(x, y, r)$ means rough mereology relation x is part of y at least degree r, also described as shown in equation (2).

$$\mu(x, y, r) = sim_{\delta}(x, y, r) \Leftrightarrow \rho(x, y) \leq (1 - r) \tag{2}$$

Computing the indiscernibility relation to get the object can be done using rough inclusion, which is of less complexity time than the indiscernibility relation computed by rough set technique. Rough inclusion from metric according to Polkowski in [11] computed by the Euclidean metric space or Manhattan space, where,

$$\mu_h(x, y, r) \Leftrightarrow \rho(x, y) \leq 1 - r \tag{3}$$

Then, the indiscernibility relation Ind can be computed as shown in equation (4).

$$Ind(x, y) = \frac{|IND(x, y)|}{|A|} \tag{4}$$

Then, equation (2) becomes equations (5) and (6).

$$\mu_h(x, y, r) \Leftrightarrow Ind(x, y) \geq r \tag{5}$$

$$IND(x, y) = a \in A : a(x) = a(y) \tag{6}$$

Where, a is an attribute(s) in an information system A, and $|A|$ is the cardinality of a set A.

III. The proposed recommender system

The proposed Ozone recommender system tested for ground-level Ozone data collected at Al-Jahra city in Kuwait during the time from January/2006 to September/2010. The architecture of the proposed system consists of three phases; namely pre-processing, classification, and recommendation phases.

A. Pre-processing Phase

In this phase, the proposed recommender system receives the ground-level Ozone data as an input, then it generates a normalized rating matrix. As shown in equation (7) with X representing the original value, the input data were mapped into a normalized dataset of the range [0,1], where the values 0 and 1, represent the smallest and the largest values in each dataset attribute, respectively.

$$Normalized = \frac{X - min.value}{max.value - min.value} \tag{7}$$

B. Classification Phase

In this phase, the proposed recommender system receives the normalized Ozone dataset generated from the pre-processing phase. Rough mereology and rough inclusion approaches were applied in order to produce rough inclusion table that reflects similarity degree among parameters. Then, the granulation mechanism that reflects a given set of inclusion data into collection of granules will be applied via voting by training objects in order to produce the optimal similarity measurement.

1) Ground-level Ozone Clustering: In this phase, the granular computing formalized within the theory of rough mereology - as proposed by Polkowski; as an application of the idea of a granular reflection of data and of classifiers induced from it [5] is used to classify the normalized data into group of similar data according to the definition of rough mereology. In this section, a brief description of the theory of rough mereology and rough inclusion will be presented.

2) Rough Mereology: Indiscernibility relations in rough set theory represents the core problem in this theory that it takes a long of time to get the classification of data. Rough mereology theory used the flexible similarity relations, which allow for huge data to be classified. The similarity relations that will be used must be satisfy properties are MON, ID, EXT. Rough inclusion technique satisfy their three properties of similarity relations; namely Monotonic (MON), Identity (ID), Extreme, or proportionality (EXT) [5].

- (MON) if similarity (x, y, 1) then for each z, from similarity (z, x, r) it follows that similarity (z, y, r).
- (ID) similarity (x, x, 1) for each x.
- (EXT) if similarity (x, y, r) and $s \leq r$ then similarity (x, y, s).

3) Rough Inclusion: Rough Inclusion is a technique that uses the Reduced Hamming Distance [11] equation to compute the similarity between vector u and v, where u represents a user and v represents an item, as shown in equation (8) [5], where, $IND(u, v) = \{a \in A : a(u) = a(v)\}$, and $|A|$ denotes the cardinality of set A.

$$ind(u, v) = \frac{|IND(u, v)|}{|A|} \tag{8}$$

After applying equation (8) on the normalized rating matrix, similarity table is produced by using rough inclusion approach. The proposed recommender system classifies the attributes of the dataset using *Gödel* t-norm technique [6], as shown in equation (9).

$$T_{min}(a, b) = min\{a, b\} \tag{9}$$

The granulation mechanism reflects a given inclusion data into collection of granules, each granule has a fixed granule radius value r, where $r \in$ the interval [0,1] [7]. Accordingly, the proposed recommender approach applies the voting by training objects to produce prediction/recommendation regarding the ground-level Ozone.

4) Voting by Training Objects: Voting by objects stage takes as an input the list of similarity measure tables for each radius resulted from the clustering stage and computes the accuracy rate for each radius as shown in equation (10). Then, the optimum radius that represent the largest accuracy measure is selected. This stage is divided into two steps: 1) accuracy measure computation, which uses equation (10) to compute the accuracy for each table representing the similarity measure at radius r and 2) optimization step, where the table containing the largest accuracy measure at radius r is selected, so the radius r called the optimum radius r_{opt}.

$$Accuracyrate = (\frac{T}{N}) * 100 \tag{10}$$

Where, T is the rate of number of valid tuples and N is the total number of tuples in each row.

Let a is an attribute, u is a test object, v is a training object, and ϵ is a selected number in an interval $[0, 0.1]$ taken every 0.01. So, the final value of ϵ is taken according to the equation of the factor, shown in equation (11).

$$q_a(u, v) = \frac{|a(u) - a(v)|}{diam(a)} \tag{11}$$

Where, diam(a) is computed as shown in equation (12).

$$diam(a) = Max(a) - Min(a) \tag{12}$$

and $a(u)$ can be computed as follows:

$$a(u) = \begin{cases} min & \text{if } u < min, \\ max & \text{if } u > max, \\ u & \text{if } otherwise \end{cases}$$

For each selected ϵ, $sel_u(c, t) = \frac{\sum w_u(v,t)}{sizeofc}$ is computed, where c is the decision category, t is selected t-norm, and $w_u(v, t) = ind_\epsilon(u, v)$. Then, u is assigned to the category with maximal $sel_u(c, t)$. The end value of ϵ is selected when $q_a(u, v) < \epsilon$. Based on the optimum radius selected from the voting by object stage, rules are generated from the rough inclusion table of selected optimum radius.

C. Recommendation Phase

In this phase the system receives the testing Ozone dataset as an input, then it outputs a recommendation/prediction value of ground-level Ozone as shown in equation (13) according to the formula that used to recommendation/prediction in collaborative filtering technique.

$$prediction = \frac{\sum_i \frac{w_{u,i} - mean}{\sigma_u} \times w_{a,u}}{\sum_i w_{a,u}} \times \sigma_a + mean \tag{13}$$

Where, $w_{u,i}$ is the Pearson's correlation coefficient shown in equation (14).

$$w_{a,u} = \frac{\sum_{i=1}^{m} (r_{a,i} - mean) \times (r_{u,i} - mean)}{\sigma_a \times \sigma_u} \tag{14}$$

IV. Experimental analysis and discussion

A. Dataset

The study used hourly air pollutants data from January 2006 through September 2010 gathered by the Environmental Monitoring Information System of Kuwait (eMISK, working under the Environment Public Authority of Kuwait). The data were collected from Al-Jahra fixed surface station. The initial three years of data was used to develop the forecasting models and the remaining data was used for testing and verifying these models. Table I represents the generated rules of Al-Jahra dataset.

B. Air Quality Index (AQI)

The Air Quality Index (AQI) is a key tool for making information about outdoor air quality as easy to find and understand as weather forecasts. It is an index for reporting daily air quality via updating how healthy or unhealthy the air is. Equation (15) shows the AQI conversion formula for the ground-level Ozone.

$$AQI = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} \times (C_{O3} - BP_{LO} + I_{LO}) \tag{15}$$

Where, I_{HI} is the index value at the upper limit of the AQI category, I_{LO} is the index value at the lower limit of the AQI category, BP_{HI} is the break-point concentration at upper limit of the AQI category, BP_{LO} is the break-point concentration at lower limit of the AQI category, and C_{O3} refers to 1-hour Ozone concentration. Also, depending on the value of the Ozone concentration C_{O3}, the values of the I_{HI}, I_{LO}, BP_{HI}, and BP_{LO} parameters are being identified. For example, if the detected value of the C_{O3} is within the range [0.00, 0.059], then the values of the I_{HI}, I_{LO}, BP_{HI}, and BP_{LO} parameters will be 50.00, 0.00, 0.059, and 0.00, respectively [12]. The AQI is considered a standard measurement that runs from 0 to 500, where the higher the AQI value, the greater the level of air pollution and the greater the health concern [13].

C. Statistical Analysis

Evaluating the recommendation quality of the proposed recommender system is mainly based on statistical precision

TABLE I
RULE GENERATION TABLE OF AL-JAHRA DATASET

Rule	Description	Accuracy
Rule(1)	if SO2 is healthy then Ozone is healthy	79.1087
Rule(2)	if NO is healthy then Ozone is unhealthy	29.5273
Rule(3)	if NOX is healthy then Ozone is healthy	62.2358
Rule(4)	if NO2 is healthy and Wind-deg is healthy then Ozone is unhealthy	26.7496
Rule(5)	if PM10 is healthy then Ozone is healthy	79.3381
Rule(6)	if CO is healthy and Wind-deg is healthy then Ozone is unhealthy	26.7496
Rule(7)	if CH4 is healthy then Ozone is healthy	76.1768
Rule(8)	if NCH4 is healthy then Ozone is healthy	73.8360
Rule(9)	if WS is healthy then Ozone is healthy	83.3083
Rule(10)	if WG is healthy then Ozone is healthy	83.8488
Rule(11)	if SOLAR is healthy then Ozone is healthy	80.2869
Rule(12)	if TEMP-IND is healthy then Ozone is unhealthy	44.2708
Rule(13)	if TEMP-AMB is healthy then Ozone is healthy	83.3141
Rule(14)	if CO2 is healthy then Ozone is healthy	69.2238

and decision supporting precision measurement methods [14]. Statistical precision measurement method adopts the MAE (Mean Absolute Error) in order to measure the recommendation quality [15]. MAE is a commonly used recommendation quality measurement method. MAE calculates the irrelevance between the recommendation value predicted by the recommender system and the actual evaluation value. Each pair of interest predicted rank is represented as $< p_i, q_i >$, where p_i is the system predicted value and q_i is the user evaluation value. Based on the entire set of $< p_i, q_i >$ pairs, MAE calculates the absolute error value $|p_i - q_i|$ and the sum of all the absolute error value. Then, the average value is calculated. Small MAE values represent a good recommendation quality indicators. The predicted values of rating sets can be represented as $p_1, p_2, ..., p_N$ and the corresponding actual testing rating set can be represented as $q_1, q_2, ..., q_N$. The MAE can be defined as shown in equation (16).

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \tag{16}$$

Based on experimental results, the ground-level Ozone depends on the values of NO, (NO2, CO), Wind direction degree, and Industrial temperature. Figure 1 depicts the monthly actual values, predicted values by the proposed recommender system, and the mean absolute error (MAE) of the ground-level Ozone in Kuwait (AL-Jahra city station) during the time from January-2009 to September-2010. As shown in figure 1, the curve presenting the values of ground-level Ozone predicted by the proposed recommender system has a similar behavior as the curve presenting the actual values of the ground-level Ozone dataset. Also, both curves existed in the healthy region of the O3 value, which is less than 0.165. For the MAE curve, it almost matches the predicted curve for the months (10/2009, and 11/2009) as the data readings gathered in the tested dataset during those months was zeros, the MAE curve may be in some points drawn over the actual values of the ground-level Ozone dataset because the actual dataset attribute's has sparsity problem and the rough mereology classifier predict this value as the lowest value in this attribute.

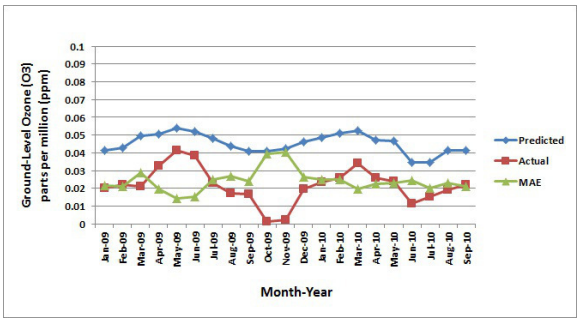


Fig. 1. Monthly values of actual, predicted, and MAE values of ground-level Ozone in Kuwait (AL-Jahra) [Jan-2009 to Sept-2010]

V. Conclusions

In this article, based on ground-level Ozone concentration data gathered from Al-Jahra station in Kuwait, a rough mereology based recommender system was presented for the prediction of ground-level Ozone pollution. The obtained results demonstrate the effectiveness and the reliability of the proposed recommender system. Resulted experimental values of ground-level Ozone predicted by the proposed recommender system showed similar behavior as the actual tested values of the ground-level Ozone dataset. Also, both experimentally resulted and actual dataset values existed in the healthy region of the O3 value, which is less than 0.165 ppm according to the reference AQI.

References

[1] S. Abdul-Wahab, W. Bouhamra, H. Ettouney, B. Sowerby, B. D. Crittenden, "Analysis of Ozone Pollution in the Shuaiba Industrial Area in Kuwait," *Int. J. Environ. Studies*, vol. 57, no. 2, pp. 207–224, 2000.

[2] E. T. Al-Shammari, "Public warning systems for forecasting ambient ozone pollution in Kuwait," *Environmental Systems Research*, vol. 2, no. 2, 2013.

[3] K. N. Jallad, C. E. Jallad, "Analysis of Ambient Ozone and Precursor Monitoring Data in a Densely Populated Residential Area of Kuwait," *J. Saudi Chem. Soc.*, vol. 14, pp. 363–372, 2010.

[4] B. Dimitriades, "Photochemical Oxidant Formation: Overview of Current Knowledge and Emerging Issues," *Atmospheric Ozone Research and its Policy Implications*, vol. 35, Studies in Environmental Science, Elsevier Science Publishers, Amsterdam, pp. 35–43, 1989.

[5] B. Telenta, N. Alfksic, M. Dacic, "Application of the Operational Synoptic Model for Pollution Forecasting in Accidental Situations," *Atmos. Environ.*, vol. 28, pp. 2885–2891, 1995.

[6] S. M. Robeson, D. G. Steyn, "Evaluation and Comparison of Statistical Forecast Models for Daily Maximum Ozone Concentrations," *Atmos. Environ.*, vol. 24B, pp. 303–312, 1990.

[7] D. Elsom, "Smog Alert: Managing Urban Air Quality," *London: Earthscan Publications Limited*, 1996.

[8] H., Noordijk, "The National Smog Warning System in the Netherlands; a Combination of Measuring and Modeling," *Air Pollution*, vol. 2, Pollution Control and Monitoring, WIT Press: Southampton, pp. 169–176, 1994.

[9] J. Yi, V. R. Prybutok, "A Neural Network Model Forecasting for Prediction of Daily Maximum Ozone Concentration in an Industrialized Urban Area," *Environ. Pollut.*, vol. 92, pp. 349–357, 1996.

[10] S. Lesniewski, "On the foundations of set theory," *Topoi*, vol. 2, pp. 7–52, 1982.

[11] L. Polkowski and P. Artiemjew, "Granular Computing in the Frame of Rough Mereology. A Case Study: Classification of Data into Decision Categories by Means of Granular Reflections of Data, rq' *International journal of intellegent systems*, vol. 26, no. 6, pp. 555–571, 2011.

[12] http://www.ncair.org/airaware/Ozone/codecalc.shtml (Accessed: March,2013)

[13] U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Outreach and Information Division, "Air Quality Index (AQI) - A Guide to Air Quality and Your Health, EPA-456/F-09-002, August 2009.

[14] R. S. Ettouney, F. S. Mjalli, J. G. Zaki, M. A. El-Rifai, H. M. Ettouney, "Forecasting of Ozone Pollution using Artificial Neural Networks," *Mgmt. Environ. Quality: An Int. J.*, vol. 20, no. 6, pp. 668–683, 2009.

[15] S. Abdul-Wahab, W. Bouhamra, H. Ettouney, B. Sowerby, and B. D. Crittenden, "Predicting Ozone Levels: A Statistical Model for Predicting Ozone Levels," *Environ. Sci. Pollut. Res.*, vol. 3, pp. 195–204, 1996.

[16] A. Ali, S. E. Amin, H. H. Ramadan, and M. F. Tolba, "Ozone monitoring instrument aerosol products: Algorithm modeling and validation with ground based measurements over Europe," *Proceedings of the International Conference on Computer Engineering and Systems, (ICCES'2011),*, pp. 181–186, 2011.

[17] A. Ali, S. E. Amin, H. H. Ramadan, and M. F. Tolba, "Data assimilation of ozone monitoring instrument images for improving aerosol optical depth prediction," *8th International Conference on Informatics and Systems (INFOS 2012)*, pp. BIO131–BIO139, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 111–114

Prediction of School Dropout Risk Group Using Neural Network

Valquíria R. C. Martinho
Department of Electro-Electronic
Federal Institute of Mato Grosso
Rua Zulmira Canavarros, nº. 95
CEP: 78000-000
Cuiabá, MT, Brazil
vribeiro@terra.com.br

Clodoaldo Nunes
Department of Informatics
Federal Institute of Mato Grosso
Rua Zulmira Canavarros, nº. 95
CEP: 78000-000
Cuiabá, MT, Brazil
cncefet@gmail.com

Carlos Roberto Minussi
Laboratory of Intelligent Systems
Electrical Engineering
Department, Campus of Ilha
Solteira, UNESP Av. Brasil 56,
PO Box 31 CEP: 15385-000
Ilha Solteira, SP, Brazil
minussi@dee.feis.unesp.br

***Abstract*—Dropping out of school is one of the most complex and crucial problems in education, causing social, economic, political, academic and financial losses. In order to contribute to solve the situation, this paper presents the potentials of an intelligent, robust and innovative system, developed for the prediction of risk groups of student dropout, using a Fuzzy-ARTMAP Neural Network, one of the techniques of artificial intelligence, with possibility of continued learning. This study was conducted under the Federal Institute of Education, Science and Technology of Mato Grosso, with students of the Colleges of Technology in Automation and Industrial Control, Control Works, Internet Systems, Computer Networks and Executive Secretary. The results showed that the proposed system is satisfactory, with global accuracy superior to 76% and significant degree of reliability, making possible the early identification, even in the first term of the course, the group of students likely to drop out.**

I. Introduction

HISTORICALLY, school dropout is one of the most complex and crucial problems in education, causing social, economic, political, academic and financial damage to all the people involved in the educational process, from the students to the governmental and promotional agences that long for efficient strategies to reduce the indexes of school dropout, since the measures adopted up to now did not have the desired effect.

In relation to higher education, school dropout is an international problem. Although its indexes show considerable variations among different nations, they show that in fact school dropout is present and strikes more and more a higher number of higher educational institutes (HEI) worldwide.

It is worth mentioning the United States - USA, with a dropout rate in colleges and universities of around 40%, representing a decline in the index of students graduated in higher education. Conversely, China and India empower higher education, increasing the conclusion index. Between these extremes lies Brazil, presenting a mean dropout index of about 20%.

Even taking into account all the differences and specificities of the (HEI) of different nations, the difficult task of solving the evasion problem is still common ground between them.

From this perspective, prevention and intervention programs are developed and structured taking into account the results of researches that identify the possible causes that generate the phenomenon of evasion. However, such measures could be more fruitful if there was prior knowledge of the students prone to evasion. And, for this, the development of methods, instruments or systems capable of previously making this identification is necessary.

To meet this need an intelligent, ambitious and innovative system was developed, for the prediction of risk groups of student dropout in presential higher education courses [1], using artificial intelligence techniques, the Fuzzy ARTMAP Neural Network [2-4]. This network has a structure in which the training is carried out in a supervised and self-organized way, with the possibility of continued learning [2].

This paper aims at presenting and making the developed intelligent system available as a possibility of identifying, in a proactive, continued and accurately the students of the traditional presential education, prone to evasion in higher education. And also to disseminate their fruitful results that contributed to the development of prevention and intervention programs, in order to improve retention of those students identified in the institution [1].

II. ART and FUZZY ARTMAP Neural Networks

The Artificial Neural Networks (ANN) [5] are computational tools that emulate the human brain and learn with the experience, trying to model and simulate its learning process, organizing its neurons in such a way that they will be capable of processing the information.

The ART network systems are able to solve the “stability-plasticity” dilemma. They are plastic because they are able to learn to adapt to a changing environment and, at the same time, preserve their previously learned knowledge while maintaining their ability to learn new patterns, therefore they are stable.

The basic structure of an ART neural network consists of two subsystems of attention and orientation, where some elements such as: two layers of neurons (F1 and F2) and their synaptic weights (Wij amd Vji), the module parameter vigilance (ρ) and the module reset are arranged and inter-linked.

Briefly, the process of classification of an ART network consists of four phases [6]:

978-83-60810-53-8/$25.00

- Recognition: recognizes the stimuli produced in layer F2 and selects the category of higher value after calculating the function choice.

- Comparison: through the vigilance parameter, tests the similarity between the input vector and the prototype vector, whether allowing or not the inclusion of the pattern input in the category. If the vigilance parameter is not met, the input vector is stored in another neuron.

- Search: for every new input vector, searches for a neuron in layer F2 to represent it.

- Training: the training only starts after the conclusion of the search process, it can occur quickly or slowly.

The Fuzzy ART neural network [7] uses the theory of the fuzzy sets, employing the minimum operator (^) AND Fuzzy, enabling the treatment of patterns of binary and analogical input, in an interval [0, 1], and increasing the generalization ability of the network.

In the Fuzzy ARTMAP model, two ART modules are interlinked through an inter-ART module, called Field Map. This module has a self-regulatory mechanism called match tracking that seeks for "marriages" or combinations among the categories of ARTa and ARTb modules, aiming to increase the generalization level and reduce the network error [2].

The architecture of the Fuzzy-ARTMAP neural network, has been designed to conduct supervised learning in an environment or set of multidimensional data. When the Fuzzy-ARTMAP network is used in a learning problem situation, it is trained until it can classify correctly all the training data.

The mathematical development and the algorithms for the processing of a Fuzzy-ART and Fuzz-ARTMAP neural network are found, respectively in [7] and [8], and applied in [9].

III. Methodology

This study was conducted under the Federal Institute of Education, Science and Technology of Mato Grosso - IFMT. The universe of interest are the students enrolled in the Colleges of Technology (CT) in Automation and Industrial Control, Control Works, Internet Systems, Computer Networks and Executive Secretary at IFMT, attending presential courses in the morning, afternoon and evening. The choice is justified in view of the high dropout rates, verified by previous statistical studies, noting that CT Automation and Industrial Control, reached a dropout rate of 62.46% from 2004 to 2010 [1].

In the implementation and pilot test of the intelligent system proposed, the neural network was fed with data belonging to all the students enrolled in the CT, from 2004 to 2009, making a total of 1650 samples for the training phase, constituting the basis historical data. For diagnosis 499 samples, of data from the students enrolled in 2010 and 2011 were used [1].

The database for prediction of the risk group prone to evasion consists of the students' characteristics such as demographic factors, and factors internal and external to the school. These characteristics were lifted from data from the selection processes at IFMT, the Q-Selection, which stores the answers of the socioeconomic questionnaire filled by the students on the day they enroll for the selection examination and the Q-Academic, system of integrated academic management, where all the academic history of the IFMT students is concentrated [1].

The input vector of the Fuzzy ARTMAP neural network is composed by 16 parameters considered as significant for the school dropout prediction and the output of the network constituted by two classes, evasion and non-evasion. The input-output vector pairs are represented in the binary coding, being the input vector composed by 41bits and, the expected response represented by 1 bit. A summary of the input and output variables of the neural network can be visualized in Table I.

IV. Fuzzy ARTMAP Neural System Proposed for the Evasion Prediction

The data that involve the study about evasion, are sometimes, complex, subjective, non-linear, inter-related and keep in themselves the specificities inherent to the different levels of teaching, courses and institutions that one can analyze, thus choosing ANN, as among its potentialities there is the possibility of processing problems where complex and unknown relations are involved among different sets of data and, also adjust the relations of non-linearity between the input and outup variables [1]. More specifically, the Fuzzy ARTMAP network, where the training is carried out in a supervised and self-organized way, with possibility of continued learning, as implemented in [10]. Its application potential aims at solving several problems of classification and of approach of non-linear functions and showing prompt reply.

The input of the Fuzzy ARTMAP network proposed is represented by vector a (input of the module ARTa) and its

TABLE I. Composition of input and output vectors

	Characteristics of the Subvectors of a and y			
	Position	***Name***	***Abbreviation***	***Size***
Variable of the Input Vector (a) of the Network	a_1	Gender	Gen	1 bit
	a_2	Age Group	Ag	3 bits
	a_3	Ethnicity	Etn	3 bits
	a_4	Marital Status	MSt	3 bits
	a_5	People/House	P/H	3 bits
	a_6	Family Income	FI	3 bits
	a_7	Has a Computer	Comp	1 bit
	a_8	Parents' Education	PE	3 bits
	a_9	School of Origin	SO	3 bits
	a_{10}	Self-Evaluation	SEv	3 bits
	a_{11}	Where From	WF	1 bit
	a_{12}	Dist. School-Residence	DistSR	3 bits
	a_{13}	Means of Transport	MT	3 bits
	a_{14}	Work	Wk	3 bits
	a_{15}	Study Shift	SS	2 bits
	a_{16}	Students/Classroom	S/C	3 bits
Output Vector (y)	y	Non-Evasion	NEv	1 bit
		Evasion	Ev	

desired output, in the training phase, represented by vector b (input of the module ARTb), being these ones described in the following way:

$\boldsymbol{a} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \boldsymbol{a}_3\ \ldots\ \boldsymbol{a}_{16}]$

$\boldsymbol{b} = [\ b\ ]$, where: $\boldsymbol{b}$ = "0" ou "1"

The subvectors $\boldsymbol{a}_1\ \boldsymbol{a}_2\ \boldsymbol{a}_3\ \ldots\ \boldsymbol{a}_{16}$ of the vector a (Table I) are lines vectors which contain the binary representation of the students' characteristics. Each bit corresponds to one component of the corresponding vector.

The network output is represented by the activity layer vector F2 (y) and provides answers in the binary coding with 1 bit, being that code "1" corresponds to students' evasion and code "0" to non-evasion, defined as follows:

y = [y] (Fuzzy ARTMAP network output)

The model proposed in this study consists of an intelligent system (flowchart shown in Fig. 1) for the study of students' evasion in the IFMT, using an Fuzzy ARTMAP Neural network [2-4], Logic Fuzzy and/or Dempster-Shafer's Theory of Evidence - TDS.

The information of the database is pre-processed and converted into a binary database. The essentially binary conception is considerably worthwhile, because the neural network presents a more efficient behavior (prompt and better quality of answers) and allows the extraction of knowledge in a continuous way (continued training), seeking for a better adaptation to the conditions of the institution and improvement with time.

In the phase of the neural analysis, if the answer is negative in relation to evasion, no action is adopted; just the register of the mentioned information is performed. If the answer of evasion is positive, the following step corresponds to a better discrimination about the quality of information (fine analysis) based on the use of Fuzzy module and/or of the Dempster-Shafer's Theory of Evidence. Later, solutions that aim to revert students' evasion will be proposed (proactive action).

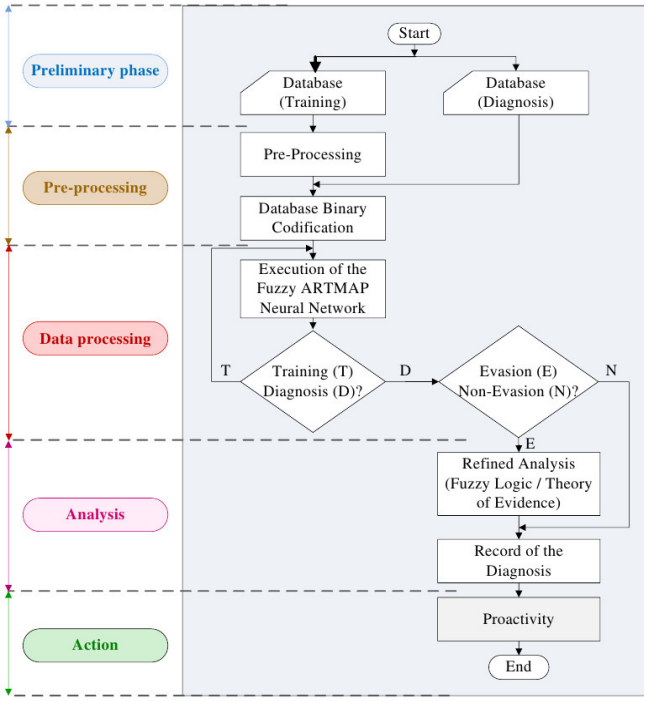


Fig. 1 Flowchart of the structure and sequence of development of the neural system proposed to perform the prediction of the evasion group risk

V. Application and Analysis of the Results

The intelligent system, using a Fuzzy ARTMAP, Neural Network proposed to make the prediction of the risk group of students prone to evasion, was implemented and tested with a database composed by 1.650 rows and 42 columns in the training phase of the network. In the validation and diagnosis phase of the network a sample with 499 lines and 41 columns was used, about 30% of the training samples. Each line represents the inputs standard vector and its corresponding desired output, in the training. The data of the columns from 1 to 41 represent the attributes correspondent to vector a, input of the module ARTa. In column 42 are represent the desired outputs, vector b (input of the module ARTb) of the Fuzzy ARTMAP neural network.

The parameters used in the database processing are specified in Table II.

After the network training five simulations were performed, based on data for the diagnosis, for the validation of the model proposed, being that, in one of them the samples were processed in a naturally way and the other in a randomized way.

The results of the processing were compared and analyzed, using a criterion, called "voting criterion" [7], "0" or "1" of higher incidence for each of the inputs. The result of higher incidence constitutes the output of the neural network.

Later, comparing the output from the network with the real situation of each sample of the group of students analyzed, it was possible to investigate the coincidence of the evasion ("1") and non-evasion ("0") among the samples processed and the reality.

After concluding the phases of the processing of database through an Fuzzy ARTMAP Neural Network and respective analyses necessary to the understanding of the behavior in relation to students' evasion and non-evasion, the results were compiled and, briefly, shown in Table III.

TABLE II. Specification of the Parameters: Fuzzy ARTMAP Network

Parameters and References Values	
Parameters	***Values***
Choice parameter ($\alpha > 0$)	0.001
Training rate ($\beta \in [0,1]$)	1.0
Vigilance parameter module ART_a ($\rho_a \in [0,1]$)	0.2
Increasing in the vigilance parameter ρ_a (ε)	0.05
Vigilance prameter module ART_b ($\rho_b \in [0,1]$)	0.999
Vigilance parameter module inter-ART_{ab} ($\rho_{ab} \in [0,1]$)	0.7
Vigilance parameter in the match tracking ($\rho_{amat} \in [0,1]$)	0.75

TABLE III. QUANTITATIVE AND PERCEPTUAL RESULTS OF THE DIAGNOSIS OF SCHOOL EVASION PREDICTION

Diagnosis of School Evasion	Quantitative and Percentages Values: Output of Network					
	Evasion		*Non-Evasion*		*Total of Samples*	
	Number	*%*	*Number*	*%*	*Number*	*%*
Samples	90	100	409	100	499	100
Corrects	88	97.8	295	72.1	383	76.7
Errs	2	2.2	114	27.9	116	23.3

The reading, interpretation and data analysis in Table III show that: of 499 samples, 90 of them corresponded to the evaded students and, 409 students who had concluded or attending a course, that is, not-evading. The proposed system identified 88 evasion possibilities and ignored 2, with a margin of success of 97.8%. Among the 409 samples of non-evasion, the Fuzzy ARTMAP network proposed recognized 295 samples in this situation and did not hit the target in 114, getting it right in 72.1% of the cases. It reached the global accuracy of 76.7%, finding correctly 383 samples of a total of 499.

The quantitative results of the previous diagnosis of the students with possibility of evasion can be perceived, more clearly in the graphs of Fig. 2.

Considering the experiment done and consistency of the results obtained, it can be inferred that the intelligent system, using Fuzzy ARTMAP, neural network proposed to identify the students prone to evasion, is a model with a significant degree of reliability and expresses accurately the situation in which the students analyzed are.

VI. CONCLUSION

This study presented an innovative method to identify, in a proactive, continued and accurate way, the students considered to belong to the risk group of school dropout, using Fuzzy ARTMAP neural network.

The analysis of the results showed that the proposed system is satisfactory, with global accuracy superior to 76%, and with a significant degree of reliability, making possible the early identification, even in the first term of the course, the group of students likely to drop out. The anticipated identification of this group of students enables the institutional education, alongside the multidisciplinary team to adopt strategic, proactive and individualized measures with the aim of reducing or even mitigating the students' evasion.

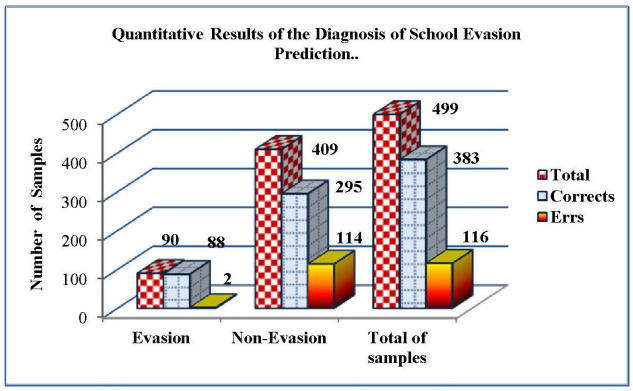


Fig. 2 Qualitative result of prediction of school evasion

REFERENCES

[1] V. Martinho, "Intelligent System for Prediction of Group Risk Evasion Student," PosGraduate Program in Electrical Engineering, Electrical Engineering Department, UNESP, Ilha Solteira - SP, 2012. (in Portuguese).

[2] G. A. Carpenter, S. Grossberg, and K. Iizuka, "Comparative performance measures of fuzzy ARTMAP, learned vector quantization, and back propagation for handwritten character recognition," International Joint Conference on Neural Networks, vol.1, pp. 794-799, 1992.

[3] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Transactions on Neural Networks, vol. 3, pp. 698-713, 1992.

[4] G. A. Carpenter and S. Grossberg, "A self-organizing neural network for supervised learning, recognition, and prediction," Communications Magazine, IEEE, vol. 30, pp. 38-49, 1992.

[5] S. Haykin, Neural Networks: Comprehensive Foundation: Prentice Hall, 1999.

[6] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," Neural networks and natural intelligence, Massachusetts Institute of Technology, pp. 213-250, 1988.

[7] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," Neural Network, vol. 4, pp. 759-771, 1991.

[8] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Transactions on Neural Networks, vol. 3, pp. 698-713, 1992.

[9] V. Martinho, C. Nunes, C. R. Minussi,, "A new method for prediction of school dropout risk group using neural network Fuzzy," XV International Conference on Artificial Intelligence, Las Vegas, 2013, in press.

[10] S. C. Marchiori, M. C. G. da Silveira, A. D. P. Lotufo, C. R. Minussi, and M. L. M. Lopes, "Neural network based on adaptive resonance theory with continuous training for multi-configuration transient stability analysis of electric power systems," Applied Soft Computing, vol. 11, pp. 706-715, 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 115–122

Semantic Explorative Evaluation of Document Clustering Algorithms

Hung Son Nguyen
Institute of Mathematics
The University of Warsaw
Banacha 2, 02-097, Warsaw Poland

Sinh Hoa Nguyen
Institute of Mathematics
The University of Warsaw
Banacha 2, 02-097, Warsaw Poland

Wojciech Świeboda
Institute of Mathematics
The University of Warsaw
Banacha 2, 02-097, Warsaw Poland

***Abstract*—In this paper, we investigate the problem of quality analysis of clustering results using semantic annotations given by experts. We propose a novel approach to construction of evaluation measure, which is based on the Minimal Description Length (MDL) principle. In fact this proposed measure, called SEE (Semantic Evaluation by Exploration), is an improvement of the existing evaluation methods such as Rand Index or Normalized Mutual Information. It fixes some of weaknesses of the original methods. We illustrate the proposed evaluation method on the freely accessible biomedical research articles from Pubmed Central (PMC). Many articles from Pubmed Central are annotated by the experts using Medical Subject Headings (MeSH) thesaurus. This paper is a part of the research on designing and developing a dialog-based semantic search engine for SONCA system[1] which is a part of the SYNAT project[2]. We compare different semantic techniques for search result clustering using the proposed measure.**

I. Introduction

CLUSTERING can be understood as an unsupervised data mining task for finding groups of points that are close to each other within the cluster and far from the rest of clusters. Intuitively, the greater the similarity (or homogeneity) within a cluster, and the greater the difference between groups, the "better" the clustering. Clustering is a widely studied data mining problem in the text domains, particularly in segmentation, classification, collaborative filtering, visualization, document organization, and semantic indexing.

It is a fundamental problem of unsupervised learning approaches that there is no generally accepted "ground truth". As clustering searches for previously unknown cluster structures in the data, it is not known a priori which clusters should be identified. This means that experimental evaluation is faced with enormous challenges. While synthetically generated data is very helpful in providing an exact comparison measure, it might not reflect the characteristics of real world data.

In recent publications, labeled data, usually used to evaluate the performance of classifiers, i.e. supervised learning algorithms, is used as a substitute [18], [6], [1]. While this provides the possibility of measuring the performance of clustering algorithms, the base assumption that clusters reflect the class structure is not necessarily valid.

[1]Search based on ONtologies and Compound Analytics

[2]Interdisciplinary System for Interactive Scientific and Scientific-Technical Information (www.synat.pl)

Some approaches therefore resort to the help of domain experts in judging the quality of the result [2], [7], [6]. When domain experts are available, which is clearly not always the case, they provide very realistic insights into the usefulness of a clustering result. Still, this insight is necessarily subjective and not reproducible by other researchers. Moreover, there is not sufficient basis for comparison, as the clusters that have not been detected are unknown to the domain expert.

There have been several suggestions for a measure of similarity between two clusterings. Such a measure can be used to compare how well different data clustering algorithms perform on a set of data. These measures are usually tied to the type of criterion being considered in evaluating the quality of a clustering method [10].

The goal of clustering is to assign objects to subsets which are coherent internally, but are dissimilar to each other. These goals are usually explicitly formulated as *internal criteria of clustering quality*. The word "internal" highlights the fact that they are based on object similarity expressed in the original feature space. Usually it may not necessarily be clear whether modeling assumptions in the underlying model (feature space and e.g. the notion of distance between objects) are valid. Hence, one may ask to validate or evaluate a clustering in a specific application, using feedback from users or experts. When a "gold standard" clustering is provided by experts, one may compare it with the result from a clustering algorithm. This approach is an *external criterion of clustering quality*.

The problem becomes even more complicated in evaluation of text clustering with respect to semantic similarity, whose definition is not precise and highly contextual. As the number of results is typically huge, it is not possible to manually analyze the quality of different algorithms or even different runs of the same algorithm.

The remainder of this paper is structured as follows. In Section II we present some basic notions and problem statement. This is followed by an overview of external clustering evaluation methods in Section III. In Section IV we present the basic semantic evaluation techniques and propose a novel evaluation method based on exploration of expert's tags which is the fundamental contribution of this paper. After this we use the proposed evaluation method to analyze the search result clustering algorithms over the document collection publish by PubMed. An analysis of the methods result representation and their interpretability is presented in Section V, followed by some conclusions and lessons learned in Section VI.

978-83-60810-53-8/$25.00

TABLE I. ILLUSTRATION OF HARD CLUSTERING DEFINED BY AN ALGORITHM AND BY AN EXPERT.

Doc.	Hard Cluster			Expert Cluster			
	C_1	C_2	C_3	E_1	E_2	E_3	E_4
d_1	1						1
d_2	1			1			
d_3		1					1
d_4		1				1	
d_5			1		1		
d_6			1			1	

TABLE II. ILLUSTRATION OF SOFT CLUSTERING FOUND BY AN ALGORITHM AND DEFINED BY AN EXPERT.

Doc.	Soft Cluster			Expert Tag				
	C_1	C_2	C_3	Cosmonaut	astronaut	moon	car	truck
d_1	1			1		1	1	
d_2	1				1	1		
d_3	1	1		1				
d_4		1	1				1	1
d_5		1	1				1	
d_6			1					1

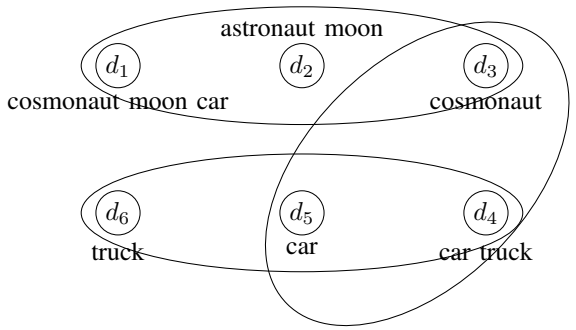


Fig. 1. Illustration of clustering from example Table II.

II. PROBLEM STATEMENT

A hard clustering algorithm is any algorithm that assigns a set of objects (e.g. documents) to disjoint groups (called clusters). A soft clustering relaxes the condition on target clusters being disjoint and allows them to overlap. Clustering evaluation measures[15], [10] proposed in the literature can be categorized as either *internal criteria of quality* or *external criteria of quality*.

An *internal criterion* is any measure of "goodness" defined in terms of object similarity. These criteria usually encompass two requirements – that of attaining high intracluster similarity of objects and high dissimilarity of objects in different clusters.

External criteria on the other hand compare a given clustering with information provided by experts. Typically in the literature it is assumed that both the clustering provided by studied algorithm and the clustering provided by experts are hard clusterings. We believe that the requirement that expert knowledge is described in terms of a hard clustering is overly restrictive. In typical applications in text mining, one faces datasets which are manually labeled by experts, but with each document being assigned a set of tags. We can think of such tags as of soft clusters. In this paper we aim to provide measures of external evaluation criteria that relax both conditions on the clustering and expert clusterings being partitions (hard clusterings).

In what follows, our focus is on clustering of documents, hence we will occasionally use terms "object" and "document" interchangeably. We stress that the notion of Semantic Explorative Evaluation (SEE) introduced in this paper is a general framework which can be used to evaluate clustering of objects of an arbitrary type.

Typically in the literature it is assumed that the input data to clustering evaluation can be described in a form similar to Table I, i.e. with exactly one valid cluster C_i and exactly one valid expert cluster E_j for each document.

We will relax this condition to allow comparison of soft clustering and a set of expert tags assigned to each document, thus allowing input data as in Table II.

III. OVERVIEW OF CLUSTERING EVALUATION METHODS

In this section we briefly review external evaluation criteria typically used in clustering evaluation. We assume that two partitions (hard clusterings) of objects are given: one by an algorithm, and another one provided by domain experts. Most external evaluation criteria can be naturally grouped in two groups:

- *Pair-counting measures*, which are defined on a 2×2 contingency matrix that summarizes similarity of pairs of objects w.r.t. both clusterings (see Table III):
  If there are k objects in the dataset, then
  $$a + b + c + d = \binom{k}{2}$$
  A typical measure that can be expressed in terms of these numbers is
  $$\textit{Rand Index} = \frac{a+d}{a+b+c+d}.$$
  However, there is a multitude of different variants of other similar measures. Pfitzner et al.[15] provide an overview of 43 measures that all fit into this scheme.
- *Information-theoretic measures* on the other hand compare distributions of $c(D)$ and $e(D)$, which denote respectively the cluster and the expert label (which induces a partition) assigned to a document D drawn randomly from the dataset. These measures can be expressed in terms of joint distribution of $\langle c(D), e(D)\rangle$, i.e. simply by counting objects belonging to each pair $\langle C_i, E_j\rangle$ as shown in Table IV. Numbers in brackets denote expected values of counts assuming independence of $c(D)$ and $e(D)$. Information-theoretic measures thus aim to to measure the degree of dependence between these two. An example such

TABLE III. ALL PAIR-COUNTING MEASURES CAN BE SUMMARIZED IN TERMS OF NUMBERS a, b, c, d IN THIS TABLE.

Pairs of documents		Same cluster? True	Same cluster? False
Same expert tag?	True	a	b
	False	c	d

TABLE IV. INFORMATION-THEORETIC MEASURES ARE DEFINED IN TERMS OF CONTINGENCY TABLE SHOWN HERE (FOLLOWING EXAMPLE IN TABLE I).

	C_1	C_2	C_3	Total
E_1	1	0	0	1
E_2	0	0	1	1
E_3	0	1	1	2
E_4	1	1	0	2
	2	2	2	

measure is *mutual information* I between $c(D)$ and $e(D)$, where

$$I(X,Y) = \sum_x \sum_y p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right).$$

A measure typically used in clustering evaluation is *Normalized Mutual Information*[10], though [15] reviews 13 different measures, all defined quite similarly.

Purity is a measure occasionally used as an external evaluation criterion. While it is not strictly an information-theoretic measure, it can be also expressed in terms of table IV.

IV. SEMANTIC EVALUATION METHODS FOR SOFT CLUSTERING

We stress two limitations of measures proposed in the literature and briefly reviewed thus far:

- The first limitation, already mentioned in the previous section, is the typical assumption that both the clustering algorithm and the experts provide partitions. We will show how all measures mentioned (whether directly or indirectly) in the previous section can be naturally extended to the case of comparing a soft clustering with expert knowledge expressed in terms of multiple tag assignment. In the first part of this section we will briefly review our proposed solution to this problem, described earlier in [12].
- A more important limitation, though, is that neither of these measures described so far resemble the thought process that the expert himself would undergo if he was faced with the task of manually evaluating a clustering. In the second part of this section we will describe a novel method of semantic evaluation that addresses this issue.
  This is the fundamental contribution of this paper.
- The third problem, that we address further in the paper, is that of comparing different clusterings.
- Finally, methods mentioned so far do not allow us to compare different clusters of a single clustering.

A. Comparing set covers.

Previous works by other authors on this problem include [3] (Fuzzy Clustering Mutual Information) and [8] (comparing set covers).

In this section we consider two types of measures defined earlier separately.

TABLE V. PAIR-COUNTING MEASURES CAN BE NATURALLY DEFINED FOR SOFT COVERS IF WE SUBSTITUTE *hard membership* (SEE TABLE III) BY THE NOTION OF *similarity*.

Pairs of documents		Cluster-similar? True	Cluster-similar? False
Expert-similar?	True	a	b
	False	c	d

TABLE VI. INFORMATION-THEORETIC MEASURES CAN BE DEFINED IF WE CAN DESCRIBE THE JOINT DISTRIBUTION OF CLUSTERS AND EXPERT LABELS (SEE EXAMPLE IN TABLE II).

	C_1	C_2	C_3
Cosmonaut	0.139	0.083	0
astronaut	0.083	0	0
moon	0.139	0	0
car	0.056	0.125	0.125
truck	0	0.042	0.208

First we describe how to extend a pair-counting measure of similarity of two partitions to a measure of similarity of set covers. Pair-counting measures only pose a tiny problem. Looking at table III, we see that for soft clusterings, rows and columns are not well defined. In order to fully characterize a pair of documents $\langle d_i, d_j \rangle$, we proposed in [12] to define notions of cluster-similarity and expert-similarity for documents and base pair-counting measures on table V. This approach naturally extends any pair-counting measure, with the focus of our prior experiments on Rand Index [16]. We defined very simple notions of similarity: we considered two documents d_i, d_j θ-expert-similar, if $\frac{|e(d_i) \cap e(d_j)|}{|e(d_i) \cup e(d_j)|} \geq \theta$, and we defined θ-cluster-similarity in the same way. This approach allows us to effortlessly apply each of the 43 pair-counting measures reviewed by Pfitzner[15].

Information-theoretic measures can be extended by counting a given document in multiple cells of Table IV whenever the document is in multiple clusters and/or multiple tags are assigned to the document. If we wish to assign an overall equal weight to each document, instead of raw counts, one may further assume that the contribution of a document is inversely proportional to the number of cells that it contributes to. This has a straightforward probabilistic interpretation. The original measures, like $I(c(D), e(D))$ are defined for deterministic functions c and e and a random document D. In the proposed extension, c and e are also random variables, with $c(d)$ uniformly distributed across clusters containing document d, and $e(d)$ uniformly distributed across tags assigned to document d. Original formulas themselves, like $I(c(D), e(D))$ remain unchanged. This approach is illustrated in Table VI.

B. Semantic Explorative Evaluation

We have mentioned that the calculation of neither of the measures reviewed so far resembles human reasoning. We propose a different approach to the problem of semantic evaluation.

If an expert faced the problem of manual inspection of clustering results, he would try to explain (describe) the contents of clusters in his own words (i.e. in terms of expert tags). In essence, a cluster should be valid for an expert if the expert can briefly explain its contents. The expert would find

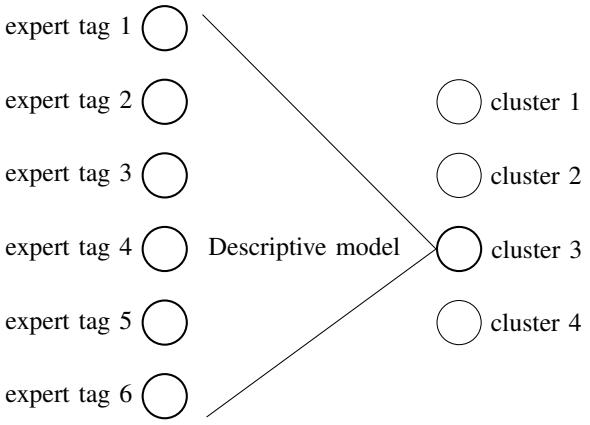


Fig. 2. For each cluster (e.g. cluster 3), we build a model describing the contents of this cluster in terms of expert tags. A measure of complexity of the resulting model thus corresponds to semantic validity of the cluster.

a set of clusters valid if he could provide a short explanation for each cluster.

In order to define a measure of semantic validity which is based on this reasoning, we need to specify three things:

- description of clusters in terms of expert concepts (i.e. a model family),
- define the length of such an explanation so that we know if it is short,
- a penalty incurred if a cluster is indescribable in terms of expert concepts,
- define the aggregate measure, so that we can evaluate a set of clusters.

We specify these three ingredients as follows:

- The explanation or description of a cluster is in essence a model of the cluster in terms of expert tags. Any classification algorithm provides such a model. The exact choice of the classifier is of secondary importance as long as the same procedure is consistently used to evaluate different clusterings. In our experiments, the classifier of choice is a decision tree with no pruning, with splits defined greedily using Gini index.
- By appealing to Minimum Description Length principle, one may then define a measure of validity of a fixed cluster as the complexity of the model describing the cluster. The measure of model complexity that we use is the average depth of the resulting tree.
- For simplicity we omit a penalty for indescribable clusters at this point, although we guarantee during tree construction that resulting leaves in descision trees contain either objects from the same decision class or objects that are indiscernible given the information about expert tags alone.
- We define the measure of validity of a clustering as the average validity of clusters.

The pseudocode of the presented idea is presented below in Algorithm 1.

Algorithm 1: SEE – Semantic Explorative Evaluation.

Input:
- $\mathbf{C} = \{C[i,j] : i = 1,\ldots,k \text{ and } j = 1,\ldots,n\}$: the document–cluster assignment matrix.
- $\mathbf{E} = \{E[i,j] : i = 1,\ldots,k \text{ and } j = 1,\ldots,m\}$: the expert–cluster assignment matrix.
- L: a decision tree construction algorithm.

Output: m: the average mean depth of decision trees describing clusters.

for $j = 1,\ldots,n$ **do**
  Construct a decision table
  $$H_j := \left[E; [C_{1,j},\ldots,C_{k,j}]^T\right]$$
  // H_j is the decision table constructed from the matrix $\mathbf{E}$ augmented with the j-th column of matrix $\mathbf{C}$ at the end as the decision variable.
  $T_j := L(H_j)$;
  // Construct the decision tree T_j by applying algorithm L on decision table H_j.
  $m_j = \text{MeanDepth}(T_j)$
end
Return $m = \frac{m_1+\ldots+m_n}{n}$;

C. Semantic Explorative Evaluation: Example

In this Section we demonstrate the proposed evaluation method (presented above in Algorithm 1) on the example introduced in Table II.

This table summarizes a small text corpus consisting of just 6 documents. Half of these documents, forming cluster C_2, concern vehicles: cars and trucks, whereas the other half concerns cosmonauts and moon: these documents form cluster C_1. Cluster C_1 is the easiest one to explain for the expert: he associates either the concept 'cosmonaut' or 'astronaut' with each document from this cluster. Document d_1 concerns a lunar rover and is an interesting "outlier" that needs to be explicitly excluded from cluster C_3 by the expert: the branch on attribute "moon" in decision tree describing cluster C_3 explicitly addresses this case.

The constructed decision trees T_1, T_2, T_3 for clusters C_1, C_2, C_3 are presented in Fig. 3, Fig. 5 and Fig. 4, respectively. According to those trees, cluster C_3 seems to be "hardest" to explain by the expert. Hence the average depth or weighted average depth of the decision tree T_3 are also higher than for T_1.

Depths of decision trees describing clusters C_1, C_2, C_3 are $1\frac{2}{3}$, 2 and $2\frac{1}{4}$, respectively. Thus SEE of the clustering equals approximately 1.97.

Doc.	Expert Tag					decision
	Cosm.	astron.	moon	car	truck	C_1
d_1	1		1	1		1
d_2		1	1			1
d_3	1					1
d_4				1	1	
d_5				1		
d_6					1	

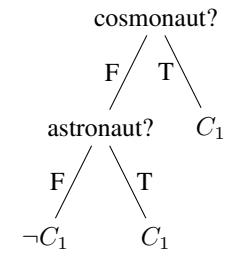


Fig. 3. The decision table H_1 (above) and the decision tree describing cluster C_1 constructed from H_1. Cluster C_1 is the easiest for the expert to explain.

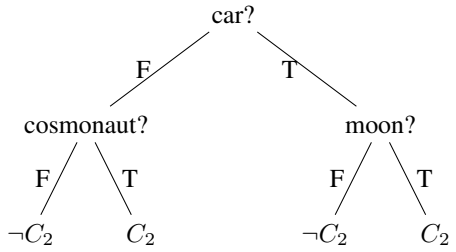


Fig. 4. Cluster C_2 is not easily describable in expert terms.

D. Randomization

The last problem we aim to address is that of comparing different clusterings. With all evaluation methods, either reviewed or introduced in this article, we face the same issue when we aim to compare different clusterings: we lack an explanation why one clustering may be better than the other one. In this section, we introduce a trick which allows us to isolate a specific sub-problem solved by a clustering algorithm, to which we can assign a measure of quality that is easily interpretable.

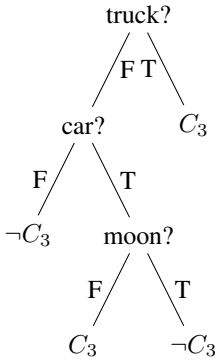


Fig. 5. Cluster C_3 does not contain document d_1, which concerns a very specific type of a car – a moon rover. This outlier forces the expert to provide a longer explanation to explicitly "remove" this object.

In what follows, we will think of a clustering algorithm as of a procedure that solves two sub-problems. For hard clustering these are:

- Determining the structure of clusters, i.e. the number of clusters and the number of documents belonging to each cluster.
- The assignment of documents to clusters, while preserving constraints on the structure.

For soft clustering, these two sub-problems are:

- Determining the structure of clusters is actually determining the number of clusters K as well as the joint (rather than the marginal) distribution of the number of documents in each cluster.
- Instead of assigning documents to clusters, an algorithm assigns documents to each of the 2^K possible partitions.

In what follows, we will focus on measuring the quality of a clustering algorithm w.r.t. the second sub-problem, while ignoring the first sub-problem. The idea is to randomize the assignment of documents to clusters while keeping the structure of clusters fixed and calculate the value of m for such randomized assignments so as to determine a meaningful "basis" or benchmark for comparison. Each measure m can thus be transformed into an m-quantile measure, which basically says how often a clustering algorithm outperforms a random assignment, while solving the second sub-problem.

V. The results of experiments

The following experiments are the continuation of our previous studies in [13], [11], [12], although in this work they merely serve as an illustration of the discussed and introduced measures.

A. Experiment Set-Up

We have applied the model-based semantic evaluation measure introduced in this paper to study clusterings induced by different document representations (lexical, semantic and structural) and using different algorithms. The document repository in our study is a subset of PubMed Central Open Access Subset[17].

TABLE VII. An example of tags assigned to the paper: "PubMed Central: The GenBank of the published literature." by Roberts R. J. ([17])

heading	subheading
Internet	
MEDLINE	economics
Periodicals as Topic	economics
Publishing	economics

The majority of documents in PubMed Central are tagged by human experts using headings and (optionally) accompanying subheadings (qualifiers) from a MeSH controlled vocabulary [20]. A single document is typically tagged by 10 to 18 heading-subheading pairs. The example of tagged document is shown in Table VII

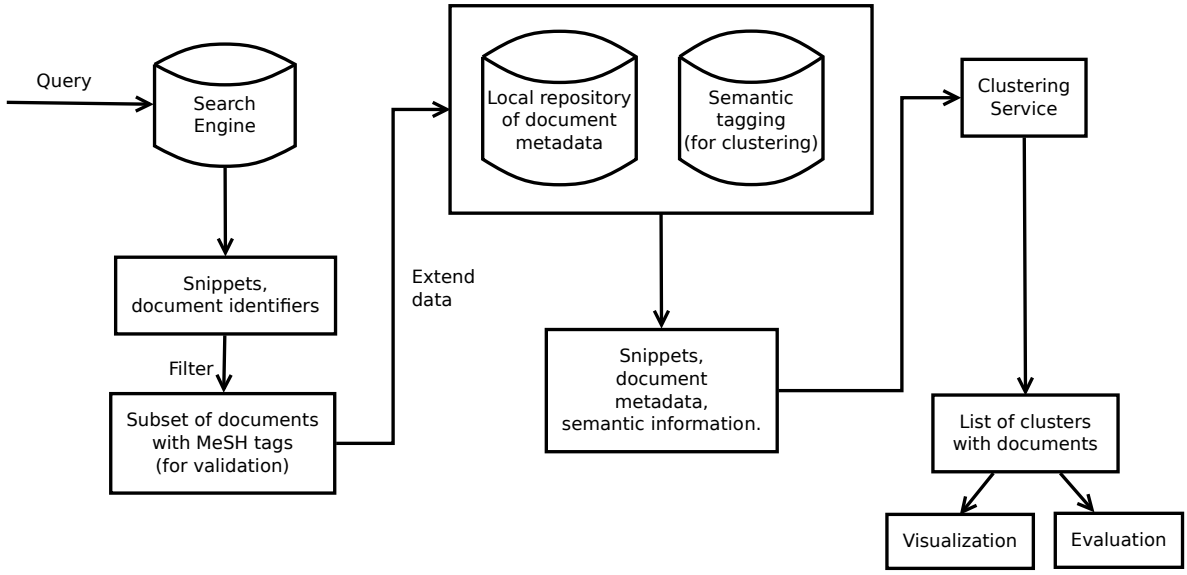


Fig. 6. Experiment diagram.

There are approximately 25000 subject headings and 83 subheadings. There is a rich structural information accompanying headings: each heading is a node of (at least one) tree and is accompanied by further information (e.g. allowable qualifiers, annotation, etc.). Currently we do not use this information, but in some experiments we use a hierarchy of qualifiers[3] by exchanging a given qualifier by its (at most two) topmost ancestors or roots.

TABLE VIII. THE NUMBERS OF POSSIBLE TAGS IN PUBMED CENTRAL OPEN ACCESS SUBSET[17]

source	**possible tags**	**expert tags assigned to example document**
headings	~ 25000	Internet, MEDLINE, Periodicals as Topic, Publishing
subheadings	83	economics
subheading roots	23	organization & administration

The choice of expert tags determines how precisely we wish to interpret expert opinion. In experiments that we describe in this paper, we interpreted subheadings as the expert tags.

The diagram of our experiments is shown in Figure 6. An experiment path (from querying to search result clustering) consists of three stages:

- Search and filter documents matching to a query. Search result is a list of *snippets* and document identifiers. Usually more than 200 documents are returned for a single query. The result set is then truncated to the top 200 most relevant (in terms of TF-IDF) documents.
- Extend representations of snippets and documents by *citations* and/or *semantically similar concepts* from MeSH ontology (these MeSH terms were automatically assigned by an algorithm[19], whereas MeSH subheadings used for evaluation were manually assigned by human experts).
- Cluster document search results.

In our experiments, we worked with three clustering algorithms: K-Means[9], Lingo[14] and Hierarchical Clustering[4].

In order to perform evaluation (and choose parameters of clustering algorithms) one needs a set of search queries that reflect actual user usage patterns. We extracted a subset of most frequent one-term queries from the daily log previously investigated by Herskovic et al. in [5] and retrieved relevant documents from PubMed Central Open Access Subset.

Roughly one fourth of these result sets was used for initial fine-tuning of parameters, whereas the remaining 71 queries were further used in evaluation.

B. Experiment results

We need to stress that we used subheadings as the source of expert tags used for semantic evaluation. There are only 83 possible subheadings in MeSH vocabulary, hence the granularity of information provided for evaluation is very limited. We have not applied pruning to resulting decision trees (the goal of algorithm Algorithm 1 is merely to provide a description, not a model for inference), and the resulting decision trees are somewhat deep, as can be seen from Figure 7.

Nevertheless, as we can see from Figure 8, the m-quantile measure is usually below 0.5 (SEE-quantile value 0.5 corresponds to a random document-to-cluster assignment). Furthermore, result sets for different queries visibly differ in how "hard" they are to cluster: m-quantile measures of different algorithms are significantly correlated. Figure 7 should not be directly interpreted in this way due to different structure of result sets corresponding to different queries (e.g. different number of documents).

[3] http://www.nlm.nih.gov/mesh/subhierarchy.html

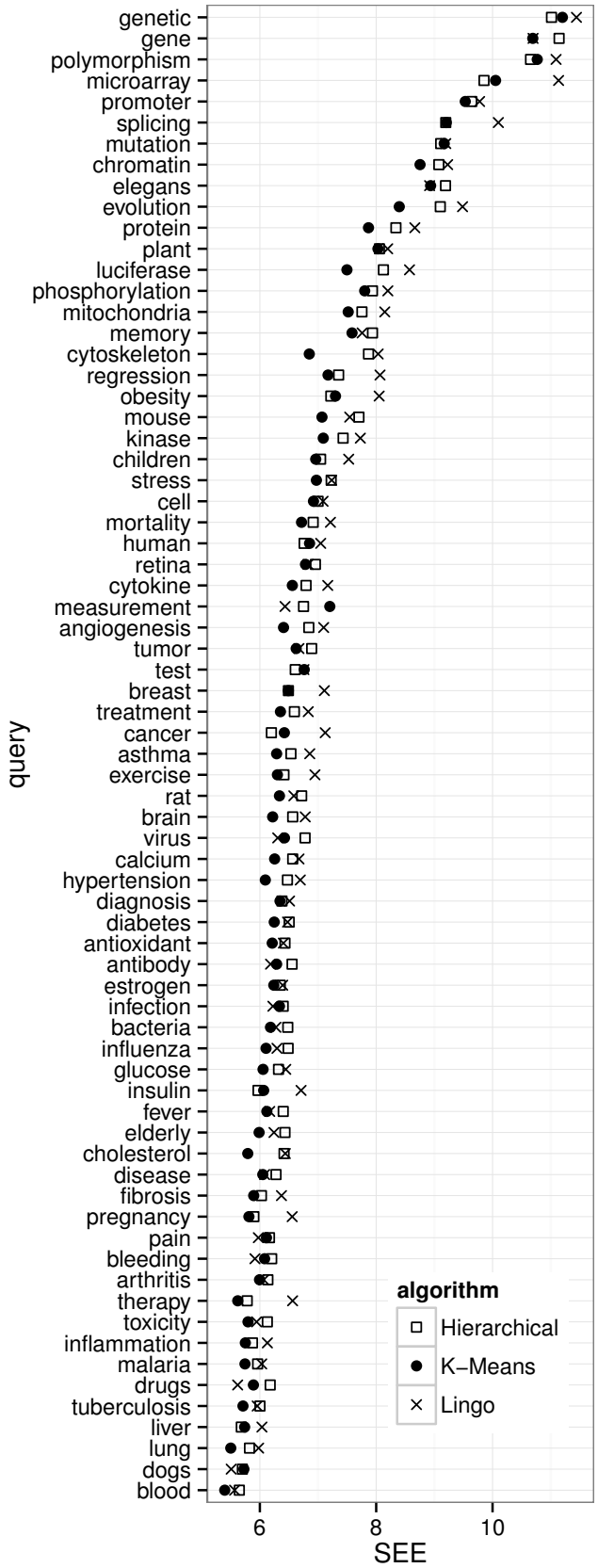


Fig. 7. Average tree depth for different result sets and clustering algorithms.

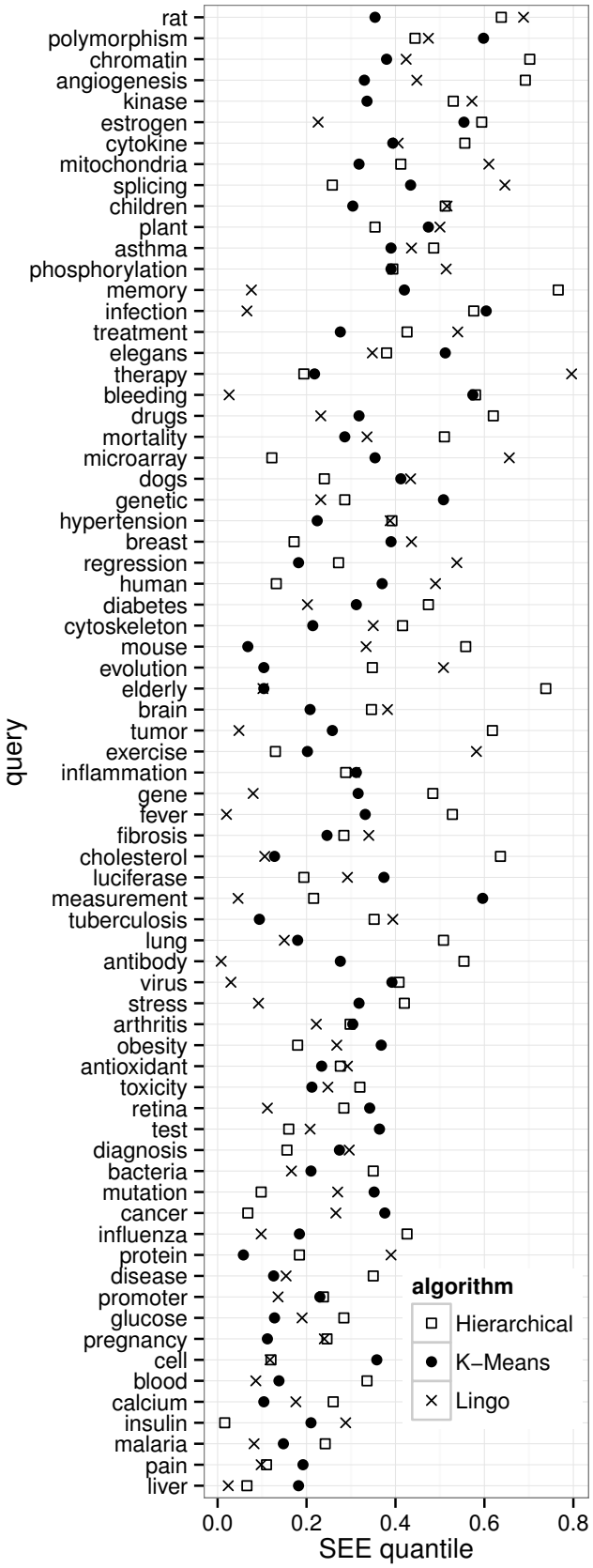


Fig. 8. SEE-quantile measure for different clusterings and queries.

VI. Conclusions and future plans

In this paper we have introduced a novel paradigm of semantic evaluation. Unlike traditional approaches, which are either measures counting pairs of objects or are variations of information theoretic approaches, our proposed procedure resembles the process of human perception, as it is based on a model describing the clustering in terms of expert knowledge. We proposed a specific implementation of this evaluation measure (i.e. a choice of the underlying model structure and optimization framework) and further demonstrated its application to online results clustering evaluation problem. We have observed that even if we only used information about MeSH subheadings assigned to documents as the source of information for evaluation, for most result sets in our experiments clusterings performed better than random assignments of documents to clusters. Furthermore, we have observed that some result sets are inherently harder to cluster than others, and the performance of analyzed clustering algorithms is usually correlated.

VII. Acknowledgements

The authors are supported by grants 2011/01/B/ST6/03867 and 2012/05/B/ST6/03215 from the Polish National Science Centre (NCN), and the grant SP/I/1/77065/10 in frame of the strategic scientific research and experimental development program: "Interdisciplinary System for Interactive Scientific and Scientific-Technical Information" founded by the Polish National Centre for Research and Development (NCBiR).

References

[1] I. Assent, R. Krieger, E. Müller, and T. Seidl. Visa: visual subspace clustering analysis. *SIGKDD Explor. Newsl.*, 9(2):5–12, Dec. 2007.

[2] C. Böhm, K. Kailing, H.-P. Kriegel, and P. Kröger. Density connected clustering with local subspace preferences. In *ICDM*, pages 27–34. IEEE Computer Society, 2004.

[3] T. Cao, H. Do, D. Hong, and T. Quan. Fuzzy named entity-based document clustering. In *Proc. of the 17th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2008)*, pages 2028–2034, 2008.

[4] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.

[5] J. R. Herskovic, L. Y. Tanaka, W. Hersh, and E. V. Bernstam. A day in the life of pubmed: analysis of a typical day's query log. *Journal of the American Medical Informatics Association*, pages 212–220, 2007.

[6] H.-P. Kriegel, P. Kroger, M. Renz, and S. Wurst. A generic framework for efficient subspace clustering of high-dimensional data. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, ICDM '05, pages 250–257, Washington, DC, USA, 2005. IEEE Computer Society.

[7] P. Kröger, H.-P. Kriegel, and K. Kailing. Density-connected subspace clustering for high-dimensional data. In M. W. Berry, U. Dayal, C. Kamath, and D. B. Skillicorn, editors, *SDM*. SIAM, 2004.

[8] A. Lancichinetti, S. Fortunato, and J. Kertész. Detecting the overlapping and hierarchical community structure of complex networks. *New Journal of Physics*, 11:033015, March 2009.

[9] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. L. Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[10] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval.* 2007.

[11] S. H. Nguyen, W. Świeboda, and G. Jaśkiewicz. Extended document representation for search result clustering. In R. Bembenik, Ł. Skonieczny, H. Rybiński, and M. Niezgódka, editors, *Intelligent Tools for Building a Scientific Information Platform*, volume 390 of *Studies in Computational Intelligence*, pages 77–95. Springer-Verlag New York, 2012.

[12] S. H. Nguyen, W. Świeboda, and G. Jaśkiewicz. Semantic evaluation of search result clustering methods. In R. Bembenik, L. Skonieczny, H. Rybinski, M. Kryszkiewicz, and M. Niezgodka, editors, *Intelligent Tools for Building a Scientific Information Platform*, volume 467 of *Studies in Computational Intelligence*, pages 393–414. Springer, 2013.

[13] S. H. Nguyen, W. Świeboda, G. Jaśkiewicz, and H. S. Nguyen. Enhancing search results clustering with semantic indexing. In *SoICT 2012*, pages 71–80, 2012.

[14] S. Osinski, J. Stefanowski, and D. Weiss. Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent Information Systems*, pages 359–368, 2004.

[15] D. Pfitzner, R. Leibbrandt, and D. Powers. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowl. Inf. Syst.*, 19(3):361–394, May 2009.

[16] W. M. Rand. Objective criteria for the evaluation of clustering methods. *J. Amer. Stat. Assoc.*, 66(336):846–850, 1971.

[17] R. J. Roberts. PubMed Central: The GenBank of the published literature. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):381–382, Jan. 2001.

[18] K. Sequeira and M. Zaki. SCHISM: A new approach for interesting subspace mining. In *Proceedings of the fourth IEEE conference on Data Mining*, pages 186–193. IEEE Computer Society, 2004.

[19] M. Szczuka, A. Janusz, and K. Herba. Semantic clustering of scientific articles with use of dbpedia knowledge base. In R. Bembenik, Ł. Skonieczny, H. Rybiński, and M. Niezgódka, editors, *Intelligent Tools for Building a Scientific Information Platform*, pages 61–76. Springer-Verlag New York, 2012.

[20] United States National Library of Medicine. Introduction to MeSH – 2011. Available online, 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 123–128

Vickrey-Clarke-Groves for privacy-preserving collaborative classification

Anastasia Panoui
School of Electronic, Electrical and Systems Engineering
Loughborough University, UK
Email: a.panoui@lboro.ac.uk

Sangarapillai Lambotharan
School of Electronic, Electrical and Systems Engineering
Loughborough University, UK
Email: S.Lambotharan@lboro.ac.uk

Raphael C.-W. Phan
Faculty of Engineering
Multimedia University, Malaysia
Email: raphael@mmu.edu.my

***Abstract*—The combination of game theory and data mining opens new directions and opportunities for developing novel methods for extraction of knowledge among multiple collaborative agents. This paper extends on this combination, and motivated by the work of Nix and Kantarcioglu employs the Vickrey-Clarke-Groves (VCG) mechanism to achieve privacy-preserving collaborative classification. Specifically, in addition to encouraging multiple agents to share data truthfully, we facilitate preservation of privacy. In our model, privacy is accomplished by allowing the parties to supply a controlled amount of perturbed data, instead of randomised data, so long as this perturbation does not harm the overall result of classification. The critical point which determines when this perturbation is harmful is given by the VCG mechanism. Our experiment on real data confirms the potential of the theoretical model, in the sense that VCG mechanism can balance the tradeoff between privacy preservation and good data mining results.**

I. Introduction

DATA mining provides a range of useful tools for data manipulation and extraction of meaningful information from large data sets, that can improve our lives. For example, collaboration among hospitals and other healthcare institutions by providing the medical record sets, and thus creating a large database, can lead to better and more reliable research results. In a different scenario, markets can share their data related to the customers' shopping preferences, in order to make future product deals and offers that will increase the income. Furthermore, cooperation in international level among governments, by merging intelligence data sets, might result in strengthening the security against terrorism. However, in all cases it is important to ensure that sensitive information must remain hidden and not be disclosed.

This paper addresses the problem of privacy preserving collaborative data mining, motivated by a paper by Nix and Kantarcioglu [1]. A brief description of the setting is as follows: a number of participants, also called agents, jointly supply their individual data sets in order to perform a data mining task and extract information from the large database that is formed. As the trustworthiness of the agents is not guaranteed, it is necessary to add incentives for good behaviour. One approach is to have penalising strategies that will prevent inappropriate behaviour. However, game theory offers a solution with positive incentives. Our work, as in [1], employs a method from a branch of game theory, called mechanism design. More specifically, we use the Vickrey-Clarke-Groves (VCG) mechanism, in which the payoff of each agent contains the agent's contribution to the 'community'. Thus, if an agent's contribution harms the overall result, this agent will be charged and hence receive low payoff. Following the setting of [1] we also choose the data mining task to be classification. However, in contrast to [1], for simulation of an agent who supplies falsified data we modify the complete data set of the agent through a controlled amount of perturbation, rather than random perturbation of certain percentage of the data. Furthermore, apart from complete randomization of the data, which corresponds to the action of an agent who lies or an agent who aims for the maximum possible privacy, we also include small deviation from the true data. The latter action models an agent who wishes to preserve the privacy of his data without damaging the overall result. We show that this strategy results in information gain while keeping the agent's data private.

This work has been supported by the Engineering and Physical Science Research Council (EPSRC) of UK, under the grant EP/J020389

II. Related work

The combination of data mining and game theory in a collaborative environment has opened a new direction for research. Halpern and Teague [2] address the problem of secret sharing and multi-party computation, under the assumption that the agents are rational, rather that being good or bad. They show that there exists a randomised secret sharing scheme in which the agents reach a Nash equilibrium that overcomes the iterated deletion of weakly-dominated strategies. In [3] Abraham at al. extend the work of [2], by introducing the notion of k-resilient equilibrium, which is similar to the Nash equilibrium, but instead of tolerating deviation from one player, it tolerates deviations by coalitions with at most k members. Examination and analysis of the multi-party computation, and specifically of the secure sum computation problem under a game theoretic framework can be found in [4]. In many scenarios, in order to simulate real world situations the involved parties are divided into good or bad.

978-83-60810-53-8/$25.00

However, under a game theoretic framework this approach is often replaced by settings where the participants are assumed to be rational, whose aim is to maximise their gain. In this context, the authors of [5] introduce the notion of rational secure computation and show that the ballot-box can be used to securely compute any function. Although security is an important issue to be addressed, the behaviour of the participants must also be examined. Thus, in order to discourage improper behaviour, [4], [6], [7] introduce penalising methods. In particular, assuming semi-honest players, [6] is concerned with the problems that arise in a sovereign information sharing setting. The goal is to ensure that the participants learn the result from the task on the shared information, without gaining any knowledge about the shared data. This is achieved by using an auditing device that will repeatedly check the players' actions, penalising inappropriate behaviour. Punishing strategies against malicious players is also examined in [7], in a setting which includes verification of the results, in addition to the information sharing. A different approach to punishing policies in order to achieve good behaviour is the use of VCG mechanism [1], [8]. In [8] this particular mechanism is employed for regression learning and in [1] for classification.

III. Mechanism Design

Mechanism design is a branch of game theory concerned with the problem of social welfare [9]–[11]. The setting involves a set of I agents, each one having their own private preferences on a set of alternatives, and a principal, whose role is to ensure that the rules of the mechanism will be followed. The aim of the mechanism is to help the agents make a collective choice that is beneficial for all. Formally, a mechanism is a collection of strategy sets $S_1, \ldots, S_I$ and an outcome function $g : S_1 \times \ldots \times S_I \to X$, where X is a set of possible alternatives. Each alternative is associated with a utility function $u_i(x)$ (known also as payoff), which denotes the gain of agent i when alternative x is chosen. As different alternatives lead to different payoffs, clearly each agent has a different preference on the alternatives. In order to model the distinctiveness of the agent's preferences, we associate each agent with a type θ_i, $i = 1, \ldots, I$. An important point is that the preference, and hence the type of each agent is private information and hence θ_i is known only to agent i. For this reason, in the game theoretic context, we are in an environment of incomplete information.

Once the agents have decided upon the preferences and their type has been determined, they report types $\hat{\theta}_i$, which might or might not coincide with θ_i (direct revelation mechanism). After $\hat{\theta}_i$ has been announced from all agents, the mechanism selects the collective choice to be

$$k^*(\hat{\theta}) = \arg\max_{k \in K} \sum_i v_i(k, \hat{\theta}_i),$$

where K is the set of possible choices, $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_I)$ and $v_i(k, \hat{\theta}_i)$ is the valuation of agent i on the choice k, when his reported type is $\hat{\theta}_i$.

A. The Vickrey-Clarke-Groves Mechanism

The Vickrey-Clarke-Groves mechanism (denoted by VCG) is a mechanism where the utility function has the following quasi-linear form:

$$u_i(x, \theta_i) = v_i\big(k^*(\hat{\theta}), \theta_i\big) + t_i,$$

where $v_i\big(k^*(\hat{\theta}), \theta_i\big)$ is agent i's valuation on the choice $k^*(\hat{\theta})$ when his type is θ_i. The term t_i denotes the payment rule and in this particular mechanism has the form:

$$t_i = \sum_{j \neq i} v_j\big(k^*(\hat{\theta}), \hat{\theta}_j\big) + h_i(\hat{\theta}_{-i}),$$

where $\hat{\theta}_{-i} = (\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \hat{\theta}_{i+1}, \ldots, \hat{\theta}_I)$. In general, h_i is an arbitrary function, but in the case of VCG mechanism is equal to the following:

$$h_i(\hat{\theta}_{-i}) = -\sum_{j \neq i} v_j\big(k^*_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j\big),$$

where $k^*_{-i}(\hat{\theta}_{-i})$ is the social choice which has resulted from a mechanism with all agents excluding agent i. This particular formula for the function h_i is called the pivotal or Clarke mechanism and reflects the contribution of agent i to the community. The utility function has the final form:

$$u_i(x, \theta_i) = v_i\big(k^*(\hat{\theta}), \theta_i\big) + \\ + \Big(\sum_{j \neq i} v_j\big(k^*(\hat{\theta}), \hat{\theta}_j\big) - \sum_{j \neq i} v_j\big(k^*_{-i}(\hat{\theta}_{-i}), \hat{\theta}_j\big)\Big) \quad (1)$$

If $k^*(\hat{\theta}) = k^*_{-i}(\hat{\theta}_{-i})$, which means that the reported type of agent i does not change the social choice, then $t_i = 0$ and hence, i is not charged. If $k^*(\hat{\theta}) \neq k^*_{-i}(\hat{\theta}_{-i})$, which means that agent i's type changes the social choice (agent i is pivotal), then $t_i < 0$. By allowing the payment rule t_i to be negative, it is possible to have a mechanism with the following properties:

1. *ex post* efficient: the social welfare is maximised
2. incentive compatible: for all agents, true revelation of their type, i.e. $\hat{\theta}_i = \theta_i$, $\forall i \in I$ is a dominant strategy.

IV. Our Scheme

Motivated by the work of Nix and Kantarcioglu in [1] we advance the potential of applying VCG mechanism in order to achieve privacy preserving collaborative classification. To comply with the game theoretic scenario, we assume a set of I agents, each one possessing a data set d_i, under the assumption that the pairwise intersection of these sets is empty. All agents share the same strategy set:

$$S_1 = \ldots = S_I = \{\texttt{true}, \texttt{perturbed}, \texttt{randomised}\},$$

where `true`, `perturbed` and `randomised` correspond to an agent providing true, perturbed and randomised data, accordingly. The set X of alternatives consists of the classification results. As explained in a previous section, the outcome of the mechanism, or in other words the collective choice, is that particular alternative which maximises the social welfare. In our scenario this is translated to achieving good classification

results. As classification is a supervised mining task, this alternative corresponds to the accuracy of the classification, which measures the performance of the classifier. Following the notation of [1] we denote the classification accuracy on a data set d by $acc(d)$. However, the lack of trust among the agents requires the introduction of privacy notions.

In our model, privacy is preserved by adding noise to the data values (perturbation). Although there are techniques to determine the distribution [12] and even to recover the true data from the noise [13], our experiment makes use of real data sets that do not have any particular trend, and thus those suggested methods for data recovery lead to poor results. For a clearer understanding of why this game theoretic approach succeeds, apart from the data perturbation, we also include complete randomization of the data, by replacing the true value with a random one. This random value is chosen from the interval formed by the minimum and maximum values of the attribute to be randomised. More formally, if x_i is the true value then the randomised value is $\tilde{x_i} = t_i$, where $t_i \in [\min \texttt{attribute\_value}, \max \texttt{attribute\_value}]$. Regarding the perturbation, the method we use depends on the type of the data. For numeric attributes we have $x_i' = x_i + r_i$, where r_i is chosen randomly from $[-a, a]$. If the attribute is of nominal type, then we use the AddNoise filter of the data mining toolset WEKA [14].

After the agents have decided on their preferences, their type is determined. The different types that we consider are: `per`, `rand`, `true`, where `per` describes an agent who provides perturbed data, `rand` corresponds to an agent who randomises the data and `true` represents an agent who is truthful. As all agents ideally prefer the extraction of information from true data, we regard their true type to be `true`, which corresponds to the accuracy of the classification on the union of the data sets $\bigcup_{i \epsilon I} d_i$ when all data is true. However, when an agent reports his type, the reported type $\hat{\theta}_i$ might not be the same as the true type θ_i.

An important feature of the mechanism design concept is a trusted third party who acts as the authority that imposes the rules. This is the role of the mediator, who will perform the mining task and distribute the payoffs to each agent. If the mediator knew the true (private) type of the agents, then he could decide the outcome of the mechanism and distribute payoffs to the agents according to their types. However, as the types are private the particular form of the Clarke mechanism serves as an incentive for the agents to reveal the true type, and thus lead to a fair payoff distribution by the mediator. Rewriting the payoff function (1) using the accuracy, agent i obtains the payoff:

$$u_i = acc(d) + \big(acc(\hat{d}) - acc(\hat{d}_{-i})\big), \tag{2}$$

with $d = \bigcup_{i \epsilon I} d_i$ being the union of the true data sets d_i, $\hat{d} = \bigcup_{i \epsilon I} \hat{d}_i$ is the union of the reported data sets $\hat{d}_i$ supplied by the agents and finally $\hat{d}_{-i} = \bigcup_{j \neq i} \hat{d}_j$, $i, j \in I$ corresponds to the data set formed from all data sets apart from the data of agent i. The expression

$$acc(\hat{d}) - acc(\hat{d}_{-i}) \tag{3}$$

calculates the loss or gain that agent i poses to the overall outcome, in other words his contribution. Using the result of (3) as a reference point, we can determine whether the agent wishes to mask his data in order to keep it private, or his aim is to harm the 'community' by providing falsified data. More specifically, if the modification of his data results in classification accuracy that leads to (3) having a negative value, then his behaviour is considered harmful. However, if from the modification we obtain accuracy that keeps (3) non negative, then we infer that agent i's intention is to preserve the privacy of his data without harming the overall outcome of the classification. Clearly, in an ideal situation agents would provide the true data and thus obtain high classification accuracy. However, as privacy is also required, the experimental results in the next section demonstrate that a controlled amount of perturbation results in both high accuracy levels and hiding of the data.

A. Measuring Privacy

Since the preservation of privacy is equally significant to the extraction of information, truth telling is not a necessarily desired strategy. On the other hand, complete falsification results in poor information gain. Perturbation of the data is a reasonable compromise, but what is the limit of the perturbation range before reaches complete randomisation, and subsequently diminishes the information gain? The answer lies in the term $acc(\hat{d}_{-i})$ of (2) which indicates the accuracy that can be achieved using data sets from all agents except agent i. As long as the expression (3) remains non negative, the perturbation of agent i's data causes insignificant reduction to the accuracy. If (3) becomes negative, then this is an indication that the perturbed data of agent i harms the overall accuracy and hence, agent i must obtain low payoff.

We suggest the following three different ways to measure privacy:

With respect to the distance from the true values:

$$\text{privacy} = \frac{|\text{perturbed value} - \text{true value}|}{|\text{randomised value} - \text{true value}|}$$

With respect to the range of the attribute values:

$$\text{privacy} = \frac{|\text{perturbed value} - \text{true value}|}{|\text{max value} - \text{min value}|}$$

With respect to the accuracy:

$$\text{privacy} = \frac{|\text{accuracy(perturbed data)} - \text{accuracy(true data)}|}{|\text{accuracy(randomised data)} - \text{accuracy(true data)}|}$$

Although in all cases the highest privacy is desirable, expression (3) poses a bound in the privacy that can be achieved without decreasing the agent's payoff.

V. Experimental Results

In support of the aforementioned model, this section presents our experimental results. The data set we used relates to the Civil War events in Africa, and was obtained from the Armed Conflict Location & Event Dataset [15]. For all data mining operations we used the toolset WEKA [14]. In particular, for the classification we applied the LibSVM to perform classification using the support vector machine method. Without loss of generality, we assumed that there are three agents with the following attributes:

Agent 1: {`year`, `source`}
Agent 2: {`actor1`, `actor2`}
Agent 3: {`latitude`, `longitude`}

All agents supply modified data, which can be either perturbed or randomised. We consider a small amount of perturbation for the attributes held by agents 1 and 2. In order to understand the sensitivity of the overall classification performance in terms of the amount of perturbation, we perform simulation study for a wide range of perturbation values while keeping the amount of perturbation on the data supplied by agents 1 and 2 fixed. We also study the classification performance when agent 3 completely randomises his attributes in order to understand the tradeoff between the privacy and the performance. The reason we choose agent 3 for greater modification of the data is due to his attributes `latitude`, `longitude` consisting of a wide range of real numbers. Moreover, the attributes of agent 3 form convex sets consisting of real numbers in the range of $[\min \texttt{latitude}, \max \texttt{latitude}]$ for the `latitude`, and $[\min \texttt{latitude}, \max \texttt{latitude}]$ for the `longitude`. Hence, the perturbed values also fall within these convex sets.

Let x_i denote the true value of an attribute and x'_i be the corresponding modified value. For perturbation of the numeric attributes, and particularly for the `year` attribute, we have that $x'_i = x_i + r_i$, with $r_i \in \{-1, 0, 1\}$. We perturbed both `latitude` and `longitude` as $x'_i = x_i + r_i$, with $r_i \in [-a, a]$. In our experiments we considered a range of perturbation as characterised by $a = 0.5, 1, 1.5, 2, 2.5, 3$. The nominal attributes (i.e. `source`, `actor1`, `actor2`) are perturbed using the AddNoise WEKA filter, with the noise parameter being 10%. For the randomisation of `latitude` and `longitude` $x'_i = t_i$, where t_i is drawn uniformly at random from $[\min \texttt{latitude}, \max \texttt{latitude}]$ and $[\min \texttt{longitude}, \max \texttt{longitude}]$, respectively. In order to prevent overfitting we applied the 0.632 Bootstrap method [16] with 200 bootstrap samples, each one having the same size as the training set.

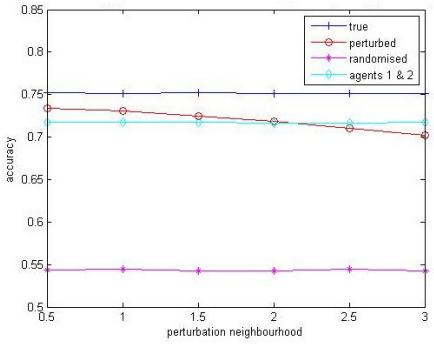
Figure 1(a) depicts the overall accuracy for four cases when the perturbation parameter a takes the aforementioned values (a) all agents provide true data (legend $-+-$), (b) all agents perturb the data (legend $-o-$), (c) agent 1 and 2 supply perturbed data and agent 3 provides randomised data (legend $-*-$) and (d) the classification is performed on the perturbed data of agents 1 and 2 only (legend $-\diamond-$). A closer look at these accuracies (Figure 1(b)) shows that between the ideal accuracy (which is achieved when all agents provide the true data) and a higher level of privacy (achieved when the data is perturbed), there is an interval where these two desired but contradictory properties are in balance. This interval lies between the accuracy of the classification on the true data and the output of (3), which is the accuracy that is achieved without the contribution of agent 3. Clearly, when agent 3 randomises, the resulting accuracy is significantly diminished.

Figures 2(a) and 2(b) depict the contribution (corresponding to the outcome of (3)) and payoff of agent 3. For a better understanding of these results, both figures present the charges and payoff, respectively, that result from the supply of perturbed data from agents 1 and 2, and true data from agent 3 (legend $-\diamond-$). As this situation offers the maximum payoff to agent 3, when he introduces perturbation in his data the charges increase and his payoff decreases. Perturbation of up to $a = 2$ (i.e., 2^0 of perturbation) results in high payoff, and at the same time the data is concealed, as 2^0 latitude is equal to 222km. Furthermore, both Figure 2(a) and 2(b) show that randomisation is not a beneficial approach due to very low payoff.

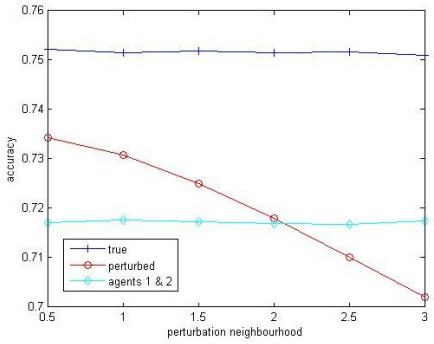
Regarding the privacy, Figure 3 shows the three different ways of measuring it. Clearly, the maximum privacy is achieved when the data is completely randomised. However, as randomisation results in poor classification accuracy, the actual maximum privacy that can be attained is represented by the line denoted by 'max-privacy-'. In all three subfigures, this line marks the critical point which separates the privacy with positive classification results from the privacy with undesirable classification results. Finally, Figure 4 presents a 3D overview of the relation among the perturbation, the accuracy and the privacy, for the three different privacy measures. The square on the figures denotes the critical point (as can also be seen in the intersection of those two curves in Figure 1(b)) where we have the maximum privacy while the accuracy of the classification is high and the perturbation of agent 3 is not harmful.

VI. Conclusions

This work examined the problem of collaborative data mining using tools from game theory, while being able to offer data privacy to individual agents. In particular, motivated by [1] we used the Vickrey-Clarke-Groves mechanism in order to offer incentives that will prompt the agents to follow the rules. The behaviours that we considered are `true`, `per`, `rand`, corresponding to agent providing true, perturbed and randomised data. Our experiment showed that indeed the use of the VCG mechanism leads to high accuracy of the data mining task, while preserving the privacy of the data by allowing the agents to supply perturbed data. The key point of the VCG mechanism is that the gain of each agent includes the agent's contribution. Hence, the agent can perturb the data, as long as his contribution does not harm the overall result.

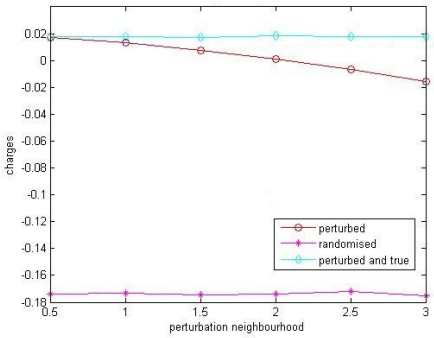


(a) The overall classification accuracy

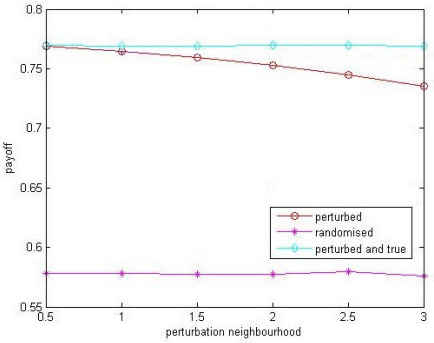


(b) Magnified part of the overall accuracy, showing the interval where accuracy is high and the contribution of agent 3 is not harmful.

Fig. 1.

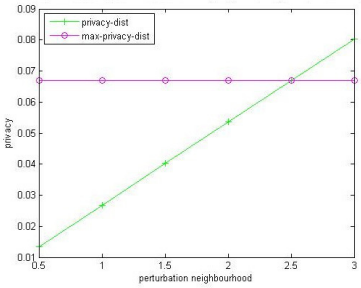


(a) The contribution of agent 3.

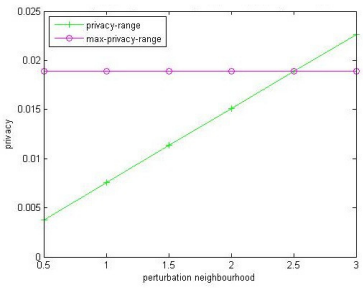


(b) The payoff of agent 3.

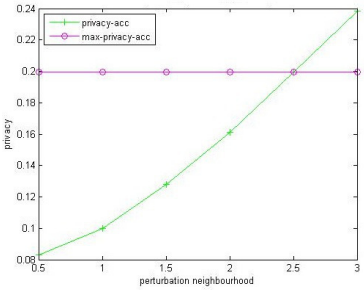
Fig. 2.



(a) With respect to the distance from the true values.



(b) With respect of the attribute values.



(c) With respect to the accuracy.

Fig. 3. Privacy

REFERENCES

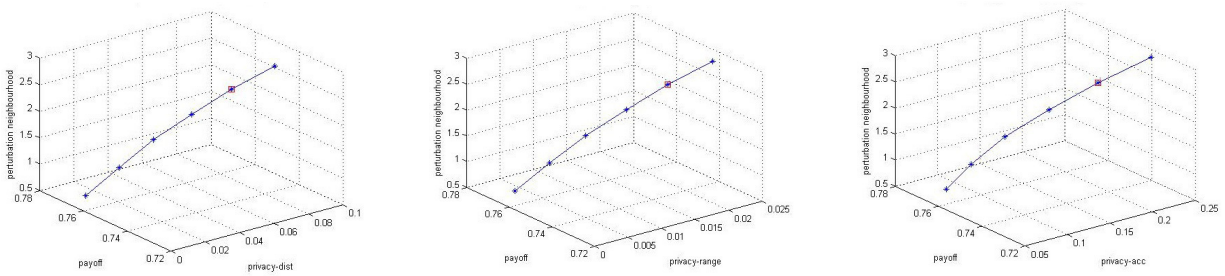
[1] R. Nix and M. Kantarcioglu, "Incentive compatible privacy-preserving distributed classification," *IEEE Trans. Dependable Sec. Comput.*, vol. 9, no. 4, pp. 451–462, 2012.

[2] J. Halpern and V. Teague, "Rational secret sharing and multiparty computation: extended abstract," in *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, 2004*. ACM, 2004, pp. 623–632.

[3] I. Abraham, D. Dolev, R. Gonen, and J. Halpern, "Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation," in *Proceedings of the 25th annual ACM symposium on Principles of distributed computing*, ser. PODC '06. New York, NY, USA: ACM, 2006, pp. 53–62.

[4] H. Kargupta, K. Das, and K. Liu, "Multi-party, privacy-preserving distributed data mining using a game theoretic framework," in *Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, ser. PKDD 2007. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 523–531.

[5] S. Izmalkov, S. Micali, and M. Lepinski, "Rational secure computation and ideal mechanism design," in *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 585–595.



(a) (b) (c)

Fig. 4. A 3D overview of the relation among the perturbation, the accuracy and the privacy, for the three different privacy measures.

[6] R. Agrawal and E. Terzi, "On honesty in sovereign information sharing," in *Proceedings of the 10th international conference on Advances in Database Technology*, ser. EDBT'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 240–256.

[7] R. Layfield, M. Kantarcioglu, and B. Thuraisingham, "Incentive and trust issues in assured information sharing," in *CollaborateCom*, 2008, pp. 113–125.

[8] O. Dekel, F. Fischer, and A. Procaccia, "Incentive compatible regression learning," *Journal of Computer and System Sciences*, vol. 76, no. 8, pp. 759–777, 2010.

[9] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*. New York: Oxford University Press, 1995.

[10] D. Parkes, "Iterative combinatorial auctions: Achieving economic and computational efficiency," Ph.D. dissertation, Univesity of Pennsylvania, 2001.

[11] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, 1st ed., ser. MIT Press Books. The MIT Press, 1994, vol. 1.

[12] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '00, vol. 29, no. 2. New York,USA: ACM, 2000, pp. 439–450.

[13] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques," in *Proceedings of the Third IEEE International Conference on Data Mining*, ser. ICDM '03, Washington, DC, USA, 2003, pp. 99–.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[15] Raleigh, Clionadh, A. Linke, H. Hegre, and J. Karlsen, "Introducing acled-armed conflict location and event data," *Journal of Peace Research*, vol. 47, no. 5, pp. 1–10, 2010.

[16] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2005.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 129–136

dotRL: A platform for rapid Reinforcement Learning methods development and validation

Bartosz Papis, Paweł Wawrzyński
Institute of Control and Computation Engineering, Warsaw University of Technology

***Abstract*—This paper introduces dotRL, a platform that enables fast implementation and testing of Reinforcement Learning algorithms against diverse environments. dotRL has been written under .NET framework and its main characteristics include: (i) adding a new learning algorithm or environment to the platform only requires implementing a simple interface, from then on it is ready to be coupled with other environments and algorithms, (ii) a set of tools is included that aid running and reporting experiments, (iii) a set of benchmark environments is included, with as demanding as Octopus-Arm and Half-Cheetah, (iv) the platform is available for instantaneous download, compilation, and execution, without libraries from different sources.**

***Index Terms*—Reinforcement learning, evaluation platform, software engineering**

I. Introduction

IN THE area of Reinforcement Learning (RL) [1] algorithms are developed that learn reactive policies for sequential decision making and control. Research in RL is based on the paradigm of *micro–worlds*: ideas are tested and demonstrated with the use of decision-making and control problems that can be defined analytically and reimplemented by others. This has forced researchers to spend a lot of time developing their experimental platforms. In order to help others and enable fair comparison of the ideas, many researchers have published their platforms: RL-Glue [2], PyBrain [3], CLSquare [4], RLT [5], PIQLE [6], lipbqrl [7], MDP Toolkit [8], MMLF [9], or QCON [10]. The general design principles for RL platforms were analysed in [11].

The purpose of this paper is to introduce another platform, dotRL, for development of RL algorithms. Although the platforms reduce the space for yet another project, it also demonstrates that a researcher developing a new idea in RL or a student getting familiar with this field still prefers writing their own platform from scratch instead of using an existing one. That is why the main principle that we adopted when designing our platform was as follows: the user should spend as little time as possible installing, getting familiar with the platform, and writing code, before they are ready to run their own agent or environment.

A. Related work

Perhaps the oldest and best-known RL platform is RL-Glue [2]. It dates back to 1996 through a project by Rich Sutton and Juan Carlos Santamaria called RL-Interface. RL-Glue has been a protocol specified by annual RL competition workshops held at ICML and NIPS. RL-Glue is basically a text communication protocol over sockets, between agents and environments. Reinforcement learning toolbox (RLT) [5] is a flexible platform for development learning algorithm in various scenarios: MDP, POMDP, and imitation learning. The price of this flexibility is the complexity of this platform and difficulty of its use. Libpgrl [7] focuses on planning and reinforcement learning in a distributed environment. Maja machine learning framework (MMLF) [9] supports not only RL but also model-based learning and direct policy search. It enables automated experimentation with the use of XML configuration files. PyBrain [3] is a general machine learning library, that also includes RL, but focuses on neural networks. Object-oriented platforms written in Java include PIQLE [6], RLPark [12], and Teachingbox [13]. Another platform, YORLL [14], is written in C++.

B. Requirements and basic assumptions

The dotRL platform is designed to minimize the time spent by its user on technical and infrastructural details. The user should focus almost all of their effort on dealing with purely scientific issues. In order to meet this requirement, the design of dotRL is based on the following assumptions and characteristics:

1) Altogether, dotRL is a *solution* written under .NET 4.5 framework, Windows operating system, and Visual Studio 2010. As a result, further development of dotRL may be based on all the tools provided with Visual Studio and .NET technology.
2) Having been downloaded and opened with Visual Studio, it is ready to be compiled and run.
3) In order to add a new agent or a new environment to the platform, one only needs to implement a class with an appropriate interface. After compilation, the platform alone is able to couple this new entity to other environments or agents.
4) Each agent and environment is designed for one particular *problem type*. The problem type defines the types of state and action spaces. They may be continuous (i.e., contain vectors of reals), discrete (contain vectors of integers), and possibly others.
5) A set of tools is provided with dotRL that enables launching many learning runs with the same setting and getting logs almost directly insertable to a scientific paper. Tools for implementing agents, such as neural networks, are also included.
6) A set of exemplary agents and environments are provided with the platform. Those include as complex en-

978-83-60810-53-8/$25.00

vironments as Octopus-Arm [15], [16] and Half-Cheetah [17], [18].

7) The platform is fully compatible with RL-Glue [2].

To our knowledge, the platform presented in this paper is the first full-featured platform written under .NET, and the first one in which adding a new agent or environment only requires implementing a single class. Especially this last feature is helpful in rapid development and validation of new algorithms.

The aforementioned notion of problem types is based on the following observation: An agent is usually applicable, without modification only to environments with compatible state and action space types. No one really implements a learning algorithm that, in the same form, is applicable to several problem types. It is possible to do so, but almost always means that the agent will do something completely different for different versions of environment it deals with at the moment.

Additional contribution of this work is RL-Glue codec for .NET platform.

dotRL is an open source software under BSD license and hosted on *sourceforge.net* [19]. We welcome anyone to contribute to the project.

C. Organization of the paper

The remaining part of the paper is organized as follows. Sec. II presents an overview of the user interface, sec. III defines basic modules and components of the dotRL platform. Another subsection presents the interaction protocol between an agent and an environment that the platform supports. Sec. IV explains how to use a new component (agent or environment) with the platform. Sec. V elaborates on integration of dotRL with the RL-Glue protocol. Sec. VI concludes the paper and indicates directions of future development of the platform.

II. User interface

Typical usage scenario of the dotRL solution, when the user wants to test an existing agent on an existing environment consists of the following steps:

1) Click the “Experiment” menu item from the “New” menu,
2) Choose an environment from the list of available environments,
3) Choose an agent from the list of available agents compatible with the chosen environment,
4) Configure parameters of the chosen environment and agent
5) Configure reporting parameters,
6) Configure experiment parameters (i.e. number of episodes, maximum number of steps in one episode),
7) Click “OK” when finished configuring the experiment,
8) Click “Background learning” or “Real time learning”,
9) Click “Present policy” and/or view the created report file.

The user can modify parameters of the ongoing experiment. Details on extending the platform’s set of components (agents or environments) are provided in Section IV. An example view of the application during configuration of an experiment is presented in Figure 1. Running experiment is presented in Figure 3, and a screen presenting functionality allowing more than one simultaneous experiments to run is shown in Figure 4.

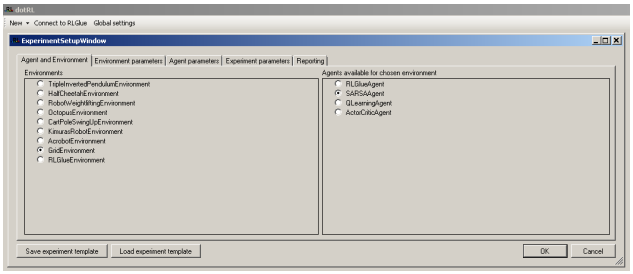


Fig. 1. Experiment configuration.

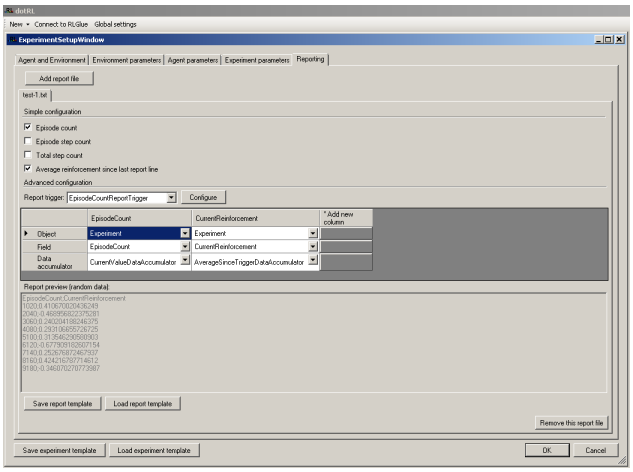


Fig. 2. Reporting configuration.

To configure the reporting parameters “Add report file” button in “Reporting” tab needs to be clicked. Then, the user can either choose to use simple configuration and choose from the standard set of report columns, or to configure their own report:

1) For each report file tab:
   a) Choose one of *report triggers*
   b) Click “Add new column” for each desired column in the output file
   c) Choose one of available data sources and a way to accumulate their values

ReportTrigger and *DataSource* objects are explained in detail in Section III-C. An example view of the application during reporting configuration is presented in Figure 2.

For interacting with RL-Glue one of these two actions must be taken:

- Choose the *RLGlueAgent* or *RLGlueEnvironment* in the component choice window after choosing to create new experiment
- Start an RL-Glue experiment to connect to RL-Glue core.

Integration with RL-Glue components is explained in detail in Section V.

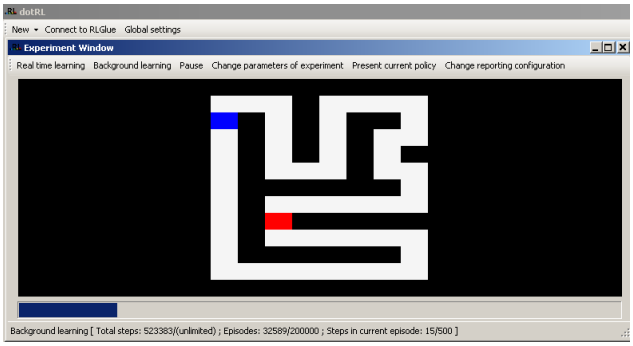


Fig. 3. An example experiment.

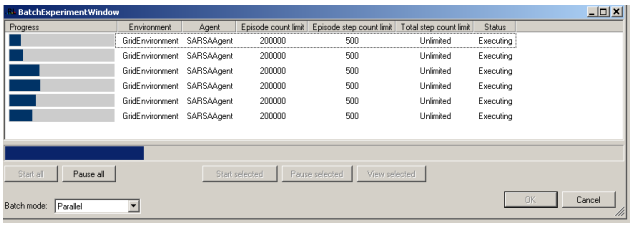


Fig. 4. An example batch experiment.

III. dotRL components

This section presents the *domain model* [20] of the dotRL solution. Section III-A presents the set of core entities which reflect key notions of an RL experiment. Section III-B presents how these components interact with each other during an experiment.

A. Classes

Learning algorithms, called *Agents* in RL are represented as subclasses of the *Agent<TStateSpaceType, TActionSpaceType>* base class. Problems to solve by these algorithms, called *Environments* in RL are represented as subclasses of the *Environment<TStateSpaceType, TActionSpaceType>* base class. Environments can have continuous or discrete state transition function and they accept real or integer vectors as actions. This divides them into four groups, three of which are commonly addressed, and which we call *problem types*: continuous state & continuous action, continuous state & discrete action, discrete state & discrete action. Each agent and environment is dedicated to one problem type and this is made explicit in dotRL in the form of generic parameters of *Agent* and *Environment* base classes. Interaction between an agent and an environment is called *Experiment*. This whole design is modelled with classes presented in Figure 5.

Agent represents the class hierarchy of all agents implemented in dotRL, with *Agent<TStateSpaceType, TActionSpaceType>* (in Figure 5 generic arguments are omitted for clarity) being their base class. Agent's responsibility is to decide which *Action* to take in given *Environment*'s *State*, and to improve its policy with received *Sample*s. Details on how to implement an agent are provided in Section IV.

Similarly, *Environment* represents the hierarchy of classes which represent RL problems to be solved by the *Agent*s. The class *Environment<TStateSpaceType, TActionSpaceType>* (in Figure 5 generic arguments are omitted for clarity) is the base class for any environment implemented in dotRL. Environment's responsibility is to simulate a designed behaviour, reacting to given *Action*s by changing it's *State* and providing a *Reinforcement*. Unlike some other solutions (like PyBrain [3]) we do not divide responsibility of modelling a behavior and assigning reinforcement between two separate objects. Theoretically, it would lead to a more accurate domain model and it is a valuable idea, but it makes development more time-consuming and this opposes our requirements. Different rewarding policies can be easily implemented using environment's parameters.

The *Experiment* models a key notion in RL research — an experiment, i.e. a continuous interaction between an *Agent* and an *Environment*. Experiment's resposibilities are: controlling the course of an experiment (i.e. informing about begining and ending of an episode, evaluating finish conditions) and passing information between an agent and an environment (*State*s, *Action*s, *Sample*s and *Reinforcement*s), and passing information to classes resposible for reporting functionality.

State<TStateSpaceType>, *Action<TActionSpaceType>* and *Reinforcement* (again, generic parameters omitted for clarity in Figure 5) are simple wrapper classes for vectors and numbers to make RL domain notions explicit in the code — they are not essential, but they make the design clear and explicit, and improve implementation's readability.

EnvironmentDescription<TStateSpaceType, TActionSpaceType> (again, generic parameters omitted for clarity in Figure 5) is a class containing information about the structure of an environment. The details about its contents are provided in Section IV.

Presentation class provides a root for hierarchy of classes that are used to visualize the state of the environment. Its responsibility is to draw a visualization of a given state on a given canvas object (.NET's *System.Drawing.Graphics*). It is used only when user chooses "Policy presentation" mode in the user interface.

Sample represents a smallest piece of information in a RL experiment. Sample consists of:

- *PreviousState*: a state in which the *Environment* was.
- *Action*: an action taken by the *Agent* for state *PreviousState*.
- *CurrentState*: a resulting state after taking action *Action* in state *PreviousState*.
- *Reinforcement*: a reinforcement received after taking action *Action* in state *PreviousState*.

The use of *sample*s allows the implementation of an agent to be stateless — no information needs to be stored between calls to various agent's methods (such as *EpisodeStarted*, *GetAction*, etc.). More details are available in Section IV.

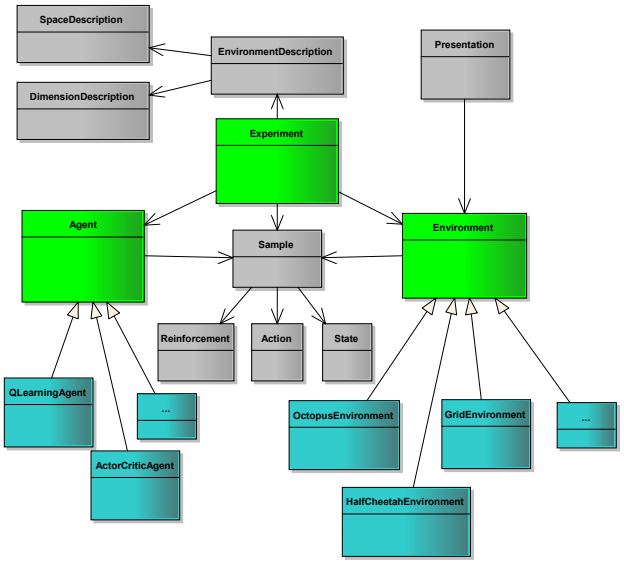


Fig. 5. dotRL main components. Green classes represent core components essential to implement the concept of RL experiment. Gray classes are useful utility classes which represent minor concepts. Blue classes represent a place for user's activity: they are concrete implementations of agent's and environment's behaviors. Ordinary arrow represents association, filled arrow represents inheritance.

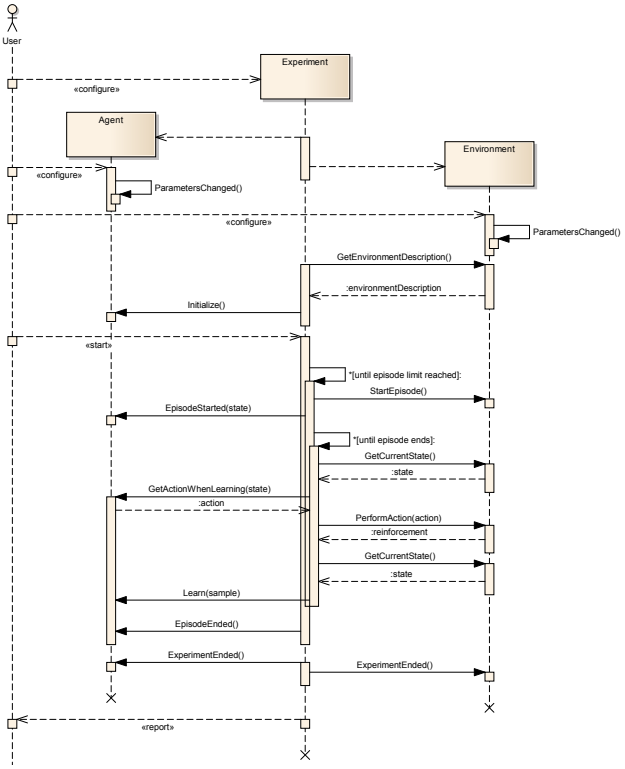


Fig. 6. dotRL interaction scenario. After user initiates a new experiment the platform's core components exchange data in a way typical for an RL experiment.

B. Operation sequence

We propose to adopt a simple interaction scenario based on explicit interfaces. Many existing RL platforms use typical setting in which subsequent method calls (*episode start*, *step*, *episode end*) implicitly rely on each other, forcing agent's implementation to be a state machine. This is not always the most convenient way, and such interface does not follow readable code guidelines [20]. The proposed sequence of method calls between components during an experiment is presented in Figure 6.

After the user initiates a new experiment instances of chosen classes are being automatically created: a subclass of the *Agent<TStateSpaceType, TActionSpaceType>* base class and a subclass of the *Environment<TStateSpaceType, TActionSpaceType>* base class (generic arguments are omitted for clarity in Figure 6). First, the user configures the parameters of the experiment (i.e. number of episodes, number of steps in each episode), agent and environment. Then, after experiment passes the information about the environment to the agent, a loop common to all RL experiments is being started. Each episode consists of a sequence of repeatedly executed steps:

1) The current state of the environment is retrieved by calling the *GetCurrentState* method.
2) If the current state is terminal or the current episode should end because of its duration limit, agent's *EpisodeEnded* method is called, and a new episode is started by calling *StartEpisode* environment's method and *EpisodeStarted* agent's method.
3) Agent's action for current state is retrieved with call to the *GetActionWhenLearning* method.
4) The *Environment* is informed what action it should execute via call to the *PerformAction* method.
5) The reinforcement and the new current state are retrieved from the *Environment* as return values from the *PeformAction* and *GetCurrentState* methods.
6) The *Agent* is informed about the consequences of executed action via call to the *Learn* method.

If the user wishes only to see how the agent behaves without changing it's policy they can choose "Policy presentation" mode. In this mode, another copy of the environment is used and agent's *GetActionWhenNotLearning* method is used (opposed to *GetActionWhenLearning*) so there is no interference with the learning process (provided that the implementation of *GetActionWhenNotLearning* is correct and truly does not influence the learning process).

C. Reporting

A valuable functionality of dotRL is provided by the reporting mechanisms. When configuring an experiment the user can setup multiple output log files. This is done through three useful notions: *ReportTrigger*, *DataSource* and *DataAccumulator*. A report trigger is a class that decides when to write a line to the output file. Lots of report triggers have already been implemented, such as: *TotalStepCountReportTrigger* which causes emiting a log file line at configured intervals of steps in an experiment, *EpisodeCountReportTrigger* which causes emiting a log file line at configured intervals of episodes in the whole expermient.

DataSource consists of an object and it's field, which's value

will be reported to the output file. There are three main objects which provide data: Experiment, Agent and Environment. The experiment provides typical experiment information, like the number of steps executed so far, or the reinforcement received by the agent. Data exposed by the agent and the environment depends on the creator of these components. Any useful piece of data can be accessed by the reporting functionality as easy, as marking any component's field with *ReportedValue* attribute. An example is provided in *QLearningAgent* which exposes it's *td* field, containing recently computed temporal-difference value.

DataAccumulator allow some simple manipulations on the data read from data sources. The most common are *no-op* data accumulator (*CurrentValueDataAccumulator*) which just outputs the returned value and *AverageSinceTriggerDataAccumulator* which accumulates the data between each report file line and calculates average.

D. Implemented components

Currently, the following components are implemented in dotRL:

- Environments:
  - *Cart-Pole Swing Up* [21]
  - *Double Inverted Pendulum on a Cart* [22]
  - *Acrobot* [23]
  - *Robot weightlifting* [24]
  - *Kimura's Robot* [25]
  - *Half Cheetah* [17]
  - *Octopus Arm* [16]
  - *Coffee task* [26]
  - *Grid*
- Agents:
  - *Actor-Critic* [27]
  - *Actor-Critic with Experience Replay* [17]
  - *Q-Learning* [28]
  - *SARSA* [29]

IV. Adding new components

We focus our design to make adding new agents and environments as simple as possible. This allows a researcher to spend most of his time on substantial work instead of dealing with technical details. Developing a new agent or a new environment is most straightforward: one needs just to subclass the *Agent*<*TStateSpaceType, TActionSpaceType*> class or the *Environment*<*TStateSpaceType, TActionSpaceType*> class, respectively. The *TStateSpaceType* and *TActionSpaceType* generic arguments need to be set to types corresponding to desired problem type (for example: setting $TStateSpaceType = double, TActionSpaceType = int$ allows creation of a continuous state & discrete action agent/environment).

Each component, once implemented, will appear automatically in the user interface. If additionally a subclass of the *Presenter* class is supplied, the environment's state will be visualized in the experiment's window. Otherwise the default presenter will be used, which just prints raw state and reinforcement information. The implementation of *Experiment*<*TStateSpaceType, TActionSpaceType*> (green component in Figure 5) is provided by the dotRL platform, and is fully configurable through the user interface.

Another convenience is automatic handling of component's parameters. Every *Agent* or *Environment* can have any of its fields or properties (doesn't matter whether private, protected, public, static or instance related) marked with one of *Parameter* or *NumericParameter* attributes. Such fields will appear in a configuration dialog window before starting each experiment, allowing the user to tune the component's behavior. Also, if any component uses another component (for example one wants to implement an environment similar to an existing one, and reuses the latter as a part of the new one) its parameters will be also handled automatically.

A. Adding a new environment

Subclassing the *Environment*<*TStateSpaceType, TActionSpaceType*> class requires implementing the following methods (for clarity, generic arguments have been omitted):

- EnvironmentDescription *GetEnvironmentDescription*(): called to retrieve information about the environment
- void *StartEpisode*(): called when a new episode begins
- Reinforcement *PerformAction*(Action action): called to execute action and retrieve reinforcement

The first method is called to transfer information about the structure of the problem to the agent. Usually agents require information about the problem's state, action and reinforcement spaces. Such information is stored in *EnvironmentDescription*<*TStateSpaceType, TActionSpaceType*> class, which has two instances of *SpaceDescription*<*TSpaceType*> classes (one for state space and one for action space) and one instance of *DimensionDescription*<*TSpaceType*> class for describing the reinforcement space. *SpaceDescription*<*TSpaceType*> consists of *DimensionDescription*<*TSpaceType*> instance for each described dimension. *DimensionDescription*<*TSpaceType*> contains: minimum value, maximum value, average value and standard deviation.

Not all of these fields are always used. Typically, state space information contains:

- Minimum value for each state variable.
- Maximum value for each state variable.
- Average value for each state variable.
- StandardDeviation of each state variable.

Action space information:

- Minimum value for each action dimension.
- Maximum value for each action dimension.

Information about the reinforcement:

- Minimum reinforcement value.
- Maximum reinforcement value.

Despite the typical setting, all values are optional but the environment should provide as much information as possible, to allow cooperation with agents that need it.

Typical behavior of the *StartEpisode* method is to initialize environment's state (to some predefined state, probably dependent on parameters or to a random state).

The last method, *PerformAction* is typical to RL environment implementations: it usually performs a simulation step, evaluating the consequences of the given *action* (calculating environment's next state) and returns a reinforcement associated with this *action* in its current state.

Technically, these methods should be implemented in the paradigm of a *stateful* protocol — environment should keep track of its current state. To facilitate this and for efficiency, the *Environment<TStateSpaceType, TActionSpaceType>* base classes exposes a protected mutable *CurrentState* property. As long as it is used by *StartEpisode* and *PerformAction* methods one needs not to bother about implementing the *GetCurrentState* method.

Additionally these methods can optionally be overriden:

- State *GetCurrentState*(): called to retrieve environment's current state
- void *ParametersChanged*(): called after user changes environment's parameters
- void *ExperimentEnded*(): called after the user closes the experiment window

The default implementation of the first method returns the *CurrentState* property as an immutable object. The default implementations of the two remaining methods do nothing.

Environment class must contain a parameterless constructor.

B. Adding a new agent

Subclassing the *Agent<TStateSpaceType, TActionSpaceType>* class requires implementing the following methods (for clarity, generic arguments have been omitted):

- void *ExperimentStarted*(EnvironmentDescription environmentDescription): called to pass the information about the environment to the agent
- Action *GetActionWhenNotLearning*: called to retrieve agent's decision about an action to take in the given *state*, when presenting current policy
- Action *GetActionWhenLearning*(State state): called to retrieve agent's decision about an action to take in the given *state*, during learning
- void *Learn*(Sample sample): called to inform the agent about a state, action that took place and the resulting next state and reinforcement

The first method is called before the start of the experiment, so the agent could prepare its internal structures accordingly to the structure of the environment(e.g., dimensions of the state, and action spaces).

The *GetActionWhenNotLearning* should return the action according to agent's current policy for the given *state*, not affected by agent's exploration policy, and in way not to interfere with agent's internal state related to learning.

The third method, *GetActionWhenLearning* should return the action according to current agent's policy for the given *state* which can be distorted for exploration. Also in this method agent can calculate or remember some additional quantities which will be needed in the *Learn* method. It is guaranteed that a call to *Learn* method will always follow a previous call to *GetActionWhenLearning*.

The *Learn* method, typical to RL agent implementations, is used to transfer experience to the agent. Agent can, for example accumulate this experience, or improve its policy at once.

Technically these methods should be implemented in the paradigm of a *stateless* protocol. Subsequent calls to *GetActionWhenLearning* or *Learn* should not be explicitly dependent. Of course, there should be an implicit dependency between these calls and calls to the *Learn* method through the agent's policy and some internal variables used for learning. Also, the *action* returned by the *GetActionWhenLearning* method will be present in a *sample* given to a subsequent *Learn* call.

The *GetActionWhenNotLearning* method should act in a completely transparent way — no assumptions should be made about the moment of its execution, as user can switch to "Policy presentation" mode at any time. The environment will remain unaffected, as in "Policy presentation" mode its copy is being used.

Similarly to environments' base class, the base class for agents exposes a mutable protected *Action* property to be modified in place for efficiency, and returned from *GetAction* as an immutable version.

Additionally these methods can optionally be overriden:

- void *EpisodeStarted*(State state): called when episode starts
- void *EpisodeEnded*(): called when episode ends
- void *ExperimentEnded*(): called after the user closes the experiment window

The default implementations of these methods do nothing.

Agent class must contain a parameterless constructor.

C. Adding a new presenter

Subclassing the *Presenter* base class is optional, however it allows a researcher to evaluate environment's behavior visually. Usually environments represent some imaginable object, or they are related to a real-world object. Having them drawn and being able to observe their dynamics makes it easier to verify their implementation, and to analyze agent's behavior. To implement a presenter one needs to:

- subclass the *Presenter* base class and implement the *Draw* method,
- provide a constructor taking one argument of type of the environment to visualize.

The *Draw* method should draw the visualization of environment's current state to the canvas held by the *Graphics* object, exposed by the *Presenter* base class. *Draw* implementation should use drawing functions also provided by the *Presenter*

base class, as they scale the drawing appropriately to window's dimensions and aspect ratio. Presenter implementations should hold the reference to the visualized environment for themselves. Implementation of a presenter can be designed to visualize more than one environment — it just needs to provide one constructor for each visualized environment. This is useful when one environment is derived from another one.

V. Integration

A. Integration with RL-Glue

In RL research field there are many agents and environments available, however they are implemented in different platforms and languages. This problem is taken care of by the RL-Glue protocol [2]. dotRL supports it, to give its users access to the vast set of agents and environments already implemented and compatible with RL-Glue.

Generally, interaction between two integrated platforms (or components from different solutions) relies on one of them managing the course of an experiment and the other acting passively as one of experiment components: an agent or an environment. This gives two options of integrating dotRL with RL-Glue components:

1) *dotRL* manages the course of an experiment employing a RL-Glue agent or environment.
2) *dotRL* acts as RL-Glue component: an agent or an environment, while RL-Glue manages the course of the experiment.

dotRL allows both scenarios: the user can instantiate a component (agent or environment) alone, and configure it to connect to a RL-Glue server application, or user can open an ordinary experiment window chosing a special agent (*RLGlueAgent*) or environment (*RLGlueEnvironment*) type which act as proxies and encapsulate RL-Glue network communication details. Example of the first scenario is presented in Figures 7 and 8. Some other platforms, like Teachingbox [13] also provide integration with RL-Glue, but dotRL is the only platform known to the authors which allows both integration scenarios.

B. Integration with other applications

Integrating dotRL with other platforms, or single-component applications is easy thanks to the mechanisms provided by the .NET framework. For example:

- C/C++ code, compiled to a *DLL* library is easily accessible from .NET through the *Platform Invoke Services* (*P/Invoke*) or *It Just Works* (*IJW*).
- Python code can be accessed with solutions like *IronPython* [30] which provide Python virtual machine implementation in .NET.
- Because *Matlab* gives access to its API through DLL interface, it also possible to run *Matlab* scripts using *P/Invoke* mechanism. *Matlab*'s COM interface can also be used.
- Interaction with *XML* based communication (e.g. configuration of the Octopus-Arm environment [15], [16]) is easy to develop thanks to provided tools ranging from simple *XML* stream processors (*XmlReader* and *XmlWriter*) to powerful *LINQ to XML* which allows comfortable operating on *XML* documents in an elegant and concise way.
- Network communication is also supported with easy to use highlevel classes and libraries like *WCF* (*Windows Communication Foundation*).

In any of these cases, user is required only to implement a wrapper class managing the interoperability using one of described mechanisms as a subclass of the *Agent* or *Environment* base class.

C. Comparison with RL-Glue

Although RL-Glue and dotRL are different types of solutions, some of their merits can be compared:

- RL-Glue being a network protocol is fully platform independent, whereas dotRL can work only on platforms with existing .NET framework implementation. However, the number of platforms which support .NET is becoming bigger. Thanks to the *MONO* project [31], .NET is supported not only on Windows, but also most Unix-like systems, MacOS X, and even Android (full list of supported systems can be found in [32]).
- RL-Glue is however tied to the infrastructure: *BigEndian* convention is obligatory[1], and there are fixed sizes of basic data types. dotRL is completely infrastructure ignorant, as long as .NET Framework is supported.
- RL-Glue, being a handcrafted protocol is not feasible for extensibility (e.g. hard-coded order of elements in *TaskSpec*). dotRL is easily extensible by design.
- In both solutions adding new components requires at minimum writing only one class required to implement one simple interface.
- RL-Glue does not offer a standard way to turn off policy improvement. dotRL handles this case via "policy presentation" mode.
- In RL-Glue reporting and visualization must be managed by the user on their own (however, see *RL-Viz* project mentioned below). dotRL offers easily extensible reporting and presentation frameworks with convenient GUI.

RL-Glue is more popular, and in fact seems a better solution if multiplatform or distributed environment is obligatory. However, dotRL is easier to maintain (no hardcoded assumptions), provides more useful tools for evaluation of new agents (built in extensible reporting framework), visualization of environments (built in extensible visualization framework), and allows easy integration with other solutions (details were provided in sec. V-B). In the context of reporting and visualization it is fair to mention the *RL-Viz* project, however it remains unreleased since 2007 [33].

[1]BigEndian is obligatory in the context of network communication with the RL-Glue Core application. Refer to the RL-Glue Core implementation: "rlBufferWrite" and "rlBufferRead" functions in the "RL_network.c" file.

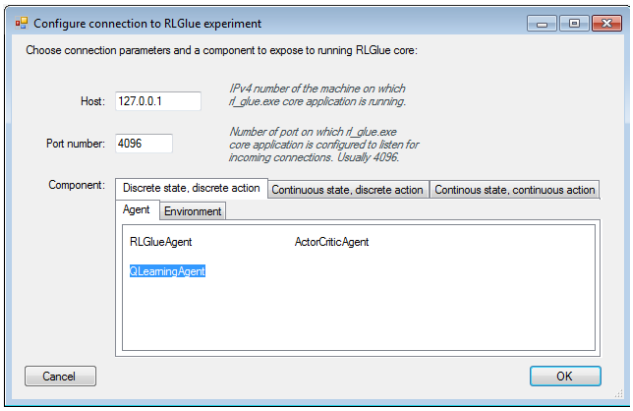


Fig. 7. RL-Glue connection configuration.

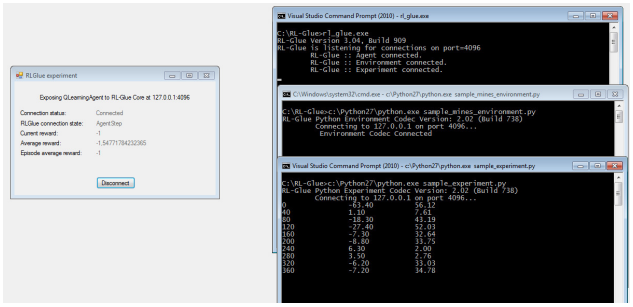


Fig. 8. RL-Glue experiment in progress, using dotRL's QLearningAgent for RL-Glue sample environment and experiment written in *Python*

VI. Conclusions and Future Work

In this paper dotRL — a platform for fast development and validation of reinforcement learning algorithms was introduced. The platform had been designed to minimize the time spent by its user on technical and infrastructural details, as they should focus on purely scientific issues. Seemingly, the platform meets this requirement.

Directions of further development of dotRL include its integration with other platforms. They also encompass enriching the set of agents and environments available within the platform.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[2] B. Tanner and A. White, "Rl-glue: language-independent software for reinforcement-learning experiments," *Journal of Machine Learning Research*, vol. 10, pp. 2133–2136, 2009.

[3] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber, "Pybrain," *Journal of Machine Learning Research*, vol. 11, pp. 743–746, 2010.

[4] R. Hafner and M.Riedmiller, "Case study: control of a real world system in clsquare," in *Proceedings of the NIPS Workshop on Reinforcement Learning Comparisons, Whistler, British Columbia, Canada*, 2005.

[5] G. Neumann, "Reinforcement learning for optimal control tasks," Master's thesis, Technischen Universitat, Graz, 2005.

[6] F. D. Comité and S. Delepoulle, "Piqle: a platform for implementation of q-learning experiments," in *NIPS workshop: reinforcement learning benchmarks and bake-offs II*, 2005.

[7] D. Aberdeen, O. Buffet, F. P. Selmi-Dei, X. Zhang, and T. Lopes, "libpgrl," 2006, http://code.google.com/p/libpgrl/.

[8] I. Chadés, M. J. Cros, F. Garcia, and R. Sabbadin, "Markov decision processes (mdp) toolbox," 2009, http://www.inra.fr/mia/T/MDPtoolbox/.

[9] M. Edgington, "Maja machine learning framework," 2009, http://mmlf.sourceforge.net/.

[10] D. Kapusta, "Connectionist q-learning java framework," 2005, http://elsy.gdan.pl/.

[11] T. Kovacs and R. Egginton, "On the analysis and design of software for reinforcement learning, with a survey of existing systems," *Machine Learning*, vol. 84, pp. 7–49, 2011.

[12] "Rlpark." [Online]. Available: http://rlpark.github.com/

[13] "Teachingbox." [Online]. Available: http://amser.hs-weingarten.de/en/teachingbox.php

[14] P. Scopes, V. Agarwal, S. Devlin, K. Efthymiadis, K. Malialis, D. T. Kentse, and D. Kudenko, "York reinforcement learning library (yorll)," reinforcement Learning Group, Department of Computer Science.

[15] Octopus-sources, 2006, http://www.cs.mcgill.ca/ dprecup/workshops/ICML06/Octopus/octopus-code-distribution.zip.

[16] B. G. Woolley and K. O. Stanley, "Evolving a single scalable controller for an octopus arm with a variable number of segments," in *Proceedings of the 11th international conference on parallel problem solving from nature, PPSN-2010*. Springer, 2010.

[17] P. Wawrzynski, "Real-time reinforcement learning by sequential actor-critics and experience replay," *Neural Networks*, vol. 22, pp. 1484–1497, 2009.

[18] P. Wawrzynski and A. K. Tanwani, "Autonomous reinforcement learning with experience replay," *Neural Networks*, in press, doi:10.1016/j.neunet.2012.11.007.

[19] B. Papis and P. Wawrzyński, http://sourceforge.net/projects/dotrl/.

[20] E. Evans, *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley, 2003.

[21] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, no. 12, pp. 243–269, 2000.

[22] A. Bogdanov, "Optimal control of a double inverted pendulum on a cart," CSEE, OGI School of Science and Engineering, OHSU, Tech. Rep. CSE-04-006, December 2004.

[23] J. H. Connell and S. Mahadevan, Eds., *Robot learning*, ser. The Kluwer international series in engineering and computer science. Boston: Kluwer Academic Publishers, 1993, index.

[24] M. T. Rosenstein and A. G. Barto, "Robot weightlifting by direct policy search," in *In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, 2001, pp. 839–844.

[25] H. Kimura and S. Kobayashi, "Reinforcement learning using stochastic gradient algorithm and its application to robots," in *IEE Japan Trans. on Electronics, Information and Systems*, vol. 119, 1999, pp. 913–934.

[26] C. Boutilier, R. Dearden, and M. Goldszmidt, "Exploiting structure in policy construction," in *IJCAI-95, pp.11041111*, 1995.

[27] H. Kimura and S. Kobayashi, "An analysis of actor/critic algorithm using eligibility traces: Reinforcement learning with imperfect value functions," in *Proceedings of the 15th international conference on machine learning*, 1998, pp. 278–286.

[28] C. J. C. H. Watkins and P. Dayan, "Technical note q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

[29] G. A. Rummery and M. Niranjan, "On-line q-learning using connectionist systems," Cambridge University Engineering Department, Tech. Rep., 1994.

[30] A. Harris, *Pro IronPython*, 1st ed. Berkely, CA, USA: Apress, 2009.

[31] X. Inc., http://www.mono-project.com.

[32] Wikipedia, "Mono (software)," 2013, [Accessed 09-May-2013]. [Online]. Available: http://en.wikipedia.org/wiki/Mono_%28software%29

[33] B. Tanner. [Online]. Available: http://code.google.com/p/rl-viz/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 137–141

An Emotional Learning-inspired Ensemble Classifier (ELiEC)

Mahboobeh Pasrapoor, Urban Bilstrup
School of Information Science, Computer and Electrical Engineering (IDE), Halmstad University

Abstract— In this paper, we suggest an inspired architecture by brain emotional processing for classification applications. The architecture is a type of ensemble classifier and is referred to as 'emotional learning-inspired ensemble classifier' (ELiEC). In this paper, we suggest the weighted k-nearest neighbor classifier as the basic classifier of ELiEC. We evaluate the ELiEC's performance by classifying some benchmark datasets.

I. Introduction

Classification methods have been widely used in the area of science, engineering, industry, business and medicine; they can be used for classification problems e.g., anomaly detection, handwriting recognition, speech recognition and medical diagnosis. Among them, the data driven classification approaches e.g., neural network-based models and neuro-fuzzy-based methods are the most popular methods due to the self-adaptive and high generalization capability. However, they have some significant issues: over fitting, model complexity and the curse of dimensionality, etc. [1]-[5]. Thus, developing new classification models to increase the classification's accuracy while resolving the mentioned issues are an open research topic in data mining. In this paper, a new classification model is suggested that can be considered as an ensemble classification with a different integration mechanism and combination algorithm. The model is an emotionally inspired model and is named 'brain emotional learning-inspired ensemble classifier' (ELiEC).

The rest of this paper is organized as follows: Section II reviews some works in classification and emotional learning–based models. Section III explains the ELiEC's structure. In Section IV, the benchmarks classification data sets are examined by ELiEC and the obtained results compared with other methods. Finally in Section V, we conclude and recommendsome possible future improvements to ElIEC.

II. A Brief Review

A. Related works to Classification methods

Numerous artificial intelligence-based methods have been proposed for classification problems. They can be categorized as: inductive or transductive, statistical-based or non-statistical–based, supervised or unsupervised methods. One popular group is supervised classification methods that include statistical methods (e.g., Naïve Bayes), non-statistical methods (e.g., neural network), instance based learning and support vector machine. Given a set of instances, these algorithms can assign an appropriate label to an unlabeled instance. Among the supervised learning methods, the support vector machine has the best performance in terms of classification accuracy; however, it has high time complexity that is a big issue for online classification applications [2] and [5].

Numerous efforts have been put into developing regularization methods to increase the generalization of supervised classification algorithms and reduce the time complexity of the learning procedure. A good example of the enhanced classification methods is the NFI model (Neuro-Fuzzy Inference Method for Transductive Reasoning) that provides a local model for each instance using a transductive reasoning system. The NFI model outperforms the neural network model in terms of accuracy and time complexity; nevertheless this model is not suitable for high dimensional classification applications [6]. Developing an ensemble-based [4] classifier is a major progress in addressing the misclassification and time complexity issues. The idea of ensemble-based classifiers is inspired by the human decision making process. The main components of an ensemble method are diversity generator and combiner. The former selects appropriate classifiers while the latter combines the classifiers' outputs. The combination mechanisms that have been developed can be divided into two subgroups: meta-combination and weighting methods. Choosing suitable diversity classifiers and combination procedures, the classification accuracy of ensemble-based classifiers outperforms the accuracy of each of the classifiers; however there is no adaptive procedure for choosing the classifiers and the combiner using information of classification problems.

Recently, a novel classification framework with the ability to adaptively tune the classifier's structure has been developed. This model, called meta-cognitive neural network (McNN), encompasses two components: a cognitive component and a meta-cognitive component [1],[5]. In McNN, the first component is a Radial Basis Neural Network and it is responsible for the change and optimization of the structure. The second component plays a role in choosing samples and the effective structure of the learning algorithm, obtaining knowledge from the training data. The McNN model and an extended version of that PBL-McRBFN [5] have shown excellent performance for classification problems.

The ELiEC model is a general purpose classification method that aims to reduce the misclassification rate and time complexity in classification applications. The ELiEC model

978-83-60810-53-8/$25.00

can be categorized as a type of ensemble classifiers that uses Wk-NN as the classifiers.

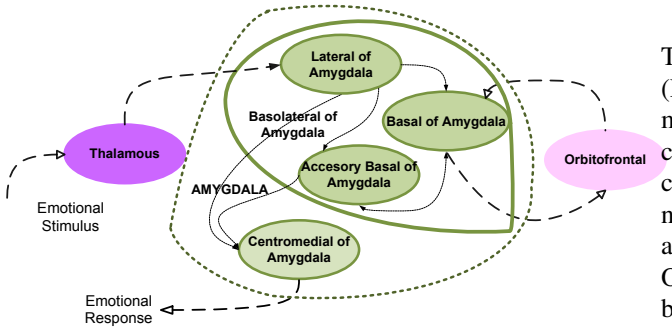


Figure1. The components of the amygdala and their connections.

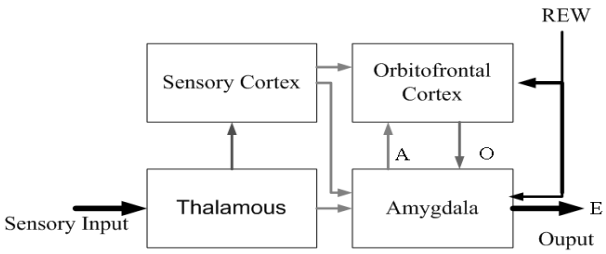


Fig. 2.The graphical description of amygdala-orbitofrontal subsystems

B. Brain Emotional Learning-based Models

Modeling the brain emotional processing and the memory to develop the artificial intelligence (AI-based) based tools has been an interesting topic in the machine learning research area. The emotionally-inspired AI tools are often referred to as the computational models of emotional learning. A well-known computational model is the amygdala-orbitofrontal system that defines emotional processing on the basis of the limbic system. The assumption of this model is that the limbic system, which consists of thalamus, sensory cortex, amygdala and orbitofrontal cortex, is mainly responsible for emotional learning in mammalian brains [8]. Fig. 1 depicts the connection between amygdala and its components. Fig. 2 describes the amygdala-orbitofrontal subsystem model that consists of four parts that interact with each other to mimic the emotional learning processing [8]. Due to its simple structure, the fundamental framework has been used for developing AI tools for control applications and and nonlinear system prediction. [12]-[17].

One popular emotionally inspired controller is BELBIC[9] that has been proven to overcome uncertainty and complexity issues of other intelligent controllers. Studies [9]-[11] have proved that BELBIC outperforms many other controllers such as PID controller and linear controllers in terms of simplicity, reliability and stability.

The emotionally based prediction models have often been applied for chaotic time series prediction and have also shown improvements in prediction accuracy [12]-[17].

III. EMOTIONAL LEARNING-INSPIRED ENSEMBLE CLASSIFIER (ELIEC)

The brain emotional learning-based ensemble classifier (ELiEC) model has a similar architecture to our two previous modelsBELRFS and BELFIS [13], [15]. The ELiEC model consists of four main parts and imitates the internal connection of the emotional system. The parts of ELiEC are named as : TH, CX, AMYG and ORBI that **are** referred to as THalamous, sensory CorteX, AMYGdala and ORBItofrontal cortex**.** The ELiEC model and the connection between these parts are described in Fig. 3.

For a classification problem, we define the set of training data as $(\mathbf{x}_1,\mathbf{c}^1),...,(\mathbf{x}_i,\mathbf{c}^i)...,(\mathbf{x}_n,\mathbf{c}^n)$, where $\mathbf{x}_i$ is an instance with m features and $\mathbf{c}^i$ determines the label class of $\mathbf{x}_i$. In a multi-class classification problem, we have n classes and the corresponding class of $\mathbf{x}_i$ which can be encoded as $\mathbf{y}^i = \ y_1^i,...,y_2^i,...,y_n^i$. If $\mathbf{c}^i$ is equal with j^{th} class, the value of y_j^i will be equal to one and other values will be zero. Using the following steps we explain how ELiEC classifies each instance in order to minimize the misclassification. The TH part evaluates the features of $\mathbf{x}_i$ and adds several extra features to $\mathbf{x}_i$. The extra features are calculated according to equation (1).

$$\mathbf{th}_i = [\max(\mathbf{x}_i), \text{mean}(\mathbf{x}_i), \min(\mathbf{x}_i)] \quad (1)$$

The CX evaluates the features of $\mathbf{x}_i$ and eliminates redundant features. The CX has a role to select the most informative features and eliminate the redundant features. Thus, the CX receives $\mathbf{x}_i$ with m features and provides $\mathbf{s}_i$ with l features ($l \le m$).

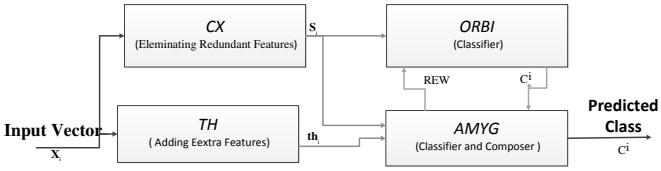


Fig. 3.The architecture of ELiEC.

The AMYG consists of a classifier and a combiner. The classifier that is represented as equation (4) predicts an appropriate class for $\mathbf{x}_i^a$ which is determined as equation (3). The combiner of the AMYG combines the outputs of the AMYG and ORBI to provide the final class as equation (5).

$$\mathbf{x}_i^a = [\mathbf{th}_i, \mathbf{s}_i] \quad (3)$$

$$C_a^i = \mathrm{Class}(\mathbf{x}_i^a) \quad (4)$$

The combiner strategy depends on the type of classification methods. In this paper, the Wk-NN method has been utilized as the classifiers of AMYG and ORBI. The combiner is another Wk-NN and determines the final class of the input vector $\mathbf{x}_i^c$ which is a vector as $\mathbf{x}_i^c = [\mathbf{x}_i^a, \mathbf{x}_i^o, C_a^i, C_o^i]$.

$$C^i = \mathrm{Class}(\mathbf{x}_i^c) \quad (5)$$

The ORBI is another classifier that can be a dependent classifier or independent classifier. For a dependent classifier the ORBI classifies the input vector $\mathbf{x}_i^o = [\mathbf{s}_i, \mathrm{REW}_i]$; while for an independent classifier the input vector is $\mathbf{x}_i^o = \mathbf{s}_i$. Finally, it forwards the classification result $C_o^i = \mathrm{Class}(\mathbf{x}_i^o)$ to AMYG. For the examples of this study an independent classifier is assigned to ORBI.

It should be noted that the classifiers of AMYG and ORBI can be defined on the basis of any supervised classification method, e.g., decision tree, single or multilayer perceptron, and support vector machine, etc. We can also form a meta-ensemble classifier by choosing an ensemble-based classifier for the AMYG and the ORBI.

IV. Weighted K- Nearest Neighbor : A Basic Classifier of ELiEC

Weighted k-nearest neighbor (Wk-NN) is a type of instance-based algorithm that has been widely used as a classification and regression method. For a given training set as: $(\mathbf{x}_1, c^1), ..., (\mathbf{x}_i, c^i) ..., (\mathbf{x}_{N_t}, c^{N_t})$, the Wk-NN determines the class of a test vector, $\mathbf{x}_{\text{test}}$,using the following steps [18]:

1) The Euclidian distance between $\mathbf{x}_{\text{test}}$ and $\mathbf{x}_i$ is calculated, $d_i = \|\mathbf{x}_{\text{test}} - \mathbf{x}_i\|_2$, in which each $\mathbf{x}_i$ is a member of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N_t}$, where N_t denotes the number of samples of the training data set.

2) The k minimum values of $\mathbf{d} = \{d_1, d_2, ..., d_{N_t}\}$ are selected as $\mathbf{d}_{\min}$. The $\mathbf{x}_i$ s that are corresponding to $\mathbf{d}_{\min}$ are the k nearest neighbors to $\mathbf{x}_{\text{test}}$ and define the local neighborhoods of the test vector, $\mathbf{x}_{\text{test}}$ [18].

3) The class label of $\mathbf{x}_{\text{test}}$ is chosen from the class labels of the local neighborhoods. Using the weighted k-nearest neighbor (W-kNN), a weight is assigned to each neighbor; the assigned weight is defined according the kernel function $K(.)$.

Any arbitrary function that holds the properties below can be considered as the kernel function [18].

1) For all d , $K(d) \geq 0$.
2) If $d = 0$ then $K(d)$ gets the maximum value.
3) If $d \to \pm\infty$ then $K(d)$ gets the minimum value.

In this paper, the kernel function is defined as (6). Thus, the closer neighbors to $\mathbf{x}_{\text{test}}$ have higher weights on estimating the class label of $\mathbf{x}_{\text{test}}$.

$$K(d) = \frac{\max(\mathbf{d}) - (d_j - \min(\mathbf{d}))}{\max(\mathbf{d})} \quad (6)$$

V. Benchmark Datasets

In this section, we test the ELiEC model to classify several data sets that have obtained from the University of California, Irvine (UCI) machine learning repository [20]. It should be noted that, each data instance has been normalized and the values of the attributes are scaled to the range [0,1]. For each benchmark, the cross validation set has the same size as the training data set. We test the classification model 100 times; however, we present the best results. To present an efficient comparison, we test the ELiEC model for both multiclass and binary class benchmark data sets. We also consider both balanced and imbalanced classification benchmark data sets. The performance of classification models are calculated using the average per class classification accuracy as (7) which is represented as η_a . The parameter n determines the number of classes and η_j indicates the classification accuracy of j class. The test error determines the number of misclassification samples of the test set; the samples of the test data set is determined by N_{test}

$$\eta_a = \frac{1}{N_{\text{test}}} \sum_{j=1}^{n} \eta_j \quad (7)$$

A. Multiclass and Well Balanced Benchmark Data Sets

Iris data set and wine determination are two well-known classification benchmarks that have been categorized under the well balanced and multiclass data sets. Iris data set waobtained from the University of California, Irvine (UCI) machine learning repository [20] and has been examined by several classification methods e.g., NFI, PBL-McRBFN, etc [1],[5],[21],[22]. This data set consists of 150 samples; each sample has 4 attributes named as: sepal length, sepal width, petal length and petal width. Iris data set is a set with multiclassand consists of three classes: iris setosa, iris versicolor and iris virginica [20].

As the first case study, 50 percent of iris data sets are randomly chosen as the training data set; while the rest of the samples are considered as the test data set. Table I compares the number of misclassifications using ELiEC, NFI and MLP. The results indicate that using ELIEC the average number of test errors is equal to 2.2, that is less than the two other methods.

Table I. Comparison of performance classification of BELBEC and other methods for Iris data set with 50% as the training data set and 50% test data set.

Classificati on Model	Specification of results		
	Structure	Test error	The number of training data
BELBEC	15 neighbor	2.2	75
NFI[6]	6.3 neighbor	3.3	75
MLP[6]	12 neuron	4.6	75

Table II. The classification accuracy of BELBEC for the Iris data set with 45 samples as training data and 105 samples test data.

Classificati on Model	Specification of results		
	Structure	The average per class accuracy	The number of training data
BELBEC without normalized	16 neighbor	99.02 %	45
McNN[1]	5 neuron	97.14 %	22
PBL-McRBFN[5]	6 neuron	98.10 %	20
BELBEC	18 neighbor	%96.24	60
McNN[1]	9 neuron	98.49 %	27
PBL-McRBFN[5]	11 neuron	98.69 %	29
ensemple of (OC-SVM) [22]	----	92.00%	----

As the second case, the ELiEC classifies iris set using 45 samples as the training data set..

The wine data set is the second data set that is used to evaluate ELiEC. This data set that is a multiclass and a well-balanced data set; it is obtained from the chemical analysis of some wines that are produced from various cultivars in the same region of Italy. The sample of the data set consists of 13 features and can be categorized into three classes. Table II compares the results of ELIEC and three methods: McNN, PBL-McRBFN and another method types of ensemple classifiers; this model was named OC-SVM that is referred to as one-class support vector machines[22]. In this experiment, we use 60 samples as the training data samples and 118 samples as the test data. It is observable that ELiEC has good results in classification of this well-balanced data set

As was mentioned, the ELiEC is based on W-kNN; thus, we compare the results of ELiEC and W-kNN for classifying the wine data set (see Table II). We also compare the ELiEC and W-kNN to classify wine data with two different structures. As Table III shows, the classification accuracy of ELiEC has no noticeable difference when the total number of neighbors of ELiEC's classifiers is only eight neighbors. However, when we define a lower number of neighbors for Wk-NN, its accuracy has a significant decrease.

Table III. Comparison between WKNN and BELBEC for the wine data set.

Classificati on Model	Specification of results		
	Structure	The average per class accuracy	The number of training data
BeLBeC	18 neighbor	%96.24	60
BeLBeC	8neighbor	%95.27	60
KNN	3 neighbor	%95.40	60
KNN	18neighbor	%91.60	60

B. Binary and Low Dimensional DatabaseLiver Disorders (LD) and PIMA Indian Diabetes are two medical data sets that have been tested by different classification methods. Liver Disorders (LD) is related to a blood test and has 345samples and 6 attributes, respectively. Pima Indians diabetes set is another medical data set and has been provided by the National Institute of Diabetes and Digestive and Kidney Diseases. It is used to diagnostic diabetes. The number of samples and number of attributes are 768 and 8 respectively. Table IV lists the results of ELIEC for the data sets and compares the results with three methods: McNN, PBL-McRBFN and a modification of adapting splitting and selection (AdaSS) [23]. It shows that the results of McNN, PBL-McRBFN and a modification of adapting splitting and selection (AdaSS)are better than the results of ELiEC; however, the ELiEC has a simpler structure than the two other methods.

Table IV. Comparison between different methods to classify the LD and PIMA. The number of training samples is equal to [5]

Classificati on Model	Specification of results		
	Structure	The average per class accuracy	DATA
BELBEC	48neuron	68.4	LD
McNN[1]	68 neurons	71.60	LD
PBL-McRBFN[5]	87 neurons	72.63	LD
BELBEC	100neuron	70.37	PIMA
McNN[1]	193 neurons	77.31	PIMA
PBL-McRBFN[5]	162 neurons	76.67	PIMA
MAD [23]	7 clusters and 5 classifiers	72.010	PIMA

V. CONCLUSION

This paper presents a new ensemble-based classifier that is inspired by the brain emotional network. The architecture is referred to as ELiEC and utilizes W-kNN as the basic classification method. However, the ELIEC differs from other ensemble methods in the way that the classifiers are fed (see Figure 2).

The performance of ELiEC is evaluated by classifying several benchmark data sets. The results indicate a fairly good performance of ELiEC for classification.

As future works, we replace W-kNN with other learning classification methods e.g., the support vector machine to address the time complexity and the curse of dimensionality issues. In addition, a random forest method can be used to form a meta-ensemble classifier.

REFERENCES

[1] G. S. Babu and S. Suresh, "Meta-cognitive neural network for classification problems in a sequential learning framework," Neurocomputing, vol. 81, pp. 86–96, Apr. 2012.

[2] S.B. Kotsiantis, I.D. Zaharakis, and P.E. Pintelas, "Machine Learning: A Review of Classification and Combining Techniques," Artificial Intelligence Review, vol.26, pp. 159-190, 2006.

[3] T.Phyu, "Survey of Classification Techniques in Data Mining", in.Proc. International Multi Conference of Engineers and Computer Scientists, vol, I IMECS 2009, 2009.

[4] L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, no. 1, pp. 1–39, 2 2010.S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, J. Informatica(31),pp. 249-26, 2007.

[5] G. S. Babu and S. Suresh, "Sequential Projection-Based Metacognitive Learning in a Radial Basis Function Network for Classification Problems," Neural Networks and Learning Systems, IEEE Transactions on , vol.24, no.2, pp.194,206, Feb. 2013

[6] Q.Song; N.K.Kasabov, "NFI: a neuro-fuzzy inference method for transductive reasoning," Fuzzy Systems, IEEE Transactions on , vol.13, no.6, pp.799,808, Dec. 2005

[7] J.Moren, C.Balkenius,"A computational model of emotional learning in the amygdala,'', in From Animals to Animats, MIT, Cambridge, 2000.

[8] C. Lucas, D. Shahmirzadi, N. Sheikholeslami, " Introducing BELBIC: brain emotional learning based intelligent controller,'' J. INTELL. AUTOM. SOFT. COMPUT., vol. 10, no. 1, pp. 11-22, 2004.

[9] R. M. Milasi, C. Lucas, B. N. Araabi, "Intelligent Modeling and Control of Washing Machines Using LLNF Modeling and Modified BELBIC," in Proc. Int. Conf. Control and Automation., pp.812-817, 2005,

[10] N. Sheikholeslami, D. Shahmirzadi, E. Semsar, C. Lucas., "Applying Brain Emotional Learning Algorithm for Multivariable Control of HVAC Systems," J. INTELL. **FUZZY.** SYST.vol.16, pp. 1–12, 2005.

[11] M. Parsapoor, C. Lucas and S. Setayeshi, "Reinforcement _recurrent fuzzy rule based system based on brain emotional learning structure to predict the complexity dynamic system," in Proc. IEEE Int. Conf. ICDIM,pp.25-32, 2008.

[12] M. Parsapoor, U. Bilstrup, "Brain Emotional Learning Based Fuzzy Inference System (BELFIS) for Solar Activity Forecasting," in Proc. IEEE Int. Conf. ICTAI 2012, 2012.

[13] M. Parsapoor, Lucas.C, Setayeshi.S, "Modifying Brain Emotional Learning Model for Adaptive Prediction of Chaotic Systems with Few Data Training Samples, '' in Proc. Int. Conf. ICAOR. pp. 328-341,2008.

[14] M. Parsapoor, M, U. Bilstrup, "Neuro-fuzzy models, BELRFS and LoLiMoT, for prediction of chaotic time series," in Proc. IEEE Int. Conf. INISTA., pp.1-5, 2012.

[15] C. Lucas, A. Abbaspour, A. Gholipour, B. Nadjar Araabi, M. Fatourechi, "Enhancing the performance of neurofuzzy predictors by emotional learning algorithm,'' J. Informatica (Slovenia)., vol. 27, no. 2 pp.165–174, 2003.

[16] T. Babaie, R. Karimizandi ,C. Lucas, "Learning based brain emotional intelligence as a new aspect for development of an alarm system,'' J. Soft Computing., vol. 9, issue 9 ,pp.857-873, 2008.

[17] G. Shakhnarovich, T. Darrell,and P.Indyk, Nearest-Neighbor Methods in Learning and Vision:Theory and Practice, MIT Press, March 2006.

[18] C. Blake and C. Merz. (1998). UCI Repository of Machine Learning Databases. Dept. Information & Computer Sciences, Univ. California, Irvine [Online]. Available: http://archive.ics.uci.edu/ml/.

[19] M. Fatourechi, C. Lucas, and A.K. Sedigh, "Emotional Learning as a New Tool for Development of Agent-based System," J. Informatica (Slovenia), vol. 27, no. 2, pp.137-144., 2004.

[20] C. Blake and C. Merz. (1998). UCI Repository of Machine Learning Databases. Dept. Information & Computer Sciences, Univ. California, Irvine [Online]. Available: http://archive.ics.uci.edu/ml/.

[21] M. Wozniak and B. Krawczyk, "Combined classifier based on feature space partitioning, " J.Applied Mathematics and Computer Science vol. 22, no4, pp. 855-866., 2012.

[22] B. Cyganek, "One-Class Support Vector Ensembles for Image Segmentation and Classification, ", J. Mathematical Imaging and Vision, vol. 42, no 2-3, pp.103-117 2012.

[23] R.Burduk, "New AdaBoost Algorithm Based on Interval-Valued Fuzzy Sets, " J. IDEAL pp. 794-801, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 143–144

Autonomous Input Management for Human Interaction-Oriented Systems Design

Michal Podpora
Opole University of Technology
Faculty of Electrical Engineering,
Automatic Control and Informatics
ul. Sosnkowskiego 31, 45-272 Opole, Poland
Email: michal.podpora@gmail.com

Aleksandra Kawala-Janik, Mary Kiernan
University of Greenwich,
School of Computing and Mathematical Sciences,
Old Royal Naval College
Park Row, SE10 9 LS London, UK
Email: {a.d.kawala-janik, m.kiernan}@greenwich.ac.uk

***Abstract*—In this paper evaluation of a policy-based algorithm for video inputs switching is presented. The term 'data quality' is not trivial for Human-Machine Interaction systems, yet a simple and efficient algorithm is needed for choosing the most valuable video source. This becomes particularly important for systems that support functional decomposition of image processing algorithm, which are designed for non-optimal working environment. In this paper an autonomous input management system is proposed, which consists of a data quality evaluation algorithm and a simple decision algorithm.**

Keywords – Human-Machine Interaction (HMI), Machine Vision, Decision Systems, Distributed Systems, Autonomic Systems

I. Introduction

COMPUTER vision has become a popular information source for computer and robotic systems interacting with humans, however the tests were conducted mostly in laboratory conditions. Real-life applications may cause some challenges as various environmental conditions are taken into account, although some biological mechanisms enable to handle both data aquisition and information processing, which exceed the capabilities of a very complex acquisition subsystems, such as – being able to see in the dark or being able to read and understand a text with the majority of illegible letters. Computer systems are supposed not only to entertain but also to support and protect humans, therefore they should contain more efficient, than the human one, information processing system. Possible future implementaion of these acquisition systems with the application of speech and/or vision include inter alia fire rescue teams supporting systems.

II. Policy-Based Input Switching

It is possible to implement all available acquisition systems and to process all the input information (thermal video, night vision), but in real-life applications it is not efficient because the data streams may be too massive and require high computing power for real-time processing in order to enable embedded system-based implementation. Transmission of multiple video streams to a remote workstation/server is also complicated due to the limited bandwidth, however it is possible to choose one input at a time and transmit it (or process it locally) [1], if the system was able to judge the value of the data input quality of its acquisition subsystems. The quality evaluation could be performed on the basis of static threshold values (e.g. any measure of image noise or edges quality), but in that case it would be just a simple condition. The Policy-Based Input Switching (PIS) becomes especially useful when adjusting/changing the policy for choosing an input depending on circumstances or environment. The proposed PIS conception is not a set of rules and threshold values, but an entire framework offering wrapper functionality (similarly to [2]) for additional modularity and as a foundation for autonomous policy reloading/changing. A semi-autonomous mobile robot designed for a fire team support could enable to change its policy for operating in a particular environment and to change its policy autonomously depending on any transient environment parameters.

III. Subjective Quality Evaluation Procedure

The Subjective Quality Evaluation Procedure (SQEP) is implemented in the core of the PIS framework and is able to decide locally on the active input (as a part of policy algorithm) and on the active policy (as a part of wrapper's autonomous module). Whilst the SQEP is intended to be run locally. It should involve only simple and efficient operations for data input evaluation. In a pilot study basic policy was implemented for choosing the best visual data input out of three available inputs: (1) video stream, (2) night vision, (3) thermography. The exemplary SQEP was implemented to choose between (1) and (2) basing on brightness and noise and then between (1|2) and (3) basing on the width of histogram of (1|2). Fig. 1 shows the value of a quality coefficient calculated for two different acquisition subsystems, acquiring visual information of the same scene and objects and in the same temporal context. Acquisition starts in darkness, and c.a. 128th frame a dim light source is turned on. The light is bright enough to turn off the infrared lighting, but not bright enough for the computer vision camera to acquire data of good quality. It is highly beneficial to have a SQEP procedure implemented in the PIS system, where the algorithm of a policy (and its rules or conditions, parameters and decisions) is intended to be easily replaceable and changable at runtime, without any

978-83-60810-53-8/$25.00

modifications to the inputs and outputs of PIS policy wrapper. The wrapper's functionality is extended to support autonomous policy exchange.

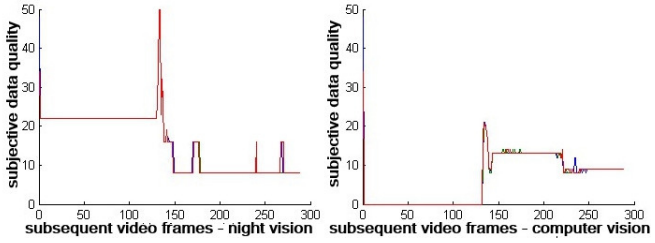


Fig. 1. Compartment of two acquisition subsystems, showing data quality during a change of lighting in the scene.

IV. SQEP for Autonomous Input Management

The main goal of using the proposed framework is to extend the functionality, versatility and robustness of the acquisition subsystems of vision-based mobile robots. It can only be achieved if the outputs of the Acquisition Subsystems and the outputs of PIS system will be developed to become modular and replaceable, so the Subjective Quality Evaluation Procedure (SQEP) should handle the evaluation of all inputs available in the hardware and all possible combinations (subsets) of these inputs and to produce the same set of predefined output variables and data structures. The SQEP should offer not only the possibility of on-demand input switching, but most of all it should perform the input selection procedures autonomously, if there is a better data quality input than the currently processed one. The decision about which input to process should be done locally to prevent transmitting of all data inputs to remote servers and the inputs should be evaluated in a simplest possible way to save the CPU cycles of local/mobile/embedded system. The most important feature of SQEP is the capability of autonomous data quality evaluation. Only one input is being transmitted and the most valuable data input is connected to the Cognitive/Decision System at a time. The coefficients and rules (the algorithm) of input quality evaluation can be changed at runtime by PIS system.

V. PIS for Autonomous Policy Management

The SQEP has proven to be useful, as it is beneficial in order to stay connected to a good data quality input, even in a changing environment. The basic PIS functionality is the possibility to change selection criterions – the policy, such as different environments as a very attractive feature. The most useful feature of PIS system, however, is the possibility to change the selection criteria in runtime (without rewriting the code, compiling, uploading, running). Thus there is no need to interrupt learning/execution or to reset the system's memory/knowledge/execution. The PIS system, in its most sophisticated version is expected to enable changing its policy 'intentionally'. In some cases it has very little effect on a mobile system, but sometimes it may be crucial feature for the accomplishment of the horizontal goal of the system.

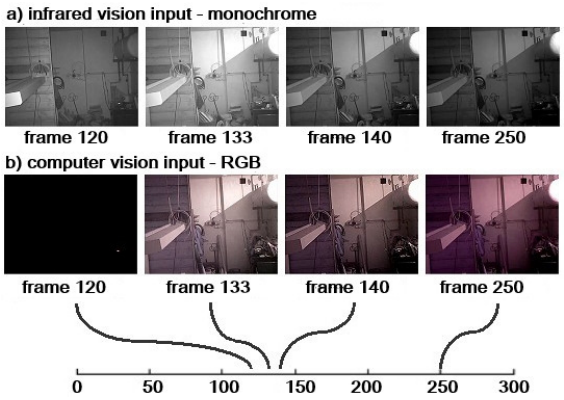


Fig. 2. The exemplary visual data input: darkness (frame 1..127) and with a dim light turned on (frame 128..288). Frames 128..140 are too bright due to the automatic white balance feature.

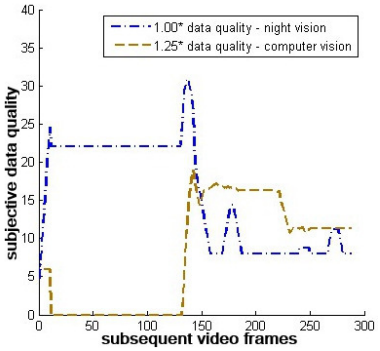


Fig. 3. An exemplary simple data quality coefficient for SQEP system for a policy-based input switching system, during a change of lighting in the scene.

VI. Advantages for HMI Systems, Future Work

Future work will implement inter alia virtual agents [3] and bio-signals. Use of virtual agents would make the system more intuitive. Initial tests were run with the application of the face mimic and the Emotiv EPOC headset. Future work will also involve development of a standalone application (AE) connecting Emotiv EPOC headset with any embedded platform. The work will be advanced by implementing various bio-signals [4]. The proposed control architecture should also be improved in order to reduce appearance of potential control errors, which is a common issue in autonomic systems.

References

[1] M. Podpora, 'Dynamic re-definition of Region-of-Interest in Vision Systems Feedback', *Proceedings of the 2nd International Students Conference on Electrodynamic and Mechatronics*, 2009, IEEE eXplore.

[2] M. Pelc, R. Anthony, 'Towards Policy-Based Self-Configuration of Embedded Systems', *System and Information Sciences Notes,* vol.2(1), 2007, pp. 20–26.

[3] M. Ochs, C. Pelachaud, D. Sadek, 'An Empathic Virtual Dialog Agent to Improve Human-Machine Interaction', *AAMAS '08 Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems,* vol. 1, 2008, pp. 89-96.

[4] F. Casacuberta, J. Civera, E. Cubel, A. Lagarda, G. Lapalme, E. Macklovitch, E. Vidal, 'Human Interaction for High-Quality Machine Translation', *Communications of the ACM – A View of Parallel Computing CACM Homepage archive,* vol. 52, 2009, pp. 135-138.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 145–151

Knowledge-based Named Entity Recognition in Polish

Aleksander Pohl
Jagiellonian University
ul. Łojasiewicza 4, 30-348 Kraków, Poland
Email: aleksander.pohl@uj.edu.pl

***Abstract*—This document describes an algorithm aimed at recognizing Named Entities in Polish text, which is powered by two knowledge sources: the Polish Wikipedia and the Cyc ontology. Besides providing the rough types for the recognized entities, the algorithm links them to the Wikipedia pages and assigns precise semantic types taken from Cyc. The algorithm is verified against manually identified Named Entities in the one-million sub-corpus of the National Corpus of Polish.**

I. Introduction

IN THE recent years the research conducted in the field of Information Extraction (IE) has brought many interesting results. First of all at the end of the twentieth century researchers have overcome the first major obstacle preventing wide-spread adoption of IE systems. Namely, by incorporating weakly supervised methods they were able to quickly adopt systems to new domains, by providing only a handful of training examples. Works of Brin [1] as well as Agichtein and Gravano [2] showed that large-scale information extraction is feasible, provided that we have access to large corpora, WWW in particular.

The ideas from the previous century were developed in two, slightly contradictory directions. First of all the term Open Information Extraction was introduced in works of Banko et al. [3]. This approach towards IE was pushed further in systems such as ReVerb [4] which does very well in the task of relation extraction. The goal of Open IE, much influenced by Information Retrieval is extraction of *all* information found in web-documents. Since it is very hard to build an ontology covering all the phenomena that might be described on the Web, these methods make very minimalistic assumptions about the world. As a result the extracted information is not transformed into some coherent semantic framework, but rather it is assumed that the user will find the relevant information by refining his/her query, the same as in traditional search engines.

The opposite approach towards IE comes from the researchers from the SemanticWeb camp. The primary difference comes from the fact that they are trying to utilize various knowledge-sources, especially taxonomies and ontologies, in order to extract the information available in the textual data and transform them into well established and broadly accepted schemas. In that form the information might be further automatically processed and consumed by intelligent systems. As a result, there are several important problems that have to be resolved, which are absent in Open IE systems. First of all, the concepts that are extracted must be identified within some ontology. The ambiguities also have to be resolved, especially among the relations. That approach is characteristic for the systems such as DBpedia Spotlight [5] and AIDA [6], both utilizing Wikipedia as the reference resource for concept identification. Recently Exner and Nugues [7] showed how relation extraction might be performed combining well known Natural Language Processing (NLP) techniques and DBpedia both as a reference conceptual scheme as well as source of examples used for training relation classifiers.

However if we look at the development of IE methods that are employed for Polish, we will discover that the state of research resembles the state of research in English in mid-nineties. To the best of our knowledge there are only two systems [8], [9] that perform template filling. Both of them employ manually constructed extraction grammars and operate in closed domains – the first one extracts information from mammogram reports and the second from patients' diabetic records. Regarding relation extraction there is only limited research in the field. [10] describes methods for automatic extraction of semantic relations aimed at development of the Polish Wordnet. The described method uses massive amounts of data in order to determine the best location for a new synset in a large semantic network. On the other hand [11] describes an ontology-based method for the extraction of relations form a bibliographical lexicon based on manually defined grammars. The only IE task that is well developed is Named Entity recognition[1], but the methods still rely on manual construction of grammars or tagging of large amounts of data.

As we know from the development of systems aimed at English these bottlenecks might be overcome if appropriate methods are employed. This work is a step in this direction. It presents an algorithm that does not require any manual construction of extraction grammars, nor tagging of large amounts of data. By utilizing Wikipedia link structure and classification of Wikipedia articles into the Cyc ontology [12] the system is able to precisely classify and link the entities to Wikipedia articles. Since there are knowledge-bases such as DBpedia [13] that are based-upon Wikipedia and Cyc is the largest known manually constructed ontology, this opens interesting possibilities for automatic processing of the extracted data by intelligent systems.

[1]The next section contains a detailed description of the available systems.

978-83-60810-53-8/$25.00

TABLE I
EVALUATION OF SPROUT-BASED NER SYSTEM PRESENTED IN [19].

Category	Precision [%]	Recall [%]
People	90.6	85.3
Locations	88.0	43.4
Organizations	87.9	56.6

TABLE II
EVALUATION OF SPROUT-BASED NER SYSTEM PRESENTED IN [20]. ONLY THE BEST RESULTS ARE REPORTED.

Category	Precision [%]	Recall [%]
People	91.0	78.0
Locations	82.0	72.0
Organizations	92.0	52.0

II. RELATED WORK

Named entity recognition (NER) is a vast topic. There are many approaches towards it and systems that actually perform NER. To constrain the review of such systems we present only the work on NER in Polish.

The research on NER in Polish starts with the works of Piskorski [14], [15], [16]. These works describe the adaptation of the SProUT platform [17] for Polish and its application in the recognition of entities such as: time, percentage, money, organizations, locations and people.

Applying SProUT to NER requires manual definition of the rules, which in general consist of a pattern/action pairs built upon typed feature structures (TFS). The LHS of a rule matches segments with particular features and the RHS defines the TFS that is produced as the result of that match. Besides basic matching of words and features, the rules allow to use variables and call other rules. As a result the formalism is both expressive and transparent.

The adoption of SProUT for Polish required integration of a morphological analyzer for Polish – Morfeusz [18]. Also gazeteers were used and the morphological variants were produced automatically. The author elaborates on problems that are specific to Slavonic languages, Polish in particular, especially the rich morphology of names as well as problems with name segmentation. This causes serious problems for NER since there are cases where it is impossible to determine the base form of a NE without appealing to data such as verb frame subcategorization, which is usually out of scope of NER. The constructed rules were evaluated on a set of 100 news articles. The values are presented in Table I.

A more elaborate evaluation of this approach is presented in [20], where it is applied in the domain of cadastral information. The described system used a more elaborated entity classification taxonomy consisting of nearly 30 classes. It it also evaluated on a much bigger corpora counting more than 4 thousand entities. The evaluation methodology was more elaborate were exact and partial matches were counted separately. The results of the evaluation are presented in Table II.

A more lightweight approach is presented in works of Graliński et al. [21], [22] and Walas et al. [23]. The authors find SProUT formalism too complex for the tasks of NE translation and anonymization. [21] identifies errors in NE translation as one of the important problems that make the translation hard to understand. What is more 10% of errors in machine translation (MT) are due to invalid processing of named entities. [22] uses NER in task of anonymization of police reports that are used to improve MT. Since the researchers are not allowed to see the names of the suspects, they provide a set of Word macros that are used to remove sensitive data (such as names of people and companies) from the documents. [23] describes NER in the context of a question answering (QA) system.

Each of these systems uses a formalism based on Spejd [24]. The rules are less complex than in SProUT so the results are less precise. For instance in [21], depending on the evaluation scheme the precision varies between 76 to 88%. On the other hand [21] and [23] report that special handling of NEs improves both MT and QA.

Only recently the researchers of Polish NER systems started to use machine learning approach. Marcińczuk at al. [25] describes application of Hidden Markov Models (HMM) for this task, which was restricted to the recognition of names of people and organizations. The authors used HMM with 7 hidden states for each NE type. The transitions between states were modeled by the maximum likelihood over the training data, while tag emissions were generated by n-gram character language models with generalized form of Witten-Bell smoothing [26]. The authors used LingPipe [27] to compute the parameters of the HMM.

The system was trained on a corpus containing stock exchange reports with 670 person mentions and more than 3 thousand company mentions. It was evaluated on the same corpus (using 10-fold cross validation) and on a corpus containing police reports. The vanilla HMM reached 86% for people names and 80% for companies names (F_1 metric) on the stock exchange corpus. In order to improve the precision of the system the authors applied two heuristics: filtering and trimming. The first one imposed some restrictions on the form of the names, the second one stripped preceding and following words if they did not start with an upper case. The first heuristic improved the precision of the method from 63% to 89% with only small reduction of recall. The authors also report the evaluation of the system on the police report corpus. The results were significantly worse (55% F_1 metric), showing that this approach towards NER is highly domain-dependent.

The latest achievement in Polish NER is described in [28]. The authors used Conditional Random Fields (CRF) model [29] to overcome the limitations of HMM model, e.g. ignorance of the right context of the name. They were detecting 5 types of entities: first names, surnames, country names, city names and road names. The developed model incorporated not only statistical data, but also knowlege, e.g. the list of names of peoples positions and titles taken from the Polish WordNet [10]. In general there were 5 types of features that were used to build the model: ortographic, binary-ortographic, WordNet-based, morphological and gazetteer-based.

The model was trained on one corpus (used in the previous works of Marcińczuk et al. [25]) and tested on that corpus following 10-fold cross-validation scheme and on two others corpora (police reports corpus and corpus of economic news). The results on the first corpus substantially outperformed the HMM model, reaching F_1 score of 92.5%. On the other hand the evaluation performed on the two other corpora showed that the drop in performance is significant – the model achieved 67.7% F_1 score on the police reports corpus and 72.3% F_1 on the economic news corpus, meaning that the model fits well to the training data, but causes problems when ported to different (even highly related, like in the case of stock exchange reports and economic news) domains.

The primary difference of the presented algorithm with the already described methods comes from the fact that this algorithm does not require any manual construction of NE recognition rules, nor tagging of the training data. As such it does not require any manual intervention, besides the *selection* of the name types that should be identified. The other difference comes from the fact that this algorithm is tested on the 1-million sub-corpus of the National Corpus of Polish [30], to the best of our knowledge the first such demonstration. This is particularly important, because the corpus contains a variety of language styles and text sources, contains a large number of NEs and allows for a comparison of different algorithms in a unified setting.

III. Entity recognition

The algorithm employed to recognize NEs is an adaptation of the Word Sense Disambiguation (WSD) algorithm described in [31]. That algorithm builds on the Wikipedia Miner disambiguation algorithm described in [32]. The original algorithm of Milne and Witten works as follows. First of all it recognizes occurrences of phrases used in Wikipedia to link to other Wikipedia articles. Since many of these phrases are ambiguous, e.g. *Washington* may refer to a state, a city, a person, a football club, a university and almost four hundred other concepts, the algorithm tries to disambiguate its meaning by employing machine-learning techniques. First of all it computes several features of each candidate target concept:

1) an average weighted *semantic relatedness* with other (unambiguous) concepts
2) a *probability* of the candidate concept
3) a *context goodness*, i.e. a measure showing if the context of the phrase is well defined

The *semantic relatedness* between two concepts[2] in the original algorithm is computed using the Normalized Google Distance (NGD) [33]:

$$sr_G(a,b) = 1 - \frac{log(max(|A|,|B|)) - log(|A \cap B|)}{log(|W|) - log(min(|A|,|B|))} \quad (1)$$

Where:

- $sr_G(a,b)$ – the measure of semantic relatedness between a and b,
- $|A|$ – the size of the set of articles that link to a,
- $|A \cap B|$ – the size of the set of articles that link both to a and b,
- $|W|$ – the number of all articles in Wikipedia.

The weights of the concepts are established according to their semantic relatedness with other concepts and the link probability of phrases they are identified by (e.g. the ratio of the number of phrase occurrences as link to the total number of phrase occurrences).

The *probability* of the candidate concept is computed as the ratio of the number of links with that phrase pointing to that concept to the total number of links containing that phrase.

When the features are computed, the algorithm tries to select the most probable meaning by employing a machine learning algorithm – C4.5 in this case [34]. The algorithm of Milne and Witten uses also Wikipedia link structure to train the classifier. Wikipedia links are used as positive examples for *phrase–concept* pairs. The negative examples are constructed from all the remaining concepts that could be linked via this phrase (i.e. there are links with the same anchor name that point to the other concepts). Since the internal link structure of Wikipedia is very dense, it is easy to generate millions of the training examples.

The primary improvement described in [31] is the usage of a Jaccard coefficient-inspired measure instead of the NGD:

$$sr_J(a,b) = \begin{cases} \frac{1}{1 - log\left(\frac{|A \cap B|}{|A \cup B|}\right)} & |A \cap B| > 0 \\ 0 & a \neq b \wedge |A \cap B| = 0 \\ 1 & a = b \wedge |A \cap B| = 0 \end{cases} \quad (2)$$

Where:

- $|A \cup B|$ – the size of the set of articles that link to a or b

The second improvement is the employment of several additional features that are already used in the original algorithm (but not in the classifier) or are trivial to compute:

- the *rank of the semantic relatedness* of the candidate concept among all the candidate concepts,
- the *rank of the probability* of the candidate concept among all the candidate concepts,
- the *link probability* of the concept.

By employing the different semantic relatedness measure and the additional features the automatically computed performance of the algorithm measured as the weighted precision and recall (F_1) jumped from 84 to 92.4% in the English dataset and from 83 to 91.7% in the Polish dataset. What is more by using the latest Polish Wikipedia dataset[3], it was measured that the performance of the algorithm was further increased to 93.5%.

Direct application of the WSD algorithm results in the identification of phrases that might be (unambiguously) linked to the Wikipedia articles. But its application in Named Entity Recognition requires several adjustments that are covered in the next section.

[2]Although this is not fully accurate, we identify the Wikipedia articles with concepts.

[3]Dump from the 2^{nd} of May 2013.

IV. ENTITY CLASSIFICATION

The regular definition of the Named Entity Recognition task [35] requires that the entities are recognized and *classified* into some well-defined conceptual scheme. Although in the time of MUC conferences [36] it was required that the classification scheme makes fine-grained distinctions allowing for e.g. distinguishing between civil and military objects[4], in recent years, e.g. in the tasks defined in the scope of ACE initiative [37], the conceptual scheme was substantially simplified. ACE 2008 defines 5 general types divided into 31 subtypes. The general types are as follows:

- facility
- geo-political entity
- location
- organization
- person

Similarly in the National Corpus of Polish [30] the following NEs are identified:

- personal names
- geographical names
- names of organizations and institutions
- words related to the above categories
- basic temporal expressions

This simplification of the conceptual scheme stems from the fact that these resources are domain-independent and as such, should make as few ontological assumptions as possible. Especially because in such domain-independent environment it is hard even for people to assign fine-grained classes, let alone force research teams to adopt one „golden" ontology.

But when it comes to the application of the results of IE in some specific domain, such sketchy schemas have limited value. If we wish to conduct intelligence analysis we are not interested in people in general, but those who have important roles in their societies. If we are considering financial investments, we are not interested in organizations „in general", but only those which operate on a specific market, e.g. mining companies. As a result, the more we are concerned with the application of the IE results, the more fine-grained conceptual scheme is needed.

In order to make the results of NEs recognition useful in practical settings, the Cyc ontology [38] is used as the taxonomy of the entities. Although there are alternatives such as YAGO [39] and DBpedia ontology [40], which are directly mapped to Wikipedia, both of them have certain deficiencies making them less useful for automatic processing of the extracted knowledge. Although YAGO's taxonomy is very dense, since its classes were extracted from Wikipedia's category scheme, it lacks features such as well defined disjointness relation. Although its authors have used heuristics to detect inconsistencies, they were error-prone. As a result there are contradictory facts in YAGO, e.g. *Gertrude Stein* is classified both as a *person* and as a *literary work*.

[4]It was assumed that a terrorism activity is targeted only at civil objects, such as churches, shopping centers and railway stations.

TABLE III
THE COVERAGE OF METHODS USED TO CLASSIFY THE ARTICLES IN THE POLISH WIKIPEDIA.

Method	Count (thousand)	Percentage
English Wikipedia [12]	283.9	25.8
Infobox mapping	534.9	48.6
People heuristic	213.1	19.4
Total	**688.6**	**62.6**

On the other hand the DBpedia ontology[5] is rather small (contains less than 400 classes) and lacks coverage. Since the entities are assigned to the classes according to the infoboxes attached to the articles and less than 2 million of articles (out of 4 million) have such infoboxes attached, many of the entities lack their type.

Cyc [38] on the other hand is a long running effort of building an ontology that would allow intelligent systems to perform common-sense inferences. As a result it contains thousands of well defined semantic classes. These classes are not only organized into subsumption relation, but they are also inter-related with disjointness relation. This feature of Cyc makes it particularly useful for ensuring high accuracy of the extracted informations, since inconsistencies in type assignment might be resolved automatically.

Although Wikipedia is not aligned with Cyc in its entirety, in our recent effort [12] we created a classification of more than 2.2 million of the English Wikipedia entities into the Cyc ontology with precision reaching 93%. That mapping was made available in the N-triple format, with links to the English DBpedia[6] and OpenCyc[7]. In order to use that mapping in the Polish Wikipedia the following procedure was employed:

1) The interlingual links from the latest DBpedia[8] were downloaded.
2) The interlingual links were used to establish a mapping between the English and the Polish DBpedia entities.
3) The Cyc types were assigned to the corresponding Polish DBpedia entities.

There were 626 thousand of links in the English DBpedia that were employed in the second step. But since the type assignment covered only 2.2 million of articles (out of 4 million) the number of classified articles in the Polish Wikipedia was 284 thousand. To extend the coverage of the method infoboxes from the Polish Wikipedia were mapped to terms in Cyc and a simple heuristic for people marking all articles with *Urodzeni w* (Eng. *Born in*) and *Zmarli w* (Eng. *Died in*) categories as instances of `Person` class was applied. The first heuristic yielded 535 thousand of classifications and the second 213 thousand. Since the methods were overlapping, the final result was a classification of 689 thousand of Polish Wikipedia articles (out of 1.1 million). The results are summarized in Table III.

[5]`http://dbpedia.org/Ontology`
[6]`http://klon.wzks.uj.edu.pl/wiki-types/`
[7]`http://www.cyc.com/platform/opencyc`
[8]`http://wiki.dbpedia.org/Downloads38`

The second adjustment which was applied to the Wikipedia-based disambiguation algorithm concerned filtering of the entities that were provided as the NER result. Since contents of Wikipedia is encyclopedic, it covers descriptions both of classes and individuals. The WSD algorithm does not make distinction between these types of entities. However, the result of NER should be a discovery of *names*, i.e. proper names (usually) of individual things. We are not expecting from the NER system to annotate occurrences of words such as *woman*, *child* or *man*, because they are instances of common nouns. Yet, the algorithm is able to properly disambiguate these words and its original version produces such results. So the filtering was concerned with rejecting the instances of common nouns and similar expressions from the final result.

In general, making the distinction between classes and individuals is not a trivial task. Since it was not the primary concern of this research, a simple, yet powerful heuristic was employed: for each Wikipedia article the total number of links starting with an uppercase letter and a lowercase letter was counted. If the number of links starting with an uppercase letter was greater or equal to the number of links starting with a lowercase latter, the article was considered as describing an individual. A small test performed on a random sample of 200 entities showed that this heuristic is able to produce distinctions with precision above 90%. This result is very good compared to the results of a similar task of distinguishing between types and instances in the Wikipedia category system [41], where several heuristics were employed, yielding similar precision.

V. Evaluation

The evaluation of the algorithm was performed on the one-million sub-corpus (ONEM) of the National Corpus of Polish [42] with manually annotated named entities [30], [43]. ONEM contains excerpts from many sources: literature, newspapers, the Internet, speech transcripts and other. It contains 3888 excerpts which amount to more than 50 thousand of NEs.

The annotation covered identification of names referring to people, locations, organizations and temporal expressions. Some of these types were subdivided into subtypes. Table IV summarizes the taxonomy of the entities. It should be noted that not only the nominal expressions were annotated, but also all other types of expressions that refer to NEs, adjectives in particular. So in the following expressions *restauracja w **Warszawie*** (Eng. *restaurant in Warsaw*) and ***warszawska** restauracja* (Eng. *Warsaw restaurant*), the boldfaced fragments were annotated as referring to *Warsaw*. These occurrences of NEs are called *relational expressions*.

Although, in general proper names do not follow the principle of compositionality, since their contents might be arbitrary, it was decided that the annotation covers also subordinate names, which are parts of larger units. E.g. in the expression *Dom dziecka w Oruni* (Eng. *Orphanage in Orunia*), the whole expression is annotated as a name of an organization, while *Oruni* is annotated as a name of a location. From the point of view of the evaluation of NER this gives some added value, since algorithms might be rewarded for accurate partial matches within longer expressions.

TABLE IV
The types of entities annotated in ONEM [30].

Type	Subtypes	Description
persName	forename surname addName	given name family name nickname
orgName		an organization
geogName		a geographical entity (excludes geopolitical entities)
placeName	district settlement region country bloc	a district within a city a city or a village a country region a country a bloc of countries
date		at least partially determined date
time		at least partially determined time

TABLE V
The counts of entities annotated in ONEM. Time and date expressions are skipped.

Entity type	Test corpus	Tuning corpus
persName	19619	336
orgName	10914	217
geogName	4001	69
placeName	15432	311
total	**49966**	**933**

Regarding the scope of the evaluation – it did not cover date and time expressions, since the WSD algorithm is not aimed at theses types of expressions. What is more, it is apparent that these expressions are domain independent and can be captured by a well developed grammar or some other NER technique. On the other hand the evaluation included the relational expressions, which are quite hard to spot using traditional NER methods.

During the developement of the algorithm, the ONEM corpus was split into two parts – one containing 100 randomly selected text excerpts, used for the tuning of the algorithm and the other containing 3788 text excerpts. The counts of entities in both of the corpora are given in Table V. The results reported in the next section are based only on the second part of the corpus.

The tuning corpus was used solely for the recognition of the Cyc types that were present in the corpus and for their mapping to the corresponding name types in ONEM. The mapping is summarized in Table VI. The mapping covers only the most general Cyc types. If a more specific Cyc type was assigned to the entity, the appropriate type was selected using the generalization relation (in Cyc called `genls`). Since some of the specific types generalize to more than one of the general types (e.g. a *Country* generalizes both to a *GeopoliticalEntity* and an *Organization*) the first matching type was selected (according to the priority presented in Table VI).

TABLE VI
THE MAPPING BETWEEN ONEM TYPES AND CYC TYPES.

Priority	ONEM type	Cyc type
1	persName	Person HumanGivenName HumanSurname Saint God FictionalCharacter PersonTypeByEthnicity PersonTypeByOccupation
2	placeName	GeopoliticalEntity PopulatedPlace
3	geogName	GeographicalPlace Territory AstronomicalBody Place
4	orgName	Organization

We should noted however that the precise Cyc type was not discarded in that procedure. The mapping was provided in order to compare the results of the algorithm with the coarse-grained annotation from the ONEM corpus. In the applications of the NER algorithm the Cyc types are easily accessible since they are attached to the Wikipedia articles, indentified by the disambiguation algorithm.

VI. RESULTS

The results of the evaluation are given in Table VII. The system was evaluated only against the whole NE units. So if the system recognized a geographical name inside a name of an organization, the result was counted as invalid, even though there was an annotation that captured the geographical name as a subordinate of the organization name. It is because we believed that the non-compositionality of NEs is their primary characteristic and usually only the whole lexical unit should be recognized and annotated. The internal lexical structure of a given NE has only anecdotal value for information extraction. What is more, if more date (such as the location of a given place or an architectural structure) regarding the entity is needed, it can be found in the DBpedia knowledge base.

The table contains the results for exact matches as well as partial matches. An *exact match* is counted as a match which is exactly the same as in the annotation, so any characters such as quotation marks have to be present in the system provided annotation, even though they do not carry much information. A *partial match* is counted for a system-provided annotation which is completely covered by the reference annotation. So this excludes matches that only partly overlap.

The direct comparison with the performance of the other NER systems shows that the primary deficiency of the algorithm is low recall. Although such comparisons should be made with care, since none of the systems described in section II was tested against ONEM corpus.

Overall recall of 41.8% for the exact matches and 48.6% for the partial matches is below expectations for a NER system. It should be noted however that the coverage of the classification of the entities was not complete (only 62.5%),

TABLE VII
THE RESULTS OF THE EVALUATION OF THE KNOWLEDGE-BASED NER ALGORITHM. P - PRECISION, R - RECALL, F_1 - $2 * P * R/(P + R)$

	Exact match			Partial match		
Entity type	**P** [%]	**R** [%]	$\mathbf{F_1}$ [%]	**P** [%]	**R** [%]	$\mathbf{F_1}$ [%]
persName	**95.2**	34.6	50.8	**93.7**	46.6	62.2
orgName	82.7	35.8	50.0	73.3	42.5	53.8
geogName	78.2	34.4	47.8	68.3	37.7	48.6
placeName	91.8	**57.0**	**70.3**	91.5	**58.3**	**71.2**
overall	**90.0**	**41.8**	**57.1**	**86.3**	**48.6**	**62.2**

so we could expect better results if all Wikipedia entities had a type assigned. What is more – the system tries not only to detect the type of the entity, but also to disambiguate it against the Wikipedia articles. As a result, only the entities that have a corresponding entry in Wikipedia might be recognized.

Regarding precision, the algorithm achieves similar results as the other described systems and in some cases (names of people and places) even outperforms the already known methods. Still the performance for organization and geographical names is below our expectations. A manual inspection of the results showed that the low precision might be at least partly a result of the organization – place name distinction used by the annotators.

It is reported [30] e.g. that a name of a country might be treated both as a name of a place and an organization, depending on the context. This causes serious problems for the algorithm, since each entity has only one type assigned. But this seems to cause problems not only for the algorithm, but also for the annotators. E.g. in the sentence ***Amerykańska*** *prasa twierdzi, że mimo oficjalnego poparcia kampanii* ***USA*** *przez rząd* ***Pakistanu***, ... (Eng. ***American*** *press says that despite the official support of the* ***US*** *campaign by the government of* ***Pakistan***, ...), *American* and *Pakistan* are marked as names of places, while *US* is marked as a name of an organization. Such annotation does not seem to be coherent. Suffice it to say that the creators of ACE NER tasks [37] introduced the geopolitical entity type to resolve this kind of problem.

VII. CONCLUSIONS

We presented a knowledge-based algorithm for named entity recognition in Polish texts. It was tested on the one-million manually annotated sub-corpus of the National Corpus of Polish. The results obtained by the algorithm show that this method, although achieving reasonably good precision, has low recall. Definitely a combination of the presented method with grammar-based method would give much better results, since the coverage of this algorithm is limited by the contents found in Wikipedia. What is more, the classification of Wikipedia entities is not complete and further worsens the recall.

On the other hand it should be stressed that the algorithm does not require any manual work, neither in the construction of multi-word grammars, nor in tagging of large amounts of textual data. What is more, the classification scheme based

on the Cyc ontology allows for a direct consumption of the results of the algorithm by intelligent systems. This feature of the algorithm together with the moderately good results show that further research in this direction should be pursued.

REFERENCES

[1] S. Brin, "Extracting Patterns and Relations from the World Wide Web," in *The World Wide Web and Databases*. Springer, 1999, pp. 172–183.

[2] E. Agichtein and L. Gravano, "Snowball: Extracting relations from large plain-text collections," in *Proceedings of the fifth ACM conference on Digital libraries*. ACM, 2000, pp. 85–94.

[3] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open Information Extraction from the Web," in *IN IJCAI*, 2007, pp. 2670–2676.

[4] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 1535–1545.

[5] P. Mendes, M. Jakob, A. García-Silva, and C. Bizer, "DBpedia Spotlight: shedding light on the web of documents," in *Proceedings of the 7th International Conference on Semantic Systems*. ACM, 2011, pp. 1–8.

[6] M. Yosef, J. Hoffart, I. Bordino, M. Spaniol, and G. Weikum, "Aida: An online tool for accurate disambiguation of named entities in text and tables," *Proceedings of the VLDB Endowment*, vol. 4, no. 12, 2011.

[7] P. Exner and P. Nugues, "Entity Extraction: From Unstructured Text to DBpedia RDF Triples," in *Proceedings of the Web of Linked Entities Workshop in conjuction with the 11th International Semantic Web Conference*, G. Rizzo, P. Mendes, E. Charton, S. Hellmann, and A. Kalyanpur, Eds., 2012, pp. 58–69.

[8] A. Mykowiecka, A. Kupść, and M. Marciniak, "Rule-based medical content extraction and classification," *Intelligent Information Processing and Web Mining*, pp. 237–245, 2005.

[9] M. Marciniak and A. Mykowiecka, "Automatic processing of diabetic patients' hospital documentation," in *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*. Association for Computational Linguistics, 2007, pp. 35–42.

[10] M. Piasecki, S. Szpakowicz, and B. Broda, *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej, 2009.

[11] W. Jaworski, "Ontology-based content extraction from polish biobibliographical lexicon," in *Recent Advances in Intelligent Information Systems*. EXIT, 2009, pp. 27–40.

[12] A. Pohl, "Classifying the Wikipedia Articles into the OpenCyc Taxonomy," in *Proceedings of the Web of Linked Entities Workshop in conjuction with the 11th International Semantic Web Conference*, G. Rizzo, P. Mendes, E. Charton, S. Hellmann, and A. Kalyanpur, Eds., 2012, pp. 5–16.

[13] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," *The Semantic Web*, pp. 722–735, 2007.

[14] J. Piskorski, "Automatic named-entity recognition for Polish," in *Proceedings of the International International Workshop on Intelligent Media Technology for Communicative Intelligence, Warsaw, Poland*, 2004.

[15] J. Piskorski, P. Homola, M. Marciniak, A. Mykowiecka, A. Przepiórkowski, and M. Woliński, "Information extraction for Polish using the SProUT platform," *Intelligent Information Processing and Web Mining, Advances in Soft Computing*, pp. 227–236, 2004.

[16] J. Piskorski, "Named-entity recognition for Polish with SProUT," in *Intelligent Media Technology for Communicative Intelligence*. Springer, 2005, pp. 122–133.

[17] W. Drozdzynski, H.-U. Krieger, J. Piskorski, U. Schäfer, and F. Xu, "Shallow Processing with Unification and Typed Feature Structures — Foundations and Applications," *Künstliche Intelligenz*, vol. 1, pp. 17–23, 2004.

[18] M. Woliński, "Morfeusz—a practical tool for the morphological analysis of Polish," *Intelligent information processing and web mining*, pp. 511–520, 2006.

[19] J. Piskorski, "Extraction of Polish Named-Entities," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC*, 2004, pp. 313–316.

[20] W. Abramowicz, A. Filipowska, J. Piskorski, K. Węcel, and K. Wieloch, "Linguistic Suite for Polish Cadastral System," in *Proceedings of the LREC*, vol. 6, 2006, pp. 53–58.

[21] F. Graliński, K. Jassem, and M. Marcińczuk, "An environment for named entity recognition and translation," in *Proceedings of the 13th Annual Conference of the European Association for Machine Translation, Barcelona, Spain*, 2009, pp. 88–95.

[22] F. Graliński, K. Jassem, M. Marcińczuk, and P. Wawrzyniak, "Named Entity Recognition in Machine Anonymization," *Recent Advances in Intelligent Information Systems*, pp. 247–260, 2009.

[23] M. Walas and K. Jassem, "Named entity recognition in a Polish question answering system," *Intelligent Information Systems*, pp. 181–192, 2010.

[24] A. Buczyński and A. Przepiórkowski, "Spejd: A shallow processing and morphological disambiguation tool," in *Human Language Technology. Challenges of the Information Society*. Springer, 2009, pp. 131–141.

[25] M. Marcińczuk and M. Piasecki, "Named Entity Recognition in the Domain of Polish Stock Exchange Reports," *Intelligent Information Systems, Siedlce*, pp. 127–140, 2010.

[26] I. Witten and T. Bell, "The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression," *Information Theory, IEEE Transactions on*, vol. 37, no. 4, pp. 1085–1094, 1991.

[27] B. Carpenter and B. Baldwin, *Text Analysis with LingPipe 4*. LingPipe Inc, 2011.

[28] M. Marcińczuk, M. Stanek, M. Piasecki, and A. Musiał, "Rich Set of Features for Proper Name Recognition in Polish Texts," in *Security and Intelligent Information Systems*. Springer, 2012, pp. 332–344.

[29] J. Lafferty, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data." Morgan Kaufmann, 2001, pp. 282–289.

[30] A. Savary, J. Waszczuk, and A. Przepiórkowski, "Towards the Annotation of Named Entities in the National Corpus of Polish," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC 2010*, 2010.

[31] A. Pohl, "Improving the Wikipedia Miner Word Sense Disambiguation Algorithm," in *Proceedings of Federated Conference on Computer Science and Information Systems 2012*. IEEE, to appear.

[32] D. Milne and I. Witten, "Learning to link with Wikipedia," in *Proceeding of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 509–518.

[33] R. Cilibrasi and P. Vitanyi, "The Google similarity distance," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, no. 3, pp. 370–383, 2007.

[34] J. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann, 1993.

[35] M. Moens, *Information extraction: algorithms and prospects in a retrieval context*. Springer-Verlag New York Inc, 2006, vol. 21.

[36] R. Grishman and B. Sundheim, "Message understanding conference-6: A brief history," in *Proceedings of COLING*, vol. 96, 1996, pp. 466–471.

[37] NIST, "Automatic Content Extraction 2008 Evaluation Plan (ACE08)," 2008. [Online]. Available: http://www.itl.nist.gov/iad/mig/tests/ace/-2008/doc/ace08-evalplan.v1.2d.pdf

[38] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[39] F. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 697–706.

[40] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia-A crystallization point for the Web of Data," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 3, pp. 154–165, 2009.

[41] C. Zirn, V. Nastase, and M. Strube, "Distinguishing between instances and classes in the wikipedia taxonomy," *The Semantic Web: Research and Applications*, pp. 376–387, 2008.

[42] A. Przepiórkowski, M. Bańko, R. L. Górski, and B. Lewandowska-Tomaszczyk, *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, 2012.

[43] A. Savary and J. Piskorski, "Lexicons and grammars for named entity annotation in the National corpus of Polish," *Intelligent Information Systems, Siedlce, Poland*, pp. 141–154, 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 153–158

Tabu Search approach for Multi–Skill Resource–Constrained Project Scheduling Problem

Marek E. Skowroński, Paweł B. Myszkowski, Marcin Adamski, Paweł Kwiatek
Institute of Informatics, Department of Artificial Intelligence
Faculty of Computer Science & Management, Wrocław University of Technology, Poland
Email: {pawel.myszkowski, m.e.skowronski}@pwr.woc.pl, {183750,180040}@student.pwr.wroc.pl

***Abstract*—In this article two approaches of Tabu Search in Multi–Skill Resource–Constrained Project Scheduling Problem (MS–RCPSP) have been proposed, based on different neighbourhood generation methods. The first approach assumes swapping resources assigned to pair of tasks, while the second one proposes assigning any resource that could perform given task. Both approaches need to respect the skill constraints. The objective of this paper is to research the usability and robustness of proposed approaches in solving MS–RCPSP. Experiments have been performed using artificially created dataset instances, based on real–world instances, got from Volvo IT and verified by experienced project manager. Presented results show that Tabu Search (TS) based methods are efficient approaches that could be developed in the further work.**

I. Introduction

RESOURCE-Constrained Project Scheduling Problem (RCPSP) is a classical problem (e.g. [3], [11]). Its objective is to assign scarce resources to given tasks to make the project schedule as short / cheap as possible. Because of its combinatorial nature, it is known as NP–hard [1]. It means that it is not possible to find optimal solution in finite, polynomial time. It also suggests using some soft computing methods, which do not always provide optimal solutions, but usually sub–optimal acceptable solutions in reasonable processing time. There are several metaheuristics [12] used for solving RCPSP and its extensions – Evolutionary Algorithms (EA) [4], [6], [7], [9], [15], [22], [26], Simulated Annealing (SA) [2], [5], [14], Ant Colony Optimization (ACO) [16] or Tabu Search (TS) [17], [18], [20], [21], [24]. The last of mentioned methods would be investigated in this paper. We decided to develop TS–based methods because of its relative simplicity in comparison to other metaheuristics. On the other hand, it could be comparably effective to other methods.

RCPSP could be extended by the skills domain to Multi–Skill Resource–Constrained Project Scheduling Problem (MS–RCPSP) [8], [19]. Each task requires given skill in specified familiarity level, while each resource disposes some skills pool. It causes that not every resource can perform every task and the schedule is more difficult to be built.

(MS–)RCPSP is very practical problem. Project managers in significant companies still often need to schedule their projects manually, what is extremely time consuming. Because of human fail-ability, manually scheduling could also cause obtaining infeasible schedule solutions, where not every constraints are satisfied. Providing computer–aided methods could save a lot of time and ensures that every designed constraints would be met. What is also important, automated AI–based methods mentioned earlier could give final solution in minutes, while experienced project manager needs a few hours to prepare a schedule for the same–sized project (with the same number of resources and tasks).

Proposed (MS–)RCPSP definition has been designed in strict cooperation with Volvo IT Department in Wroclaw. Provided requirements, assumptions and constraints have been considered, approved and then included in problem description, what enlarges the practical value of this paper.

The rest of the paper is organised as follows. Section II describes other approaches to solve the (MS–)RCPSP using metaheuristics, especially TS. Section III presents the MS–RCPSP problem statement, while Section IV describes the approaches proposed in this paper. Section V provides conducted experiments of proposed methods in a given dataset. Finally, section VI presents the conclusions of obtained results and suggests some ideas of future work.

II. Related work

To make a structured overview of related work, we divided it into groups of methods. EA–based [4], [6], [7], [9], [15], [22], [26] methods mostly use task–vector representation of an individual. The order in a vector describes the order of tasks' performance in a project. Schedule can be generated in a serial [7], [22], [25], [26] or parallel [12] generation scheme. Because of proposed individuals representation, mostly semi–blind crossover and mutation operators could be applied, like swap [22], [23], [26] or insert [7], [25] mutation and one–point [7], [9], [25], two–point [7], position–based [26] or uniform [15] crossover. Some papers presents more dedicated operators, like peak crossover in [22].

SA–based approaches [2], [5], [14] are also often investigated. In [2] classical precedence–feasible activity list representation with the serial generation scheme is used. The neighbourhood is created by insertion method of regarded task within its last predecessor and the first successor. Cooling scheme assumes annealing of multiple starting solutions. The other approach of generating a neighbourhood is presented in [5], where the new solution in neighbourhood is created randomly, preserving precedence– and resource constraints. In [5] some hybrid of TS and SA has been proposed, where the memory of moves has been added to traditional SA approach.

978-83-60810-53-8/$25.00

The last of mentioned approaches, presented in [14] assumes investigating set of SA–based methods with multiple start, where the initial solution is prepared in various ways, e. g. randomly or using scheduling priority rules.

In TS–based methods [17], [18], [20], [21], [24] several elements can be variously implemented. The neighbourhood can be created by swapping or inserting activities [20] in the task–order list. In [20] the tabu list size is dependent on the number of critical tasks to be scheduled in a project. What is more, several strategies of diversification have been proposed there, while not better solutions have been recently found and strategies of intensification, where a solution with very good quality has been found. In [21] a starting solution is obtained using the minimal slack heuristic rule (MINSLK). The same approach assumes regarding move as an exchange of two activities positions in the activity list and the *difference move* measure is introduced, computed as the difference between solutions which are regarded by given move. An aspiration criterion is also proposed in [21] – if the move is on a tabu list, but produced solution (S) is better than the best (B) found so far, S replaces B. The termination criterion is set as a number of iterations which could not find better solution. Similar starting solution, using priority rule heuristics, is proposed in [17], [18], where classical neighbourhood generation methods have been also proposed: swap and insertion. In an approach presented in [24] classical TS method has been extended by local search procedures, to find better solutions.

Other (MS–)RCPSP heuristic solutions have been included in [12], [13], where local search methods and other population based methods have been also presented, e. g. ACO–based approach, that was also investigated in [16].

III. Problem statement

In MS–RCPSP we assume that project consists of several main elements: tasks, precedence relations, resources and skills. **Tasks** are described by their start and finish dates, duration and skill required to be performed. Tasks are often related by **precedence relations** – some tasks cannot be started before their potential predecessors would not be finished. Analogously, task's successor cannot be started before it's finish date. **Resources** are described by their salary and **skills** pools they own. As it was mentioned earlier, not every resource can perform each task, when given resource does not own skill required by given task. Only one resource can be assigned to given task.

A. Conflicts, solution's feasibility

Furthermore given resource cannot be assigned to more than one task in an overlapping period of time – if such a situation occurs, we defined it as a **conflict** and it has to be resolved. Conflict solving is made by shifting one of conflicted tasks just after the other one in the timeline. Because conflict resolving could disturb the precedence relations constraints, they should be preserved after each conflict resolution. Without resolving conflicts and preserving critical path constraints, produced solutions would be infeasible and could not be regarded as final, correct schedules. Solutions where resource is assigned to task when it does not own required skill in specified level is also regarded as infeasible.

B. Evaluation function

The (MS–)RCPSP objective is to schedule the project as quick or / and cheap as possible. It could be presented as **multi–objective optimization problem**: project schedule's **duration minimization** and project's performance **cost minimization**. Those objectives are generally in opposition. Reducing a project duration could cause enlarging a project's cost and vice versa. That is why, the *happy medium* is often sought – how to reduce the value of first objective, to get larger but still acceptable value of the second objective. In project scheduling problem domain it is often called *time & cost trade–off problem*.

A single project schedule (PS) solution is represented as a resource–to–task assignments vector $A = [a_i^j] : i = 1, 2, ..., t$, $j = 1, 2, ..., r$, where a_i^j represents the assignment of j-resource to i-task, t-number of tasks and r - number of resources.

To evaluate a solution, we proposed weighted, evaluation function:

$$\min f(PS) = w_\tau f_\tau(PS) + (1 - w_\tau) f_c(PS) \tag{1}$$

where: w_τ - weight of duration component, $f_\tau(PS)$ - duration evaluation component, $f_c(PS)$ - cost evaluation component.

Components' weights are applied to tune up the importance of time and/or cost factor in the given project optimization. The duration–aided optimization means setting time weight close to value 1 that automatically reduce the cost weight near zero. Analogously weights in the cost–aided optimization are tuned.

The time component $f_\tau(PS)$ is calculated as follows:

$$f_\tau(PS) = \frac{d_{PS}}{\tau_{max}} \tag{2}$$

where: d_{PS} - duration of schedule PS, τ_{max} - maximal possible duration of schedule PS, computed as the sum of all tasks' duration. The cost component $f_c(PS)$ is defined as follows:

$$f_c(PS) = \frac{\sum_{i=1}^{t} c_i^j}{c_{max} - c_{min}} \tag{3}$$

where: c_i^j - standard cost of performing task i by resource j, c_{min} - minimal schedule cost – a total cost of all tasks assigned to the cheapest resource, c_{max} - maximal schedule cost – a total cost of all tasks assigned to the most expensive resource. c_{max} and c_{min} do not involve skill constraints. It means that c_{min} value could be reached only for non–feasible solution. Analogously to c_{max}.

C. Solution space size

Given number of tasks and number of resources, we can estimate the solution space size, as:

$$SS(t, r) = t! * r^t \tag{4}$$

However, that estimation takes also into account non–feasible solutions, because skill–constraints are not satisfied. To give an example, let's assume $t = 10$ and $r = 5$ – without any precedence relations we get $SS(10, 5) = 3.54 * 10^{13}$ combinations. It is worth mentioning, that each task can be placed only once in schedule, but resources could be assigned more often. An extreme situation occurs if the same one resource would be assigned to perform each task.

Large solution space size makes impossible checking each of the combinations manually. However, space includes also non–feasible solutions that do not satisfy defined conditions. Moreover, given example is a simplification and in real world problems we meet a higher number tasks (about $t = 100$) and resources ($r = 20$) – it gives $SS(100, 20) = 1,19 * 10^{288}$ of all solutions. As solution space is constrained, relatively large and the MS–RCPSP problem is NP–hard it proves the legitimacy of metaheuristics usage.

IV. Proposed approaches

The overall TS approach is to avoid entrainment in cycles by forbidding moves which take the solution, in the next iteration, to point in the solution space previously visited (hence tabu). TS proceeds according to the supposition that there is no point in accepting a new (poor) solution unless it is to avoid a path already investigated. This insures new regions of a problems solution space will be investigated in with the goal of avoiding local minima and ultimately finding the desired solution.

To perform, TS needs some parameters to be set. The neighbourhood size defines the size of neighbourhood, that is created based on currently best solution and from which the better solution is sought. Number of iterations stands how many times the neighbourhood would be generated and sought for better solution (stopping criterion). Tabu list size tells, how many recent swaps should be recorded in the list of forbidden ones. For simplicity in this paper, we decided not to introduce any aspiration criteria for TS in our investigations.

A. Initial solution generation

Initial solution is straightforwardly loaded from a file that contains the project data. The project schedule had been previously prepared using EA with classical crossover (one–point) and mutation (swap–based) operators that satisfies all constraints (precedence, skills and conflicts resolving). Therefore, it could be regarded as a feasible solution. The feasibility of initial solutions has been confirmed and approved by experienced project manager from Volvo IT.

B. Neighbourhood generation

To generate a neighbourhood, a new solution generation method has to be provided. We designed new solution as generated in two general steps: setting assignments to tasks and then build the schedule, respecting precedence constraints and resolve conflicts. We proposed and compared two methods of setting assignments to tasks. The first one bases on the swapping resources within the pair of tasks, while the second approach assumes assigning any resource that is capable of performing given task. Above mentioned rules provide the feasibility of generated solution.

Swap–based neighbourhood (SBN): In this approach potential solution is created in the following way. For given assignment another is sought that enables swapping resources between those assignments. Swapping is possible only if both resources are capable of being assigned to tasks related to chosen assignments – skill constraints are preserved. If no other assignment, which can be used for swapping with given one, is found, new assignment is selected and the procedure of searching *swapping mate* for it is repeated.

Random–based neighbourhood (RBN): A new potential solution, that can be visited by the TS procedure, is created by changing the assignment of given task in following way. List of resources that can perform given task (dispose skill required by task) is obtained and then any different resource than currently assigned is chosen.

Fig.1 presents schematically ways of generating new solutions that can be included into new neighbourhood. OK signs in this figure presents which resources can perform given tasks. In RBN, we consider only one task, for which new resource is sought. In SBN, we need to find two tasks and resources that are capable to swap assignments between them. The *QX.Y* notation describes skills owned by resources or required by particular task to be performed. In provided example task *T1* requires skill *Q2* at proficiency level *2* to be performed. Moreover, resource *R1* owns skill *Q1* at proficiency level equal *3* and skill *Q2* at proficiency level equal *2*. Thus R1 is able to perform *T1*. *R1* is also able to perform task *T4* because *R1* owns skill (*Q1*) required by *T4* (level 1) with higher proficiency level (*3*). Analogously R1 is also capable of performing *T3*.

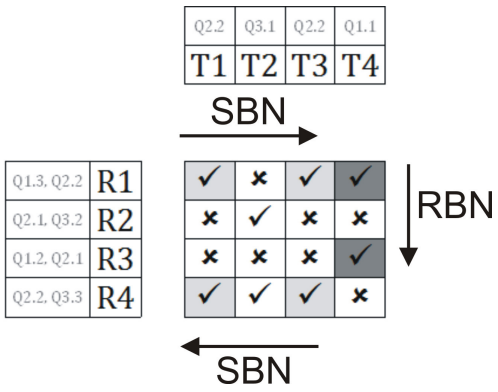


Fig. 1. Illustration of neighbourhood generation methods

Example results of neighbourhood generation methods have been presented in the Fig.2. For SBN only resources for tasks $T1$ and $T3$ could be swapped. Hence, as a result, swapping of mentioned pair of resources assigned to indicated tasks has been performed. For RBN other resource could be assigned to every task. In this example, we decided to change the assignment of $T4$. This example explains the main difference – SBN involves two tasks, while RBN changes assignment of

only one task.

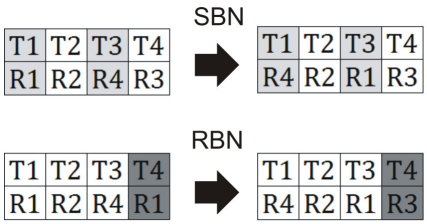


Fig. 2. Example of neighbourhood generation methods

After setting an assignments' vector, the schedule is built in a few steps. First, all tasks have their start time dates set to the project's start date. Then the conflict fixing mutually with the critical path preserving methods are launched, until no conflicts would be found and the schedule could be regarded as a feasible. Neighbourhood generation methods are designed not to allow generating infeasible solutions.

C. Move description

A move describes how the solution B has been created in the base of solution A. In our approach, the move stores the information about task and resource related to assignments that have been changed in neighbourhood generation procedure. In SBN the information about pair of tasks and pair of resources within which the swap has been performed.

Let's see some example: task A has been assigned to resource 1 - *A(1)* and task B has been assigned to resource *B(2)*. After the swap the situation is as follows: *A(2)* and *B(1)*. The information stored in move is as follows: *A(1),B(2)→A(2),B(1)*.

In RBN less information is stored in particular move. It is because this neighbourhood creation method involves only one task. Hence, using previous example, the information stored in a move looks like: *A(1)→A(2)*.

V. Experiments and results

The goal of conducted experiments was to compare two different approaches of creating neighbourhood in TS and investigate whether proposed TS approaches could be used in solving MS–RCPSP in effective way. To evaluate solution – the resulted project schedule – its duration time ([days]) and performance cost ([c.u][1]) were investigated.

A. Dataset

Due to evaluate not only the project schedule duration, but also the cost of the schedule, we cannot use the standard PSPLIB benchmark dataset [10], that does not contain any information about the task performance cost. What is more, PSPLIB dataset instances do not reflect the MS–RCPSP. Hence, we prepared the dataset, containing six project instances, that have been artificially created[2], in a base of real–world instances, got from the Volvo IT Department in Wroclaw.

[1]Currency units

[2]http://www.ii.pwr.wroc.pl/˜myszkowski/scheduling

TABLE I
MS–RCPSP dataset description

Property	D1	D2	D3	D4	D5	D6
Tasks	100	100	100	200	200	200
Resources	20	10	5	40	20	10
Skills	9	9	9	9	9	9
Relations	20	26	22	133	148	129

The dataset summary has been presented in the Table I. There are two groups of created project instances: one contains 100 tasks and the second – 200 tasks. Within the group, project instances are varied by number of available resources and the precedence relationship complexity. It led to create three different project instances both with 100 and 200 tasks. The skill variety has been set up to constant 9 different skill types for each project instance, while any resource can dispose no more than six different skill types. Because of the different resources and relations number, the scheduling complexity for each project is varied.

B. Experiments' set–up

The experiments have been divided into investigating the influence of TS parameters' configurations for project duration and performance cost in three various components' weights in evaluation function: duration optimization (DO: $w_t = 1$), balanced optimization (BO: $w_t = 0.5$) and cost optimization (CO: $w_t = 0$). Each experiment for given parameter configuration has been repeated ten times.

To present the average results in detail, a standard deviation measure should be introduced and applied to each average value. However, to keep the results' presentation clear and simple, we decided not to present it. It is because we have obtained the standard deviation measure values and they generally were not much significant. Only 2 of conducted experiments provided the standard deviation value bigger than 10% of the average.

The processing time of both approaches was varied within the range from about 40 seconds (D1 project instance with neighbourhood size = 10) to about 900 seconds (D6 project instance with neighbourhood size = 45)[3].

C. Experiments' performance

During the experiments, following parameters have been examined: tabu list size and neighbourhood size. Each parameter configuration was ran in both neighbourhood creation methods. Number of iterations has not been investigated. During experiments we noticed that after, in average, 200 iterations, no better solutions have been produced, regardless the remaining parameters' configuration and used dataset instance. Based on that observation, we set up the iterations size as a constant value equal to 250 iterations. Neighbourhood size and tabu list size parameters have been chosen experimentally.

For DO and CO, the best result has been indicated as the one with the smallest value of duration and cost respectively.

[3]Intel Core 2 Duo P8700 (2.53 GHz at each core) and 4 GB memory RAM.

If there was more than one result with the smallest value of time or cost, the one with the smallest value of the second optimization aspect has been chosen as the best. For BO the best result has been obtained based on the smallest value of evaluation function.

D. Results' discussion

Analysing results we can notice that optimization direction (more duration– or more cost–oriented) is related with the neighbourhood creation method. Looking at the Tab.**??** the best results for DO have been found in SBN method part in 5 from 6 dataset instances. On the other hand, the best CO results have been all obtained using RBN. An interesting observation regards the results for BO optimization mode. Project instances consisting 100 tasks have been optimized the best using SBN method, while the best optimization results for project instances consisting 200 tasks have been found in the part of table regarding the RBN.

Analysing the influence of examined TS parameters to obtained results, some following observations could be indicated. Taking into account the influence of the neighbourhood size to the DO results, we can suggest using smaller neighbourhood size values. It is because that the best DO results have been obtained for the $N = 10$ (4 best solutions – *winners*). $N = 20$ and $N = 45$ provided best results only once per each N value parameters' configurations. On the other hand, all best solutions for CO have been found for $N = 45$. It could lead to some general conclusion that enlarging the search space could be beneficial for cost optimization but does not have to help obtaining better solutions, where project duration is the core aspect of optimization. Above mentioned observation could be derived to BO mode but with a respect to a conclusion made in previous paragraph – for BO best solutions of project instances consisting 200 tasks have been obtained for RBN with the $N = 45$. BO results for project instances consisting 100 tasks cannot allow us to make any general conclusions.

We have not found any interesting relationship between the tabu list size and the optimization's robustness, regardless the configuration of remaining parameters, chosen optimization mode or even project instance. We have found 4 best obtained solutions in DO or BO for the $TL = 10$, but that value has not been confirmed in remaining results as potentially good in optimization. It is very difficult to make any general conclusions and assumptions regarding the influence of tabu list size into the potential of optimization.

VI. Conclusions and further work

The best obtained solutions for each dataset instance in every optimization mode have been presented in Tab.II. For each solution a project duration and performance cost has been presented with the description of configuration, for which the solution has been found. The notation for that description is as follows: *neighbourhood creation method (neighbourhood size, tabu list size)*. E.g., SBN(10,15) means that the solution has been obtained for swap–based neighbourhood solution with neighbourhood size equal to 10 and tabu list size set to 15. That summary table briefly summarizes which neighbourhood creation mode is preferred for a given optimization mode.

Obtained results could lead to a conclusion that neighbourhood creation strategy has significant influence on the optimization ability. Using SBN, better solutions in duration optimization were obtained. SBN provided the best duration optimization results in 5 of 6 project instances. On the other hand RBN generally is more effective in cost optimization. Hence, different approaches could be used for different optimization modes. Despite that, end user would have to be always aware of opposite–character of project duration and performance cost objectives, which cause enlarging one aspect where the optimization is focused on the second one.

TABLE II
SUMMARY TABLE – BEST OBTAINED RESULTS IN INVESTIGATED OPTIMIZATION MODES

ID	DO		BO		CO	
	days	cost	days	cost	days	cost
D1	SBN(10,15)		SBN(45,10)		RBN(45,10)	
	32	40656	37	38939	129	30750
D2	RBN(10,10)		SBN(45,7)		RBN(45,10)	
	33	43542	49	34240	179	26444
D3	SBN(10,7)		SBN(20,7)		RBN(45,7)	
	51	40054	61	36100	133	31645
D4	SBN(45,10)		RBN(20,15)		RBN(45,7)	
	92	88720	125	50438	254	46371
D5	SBN(20,15)		RBN(45,10)		RBN(45,10)	
	179	80448	184	54181	481	52425
D6	SBN(10,15)		RBN(45,15)		RBN(45,15)	
	199	97978	222	75996	330	73126

To lift end–user of setting weights in evaluation function, both TS approaches could be merged in Pareto–based related mechanism. Mixing solutions from both approaches and choosing only the best, non–dominated ones could give the end user more flexibility of choosing the most appropriate schedule of proposed solutions pool. Furthermore two interesting future work directions could be indicated: investigating and applying aspiration criteria for both proposed approaches and creating an initial solution in more directed ways. To cope with the second mentioned idea, scheduling priority rules could be applied, while the first of proposed research directions could enhance the proposed method's robustness.

It would be also useful to compare obtained results with simple hill–climbing method – like TS without storing information about moves. It would provide information about the real usability of applying TS for MS–RCPSP.

References

[1] Blazewicz J., Lenstra J.K., Rinnooy Kan A.H.G.; Scheduling subject to resource constraints: Classification and complexity, Discrete Applied Mathematics (5), pp. 11-24, 1983.

[2] Bouleimen K., Lecocq H.; A new efficient simulated annealing algorithm for the resource-constrained project scheduling problem and its multiple mode version, European Journal of Operational Research (149), pp. 268-281, 2003.

[3] Brucker P., Drexl A., Mohring R., Neumann K., Pesch E.; Resource–constrained project scheduling: Notation, classification, models, and methods, European Journal of Operational Research (112), pp. 3–41, 1998.

[4] Chen Z., Chyu C.; An Evolutionary Algorithm with Multi–Local Search for the Resource-Constrained Project Scheduling Problem, Intelligent Information Management (2), pp. 220–226, 2010.
[5] Das P. P., Acharyya S.; Simulated Annealing Variants for Solving Resource Constrained Project Scheduling Problem: A Comparative Study, Proceedings of 14th International Conference on Computer and Information Technology, pp. 469–474, 2011.
[6] Hartmann S.; A competitive genetic algorithm for resource–constrained project scheduling, Naval Research Logistics (45), pp. 733–750, 1998.
[7] Hindi K. S., Yang H., Fleszar K.; An Evolutionary Algorithm for Resource–Constrained Project Scheduling, IEEE Transactions on evolutionary computation (6), pp. 512–518, 2002.
[8] Kadrou Y., Najid N.M.; A new heuristic to solve RCPSP with multiple execution modes and Multi-Skilled Labor , IMACS Multiconference on Computational Engineering in Systems Applications (CESA), pp. 1302–1309, 2006,
[9] Kim J.L.; Permutation–based elitist genetic algorithm using serial scheme for large–sized resource–constrained project scheduling, Proceedings of the 2007 Winter Conference Simulation Conference, pp. 2112–2118, 2007.
[10] Kolisch R., Sprecher A., PSPLIB - A project scheduling problem library, European Journal of Operational Research (96), pp. 205–216, 1996.
[11] Kolisch R., Serial and parallel resource-constrained project scheduling methods revisited: Theory and computation, European Journal of Operational Research (90), pp. 320–333, 1996.
[12] Kolisch R., Hartmann S., Experimental evaluation of state-of-the-art heuristics for the resource-constrained project scheduling problem, European Journal of Operational Research (127), pp. 394–407, 2000.
[13] Kolisch R., Hartmann S., Experimental investigation of heuristics for resource-constrained project scheduling: An update, European Journal of Operational Research (174), pp. 23-37, 2006.
[14] Liu S., Tukel O.I., Rom W.; Flexible Scheduling Approach for Resource-Constrained Project Scheduling Problems, Proceedings of the 7th World Congress on Intelligent Control and Automation, pp. 3522–3526, 2008.
[15] Mendes J.J.M., Gonvalces J.F., Resende M.G.C.; A random key based genetic algorithm for the resource constrained project scheduling problem, Computers & Operations Research (36), pp. 92-109, 2009.
[16] Merkle D., Mittendorf M., Schmeck H.; Ant Colony Optimization for Resource–Constrained Project Scheduling, IEEE Transactions on Evolutionary Computation (6/4), pp. 333–346, 2002.
[17] Pan H.I., Hsaio P.W., Chen K.Y.; A study of project scheduling optimization using Tabu Search algorithm, Engineering Applications of Artificial Intelligence (21), pp. 1101-1112, 2008.
[18] Pan N.H., Lee M.L., Chen K.Y.; Improved Tabu Search Algorithm Application in RCPSP, Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol I), 2009.
[19] Santos M., Tereso A. P.; On the multi-mode, multi-skill resource constrained project scheduling problem - computational results, Soft Computing in Industrial Applications, Advances in Intelligent and Soft Computing (96), pp. 239–248, 2011.
[20] Thomas P. R., Salhi S.; A Tabu Search Approach for the Resource Constrained Project Scheduling Problem, Journal of Heuristics (4), pp. 123-139, 1998.
[21] Tsai Y.W., Gemmill D. D.; Using tabu search to schedule activities of stochastic resource–constrained projects, European Journal of Operational Research (111), pp. 129–141, 1998.
[22] Valls V., Ballestin F., Quintanilla S.; A hybrid genetic algorithm for the resource–constrained project scheduling problem, European Journal of Operational Research (185), pp. 495-508, 2008.
[23] Valls V., Ballestin F., Quinanilla S.; An Evolutionary Approach to the Resource-Constrained Project Scheduling Problem, MIC2001 - 4th Metaheuristics International Conference, pp. 217–220, 2001.
[24] Verhoeven M. G. A.; Tabu search for resource-constrained scheduling, European Journal of Operational Research (106), pp. 266–276, 1998.
[25] Wang H., Lin D., Li M.Q.; A competitive genetic algorithm for Resource–Constrained Project Scheduling Problem, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, pp. 2945–2949, 2005.
[26] Zhang H., Xu H., Peng W., A Genetic Algorithm for Solving RCPSP, 2008 International Symposium on Computer Science and Computational Technology, pp. 246–249, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 159–166

Novel heuristic solutions for Multi–Skill Resource–Constrained Project Scheduling Problem

Paweł B. Myszkowski, Marek E. Skowroński, Łukasz Podlodowski
Institute of Informatics, Department of Artificial Intelligence
Faculty of Computer Science & Management, Wrocław University of Technology, Poland
Email: pawel.myszkowski@pwr.wroc.pl, m.e.skowronski@pwr.woc.pl, 168037@student.pwr.wroc.pl

***Abstract*—In this article some novel scheduling heuristics for Multi–Skill Resource–Constrained Project Scheduling Problem have been proposed and compared to state-of-the-art priority rules, based on task duration, resource salaries and precedence relations. New heuristics stand an aggregation of known methods, but are enhanced by skills domain. The goal of the paper is to investigate, whether evaluated methods can be used as robustness enhancement tools in metaheuristics, mostly evolutionary algorithms. Experiments have been performed using artificially created dataset instances, based on real–world instances. Obtained results prove that such methods stand interesting feature that can be included to more complex methods and increase their robustness.**

I. Introduction

RESOURCE-Constrained Project Scheduling Problem (RCPSP) is one of the most widely described [1], [4], [10], [11], [12] combinatorial problems in the literature. RCPSP describes the problem [4], where the set of predefined tasks and resources are given. The objective of RCPSP is to assign tasks to resources in the way to make the overall project schedule as cheaper and shorter as possible – time & cost optimization. However, there are many constraints that have to be satisfied to make the schedule feasible. Given resource cannot be assigned to more than one task in specified time. Moreover, the precedence relations between tasks have to be obeyed.

The RCPSP can be extended to Multi–Skill RCPSP (MS–RCPSP) [16], where the set of predefined skills pool is given. In MS–RCPSP each task requires some skill in specified level to be performed. Each resource disposes some subset of skills. Hence not every resource can perform every task. The ability of performance specified task has to be checked before the assignment. If specified resource is assigned to given task, even it does not cover required skills, the project schedule become infeasible.

As MS–RCPSP is a combinatorial, NP–hard problem [1], there is no optimal solution that can be found in acceptable, finite time. That is why the soft computing methods are often used, mostly metaheuristics. However, simpler heuristics are often used. Usually they obtain suboptimal solution, but their advantages are different - they are much faster than more complex metaheuristics, like Evolutionary Algorithms (EA) [8], [14] or Ant Colony Optimization [13]. What is more, heuristics provide repeatable results, while the non-deterministic character of metaheuristics can be regarded as one of the most important disadvantage.

Heuristics used to solve the RCPSP and its extensions are called *scheduling priority rules* (SPR) [10]. They are usually related to the main elements that create the problem description - tasks with their precedence relations, resources or skills. In this paper five SPR will be presented and evaluated on proposed dataset: known from the literature, like based on a task duration, a resource salary, and task successors' list size. Some novel approaches are proposed, based on an adjustment between required and available skills.

Proposed SPR could be applied to more complex methods, like metaheuristics, i.e. as a method of generating initial population in EA, to make the search more effective. Furthermore SPR can be combined with the EA–based approaches, constructing hybrid metaheuristics, where those methods could be regarded as a local search methods. The goal of this paper is to investigate whether proposed SPR are effective enough to recommend them to such applications.

The rest of the paper is organised as follows. Section II describes other approaches to solve the (MS)–RCPSP with the SPR paradigm. Section III describes proposed SPR, while Section IV provides experiments made with described approaches, based on proposed dataset. Finally, Section V presents discussion of obtained results, while Section VI provides the conclusions and some ideas of future work.

II. Related work

SPR have been widely investigated in the literature (e.g. [2], [3], [10]). In [2] some standard priority rules has been presented: minimum total slack (MINSLK), minimum late finish time (MINLFT) and minimum processing time (MINPTM). Furthermore some less popular rules have been introduced: maximum number of immediate successors (MAXNIS), maximum remaining work (MAXRWK) and maximum processing time (MAXPTM). Proposed priority rules has been tested on a randomly created dataset. Results have shown that the most sufficient priority rule for proposed dataset is MAXNIS.

In [3] most of above mentioned priority rules have been examined, like MINSLK, MINLFT, MINPTM and MAXPTM. The work–related priority rules has been changed to total work: minimum total work (MINTWRK) and maximum total work (MAXTWRK). The MINPTM and MAXPTM has been

978-83-60810-53-8/$25.00

examined, however they have been called in different ways—shortest operation first and longest operation first, but their meaning is the same to above mentioned. Authors presented some novel heuristics, based on tasks criticality or load balancing factor, which became more suitable for solving RCPSP than standard ones, as it was presented in the results section.

An approach presented in [5] proposes priority rule based on resource availability. After each task is scheduled, the resource availability is computed and it influences on further resource–to–task assignments. An opposite measure for resource availability - moving resource strength (MRS) is introduced. The bigger the MRS value is, the more busy the resource is and the probability of choosing it for further tasks is smaller. MRS is computed locally – for given time window in a project. The MRS is included for most of typical priority rules (LFT, SLK, MTS).

Some researchers also proposed combined approaches, where priority rules are included in EA [11], [17]. In those approaches priority rules are used to generate the initial solution from which the whole initial population could be generated. Other approach [6] proposes using priority rule as a general method of creating the whole schedule from the resource–to–task assignments. Resources are assigned to given tasks and then the schedule is built by adding following tasks in given order, preserving resource– and precedence–constraints. In that approach SPR based on successors definition has been investigated.

The standard notation for scheduling has been presented in [4]. It introduces the formal language for describing the scheduling problem using SPR in comparison to metaheuristics and other more complex methods. It also describes other known priority rules and methods to solve various RCPSPs with different objective functions.

Most of state-of-the-art scheduling priority rules has been tested on the PSPLIB [9] benchmark dataset. The experiments and results have been widely described in [10], [11]. An update of performed evaluation has been performed in [12] and some novel methods have been also proposed there. Moreover, priority rule heuristic methods have been compared to metaheuristics, i. e. genetic algorithms (GA) [7]. Even if GA produces comparative results, priority rule heuristics are indispensable to create initial solution.

Solutions based on priority rules for multi–mode RCPSP (MM–RCPSP) are presented in [15]. MM–RCPSP assumes that task can be performed in several modes while every mode could cause various task's performance cost or its duration. For that problem typical priority rules have been proposed: LFT, SLK, NIS, LPM and SPM. Obtained results could lead to conclusion that LFT and LST priority rules are the most suitable.

III. Scheduling priority rules

A priority rule contains information to construct a list of tasks that ranks all project tasks in a certain order to determine the priorities in which the tasks are assigned to the project schedule. Such a list is constructed in order to assign priorities to tasks based on the following project information:

- task description: information about time or cost estimates of the tasks also determines the task priorities.
- precedence constraint information: information obtained from simple critical path scheduling tools determines the task priorities.
- resource skills information: information about the project resources and skills pools they cover, determines the task priorities.

The constructed list of tasks is then used and tasks are removed one by one from the list and are put in the schedule in the heuristic scheduling process.

Constraints that have to be preserved for each priority rule:

- Resource can be assigned to specified task only if it owns the required skill in required or higher familiarity level.
- If task has predecessors, it cannot be started before the last of predecessors would be finished.
- If a *conflict* occurs during the scheduling the specified task t, it has to be removed by shifting the start time of t just after the finish time of conflict-related task. The *conflict* is defined as a situation, when given resource is assigned to more than one tasks, which are performed in overlapping periods of time.

A. Simple priority rules

We proposed three simple priority rules, where tasks are ordered to be performed by satisfying simple conditions, presented below.

1) Task duration–based: In task duration–based priority rule (TD) tasks are ordered by their duration (ascending or descending). Then resources are assigned to tasks in predefined order. If there is more than one resource that can be assigned to specified task, the cheapest one (with the smallest standard rate) is selected. If there is more than one resource with the smallest standard rate, the first one from the pool is taken to be assigned. The TD bases on MINPTM and MAXPTM state-of-the-art priority rules.

2) Resource salary–based: In resource salary–based priority rule (RS) resources are sorted by their standard rate salary and then are assigned to tasks in an order the tasks were added to the project description. If there is more than one resource with the smallest standard rate, the first one from the pool is taken to be assigned. The tasks' order is taken directly from the dataset instance.

3) Successors' list size–based: In successors' list size–based priority rule (SLS) tasks are sorted by the size of successors' list. Then resources are assigned to tasks in defined order. The SLS bases on the state-of-the-art MAXNIS priority rule. If there is more than one resource that can perform given task, the cheapest one is taken (the same like in TD). If there is still more than one resource, that can be chosen, the first one from the pool is taken.

B. Complex heuristics

Above mentioned methods base on one or maximum two sorting criteria that define the tasks order. However, those methods do not utilize sufficiently the information about the resource load balancing or skills covered by resource and skills required by tasks. Thus we propose novel SPR approaches, based on mentioned aspects.

1) Skill adjustment–based: In skill adjustment–based heuristic (SA) skills are ordered by the adjustment measure (π), compared as a difference between number of tasks requiring specified skill and number of resources owning it and then normalized to [0;1] values. For each skill, the list of tasks that require it in specified level is obtained. If the list size is bigger than one, tasks are sorted by their duration time and the first one from the ordered list is taken to be assigned by the first resource in ordered resource list by standard rate salary. If there is more than one resource than can perform specified task in the same cost, the first one from the list is used.

This priority rule has been extended by one decision to made - whether the adjustment measure should be computed only once where the following skill is taken into consideration or should be computed after each task would be assigned. In extreme, if each task had different required skill, this decision would be useless. However, such situation occurs very rarely. Hence, it was useful to check the importance of that decision in the context of obtained results. This method has been presented as a pseudo–code in the Algorithm 1.

Because the sorting order for π, task duration and resource salary can be ascending or descending and the dynamic adjustment decision had to be included, we proposed four two-state parameters: P_π regards the sorting order of π, P_d concerns task duration sorting order, P_s declares the sorting order for resource salary and the last one - P_a treats whether the dynamic adjustment feature is active or disabled. Each of those parameter can be set to A (ascending for sorting order and active for dynamic adjustment feature) or D (descending sorting order or disabled dynamic adjustment feature).

2) Resource properties – based: In resource properties – based heuristic (RP) resource that should be assigned to specified task is selected by using a complex method, where many conditions are checked. First, only valid resources are obtained – who dispose required skill in specified level. Then the subset of obtained resources is created that contains only resources, who have specified skill in the highest / lowest acceptable level. If there is more than one resource that satisfies above condition, resources are compared by their free time during the project lifetime. The free time is computed for each resource as a difference between the project duration and sum of hours that reflects all tasks assigned to specified resource. If there are still more than one resource that can be assigned, the cheaper / more expensive one is chosen, by comparing their standard rate. If there are more than one resources with the smallest / biggest resource standard rate, the overtime rate is compared. The last comparison stage regards the number of skill types that are owned by resource. Finally, if there is still more than one possibility of choosing resource, the first one from the list (regarding the order from creating the project instance) is taken. The whole RP procedure is shown as a pseudo-code in the Algorithm 2.

Algorithm 1 SA(P_π, P_d, P_s, P_a) priority rule

Require: Defined tasks (T), resources (R), relations and skills
Ensure: Feasible schedule (set of task–to–resource assignments (A)

$Q \leftarrow sortByAdjustmentMeasure(P_\pi)$
for $q \in Q$ **do**
  $T_q \leftarrow tasksWithExpectedSkill(q)$
  $T_q \leftarrow sortByDuration(P_d)$
  for t to T **do**
    for r to R **do**
      if $resourceCanDoTask(r, t)$ **then**
        $R' \leftarrow add(r)$
    $R' \leftarrow sortByStandardRateSalary(P_s)$
    $R' \leftarrow getCheapestResources()$
    $r' \leftarrow getFirst(R')$
    $a_i \leftarrow t(r')$
    $A \leftarrow add(a_i)$
  if $dynamicAdjustment(P_a)$ **then**
    $\pi \leftarrow adjustment()$
    $Q \leftarrow sortByAdjustmentMeasure(\pi)$

In analogy to SA, some parameters have to be included, to determine the sorting order of investigated elements. Hence P_{sl} concerns ordering skills by their level, P_t regards the order of resources by their free time in the project, P_{sr} sorts the resource in the order of their standard rate, P_{or} - the same like previous but regards overtime rate and P_{st} defines the sorting order for number of skills owned by resource.

Algorithm 2 RP($P_{sl}, P_t, P_{sr}, P_{or}, P_{st}$) priority rule

Require: Defined tasks (T), resources (R), relations and skills
Ensure: Feasible schedule (set of task–to–resource assignments (A)

for t to T **do**
  $skill \leftarrow getSkillName(t)$
  for r to R **do**
    if $resourceCanDoTask(r, t)$ **then**
      $R' \leftarrow add(r)$
  $R_{HSL} \leftarrow resourcesOrderedBySkillLevel(skill, P_{sl})$
  if $size(R_{HSL}) > 1$ **then**
    $R_{MFT} \leftarrow resourcesOrderedByFreeTime(P_t)$
    if $size(R_{MFT}) > 1$ **then**
      $R_{MSR} \leftarrow resourcesOrderedByStandardRate(P_{sr})$
      if $size(R_{MSR}) > 1$ **then**
        $R_{MOR} \leftarrow resourcesOrderedByOvertimeRate(P_{or})$
        if $size(R_{MOR}) > 1$ **then**
          $R_{MST} \leftarrow resourcesOrderedBySkillTypes(P_{st})$
          if $size(R_{MST}) > 1$ **then**
            $a_i \leftarrow getFirst(R_{MST})$
          else
            $a_i \leftarrow R_{MST}$
            ...
  $A \leftarrow add(a_i)$

Similarly to SA priority rule, the set of values for parameters in RP priority rule is the same: A – ascending, D – descending.

IV. Experiments and results

The goal of conducted experiments was to investigate whether proposed SPR stand a robust way of solving MS–RCPSP and thus – whether they can be used in combination with EA. To evaluate solution – the resulted project schedule – its duration time ([days]) and performance cost ([c.u][1]) were investigated.

A. Dataset

Because not only the project schedule duration, but also the cost of the schedule should be evaluated, we cannot use the standard PSPLIB benchmark dataset [9], that does not contain any information about the task performance cost. What is more, PSPLIB dataset instances do not reflect the multi–skill model. We propose the dataset containing six project instances that has been artificially created[2], in a base of real–world instances, got from the Volvo IT Department in Wroclaw. The dataset instances have been verified by experienced project manager from mentioned enterprise.

TABLE I
MS–RCPSP dataset description

Property	D1	D2	D3	D4	D5	D6
Tasks	100	100	100	200	200	200
Resources	20	10	5	40	20	10
Skills	9	9	9	9	9	9
Relations	20	26	22	133	148	129

The dataset summary has been presented in the Table I. There are two groups of created project instances: one contains 100 tasks and the second – 200 tasks. Within the group, project instances are varied by number of available resources and the precedence relationship complexity. It led to create three different project instances both with 100 and 200 tasks. The skill variety has been set up to constant 9 different skill types for each project instance. Because of the different resources and relations number, the scheduling complexity for each project was varied.

B. Set-up

For SA priority rule 4 parameters has been investigated – $2^4 = 16$ experiments with different parameters' configurations have been performed (see Tab. III). In RP summary table (see Tab. IV) records containing the same values for each project instance with different parameters' configuration have been filtered out. The number of experiments is equal to the number of parameters' configuration: $2^5 = 32$.

C. Experiments

Experiments have been divided into the following parts. Evaluation of project duration and performance cost (see Table II) for first three priority rules. Furthermore, an evaluation of the same project properties (duration and cost) has been performed for SA (Tab. III) and RP heuristics (Tab. IV).

[1]Currency units

[2]http://www.ii.pwr.wroc.pl/˜myszkowski/scheduling

All experiments have been performed on a personal computer equipped with Intel Core 2 Duo P8700 (2.53 GHz at each core), 4 GB memory RAM and Windows 7 Professional with Service Pack 1. For such configuration, SPR processing times were negligible small (1-3 [s]), thus they were not presented in details in this paper. The experimental environment was provided by the Eclipse IDE and Java programming language. We used some specific libraries[3] to process MS Project files.

We decided to highlight the best obtained results in specified tables (see Tab. II, III, IV). However, as described problem is multi-objective, we highlighted the best duration– and cost oriented optimization results for given project instance. If there were more than one best result for specified objective, that with smaller value of second objective was highlighted as the best one.

D. Results obtained for TD, SLS, RS

The experimental results for TD, SLS and RS has been presented in Table II. Taking into account the consideration of obtained results for evaluation of project duration for TD, SLS and RS, we can see that that the SLS became *a winner* in 5/6 times. Both ascending and descending mode for SLS priority rule got the best results four times. RS priority rule has never obtained the best result, what can be explained by the fact that this rule does not *care about* the duration of the project. It is focused only on the cost aspect.

As RS priority rule became the worst approach for scheduling focused on the duration aspect, it turned out to be the best way when the cost–oriented scheduling mode is required. When the ascending mode for RS priority rule was set, the best results for each project instance was obtained (6/6). However, changing the order of this priority rule from ascending to descending make it the less effective approach. It suggests that this is one of the most performance cost–sensitive SPR.

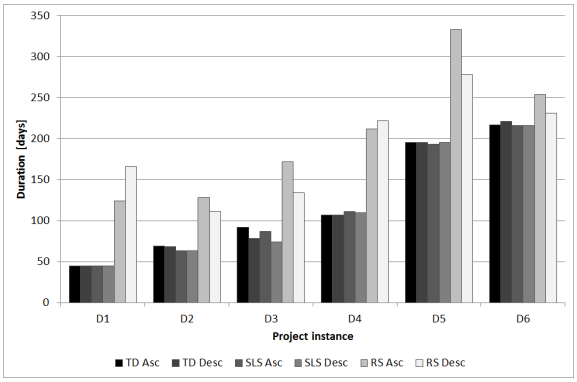


Fig. 1. Project schedule duration for TD, SLS and RS priority rules

Project duration and performance cost results for TD, SLS and RS priority rules have been also illustrated in the Fig. 1 and Fig. 2 respectively. Indicated figures gives as a clue how changing the number of resources in project could influence on the potential of scheduling optimization. The more tasks

[3]Microsoft Project Exchange - http://mpxj.sourceforge.net

TABLE II
PROJECT DURATION AND PERFORMANCE COST OBTAINED FOR TD, RS, SLS PRIORITY RULES

Method		Dataset instance											
		D1		D2		D3		D4		D5		D6	
		days	cost	days	cost	days	cost	days	cost	days	cost	days	cost
TD	Asc	45	52806	69	43262	92	40677	107	103102	195	93285	217	105686
	Desc	45	52607	68	43892	78	40862	**107**	**101319**	195	93535	221	105205
SLS	Asc	**45**	**47530**	63	43329	87	40346	111	73157	**193**	**59627**	**216**	**75033**
	Desc	45	48257	**63**	**43221**	**74**	**40286**	110	73749	195	59339	216	75141
RS	Asc	**124**	**30104**	**128**	**26323**	**172**	**30164**	**212**	**46133**	**333**	**51496**	**254**	**71986**
	Desc	166	83312	111	60384	134	52957	222	148407	278	142072	231	131272
	Avg	78	52436	84	43402	106	40882	145	90978	232	83226	226	94054

are statistically assigned to resource, the longer the project schedule could be.

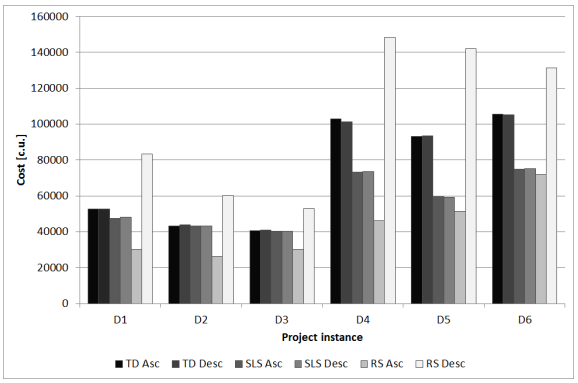


Fig. 2. Project schedule cost for TD, SLS and RS priority rules

E. Results obtained for SA

However duration results obtained for the dataset using SA heuristic are much diversified (see Tab. III), an interesting observation has been made. For 4 from 6 project instances, the best solution was found where the second parameter value, which was reflected to task duration sorting order, had been set to D. In other words - sorting tasks by their descending duration provides the best duration results. What is also interesting, for two remaining project instances the best duration results were obtained when task duration sorting order had been set to ascending, but the skill adjustment order had been set to descending.

The cost–oriented scheduling (see Tab. III) with the use of SA priority rule leads to the conclusion, that cost–related parameter (resource standard rate) is the most important. Moreover, results show that this is the only one parameter that influences to the method's robustness. If cost sorting order is set to ascending, provided results as the best, while changing the parameter to the descending value makes often results even the worst from all of investigated methods. What is also interesting, for all of performed experiments related to evaluating the cost for SA priority rule, only two resulting values were obtained.

What is also worth mentioning, duration optimization results for SA priority rule were significantly worse in comparison to those obtained for duration optimization in TD, SLS, RS and RP priority rules. While the average duration for sample project instance (D1) is equal to 78 for TD, SLS, RS (see Tab.II) or 70 (see Tab.IV) for RP priority rules, the average duration value for the same project instance in SA (see Tab.III) is about two times bigger - 149 days! On the other hand, cost aspect for D1 has been changed from average 52436 [c.u.] (see Tab.II to average 56708 [c.u.] (see Tab.III).

F. Results obtained for RP

As it is presented in the Table IV, ascending skill level criterion generally leads to get better results. Most often the best results has been placed into the parameters' configuration pool consisting skill level criterion ascending and resource availability criterion descending. It could lead to the conclusion that those two parameters influence the heuristic optimization potential the most. However, for the D6 project schedule instance, the best configuration regards ascending skill level criterion and ascending resource availability. On the other hand, for two from six project instances the best duration optimization results were obtained also for descending skill level criterion, while the resource availability criterion still remains descending.

Looking at the cost optimization results (see Tab. IV) the first impression can be made that those results are in opposition to project duration results. It proves the previous assumption that project cost and duration are in opposition. It means: reducing the performance cost leads to enlarging the project duration. The best cost optimization results were obtained when the second parameter – resource availability – has been set to ascending. An interesting fact is that steering remaining parameters does not influence on the final result. However it involves only the parameter configuration, when the skill level criterion is set descending.

Nevertheless the average values for cost optimization in SA is generally smaller than those relevant values for SA, the better optimization results were obtained for SA. The differences in average values come from smaller standard deviation of performance cost values in parameter configurations of RP. Cost obtained for RP are relatively small, but the smallest values were obtained in SA. On the other hand, invalid parameter configuration in SA could lead to increase the performance cost drastically.

TABLE III
PROJECT DURATION AND PERFORMANCE COST FOR SA PRIORITY RULE

Parameter configuration	Dataset instance											
	D1		D2		D3		D4		D5		D6	
	days	cost	days	cost	days	cost	days	cost	days	cost	days	cost
AAAA	**122**	**30104**	169	26322	189	30163	**213**	**46132**	356	51495	294	71986
AAAD	128	30104	135	26322	179	30163	237	46132	347	51495	311	71986
AADA	166	83311	126	60383	151	52957	258	148407	294	142071	258	131272
AADD	177	83311	119	60383	144	52957	234	148407	313	142071	281	131272
ADAA	**121**	**30104**	144	26322	**171**	**30163**	225	46132	349	51495	294	71986
ADAD	135	30104	140	26322	178	30163	237	46132	**329**	**51495**	292	71986
ADDA	162	83311	120	60383	138	52957	248	148407	**275**	**142071**	244	131272
ADDD	176	83311	127	60383	**135**	**52957**	245	148407	335	142071	273	131272
DAAA	127	30104	144	26322	174	30163	236	46132	342	51495	288	71986
DAAD	127	30104	144	26322	176	30163	**212**	**46132**	387	51495	292	71986
DADA	176	83311	111	60383	142	52957	241	148407	322	142071	**239**	**131272**
DADD	178	83311	111	60383	142	52957	242	148407	324	142071	247	131272
DDAA	**122**	**30104**	**131**	**26322**	177	30163	229	46132	349	51495	**259**	**71986**
DDAD	**122**	**30104**	**131**	**26322**	195	30163	229	46132	356	51495	264	71986
DDDA	166	83311	**105**	**60383**	**135**	**52957**	248	148407	309	142071	259	131272
DDDD	176	83311	**105**	**60383**	**135**	**52957**	224	148407	305	142071	259	131272
Average	149	56708	129	43353	160	41560	235	97270	331	96783	272	101629

TABLE IV
PROJECT DURATION AND PERFORMANCE COST FOR RP PRIORITY RULE

Parameter configuration	Dataset instance											
	D1		D2		D3		D4		D5		D6	
	days	cost	days	cost	days	cost	days	cost	days	cost	days	cost
AAAAA	85	37657	100	37843	118	42512	200	51905	223	58549	**216**	**97984**
AAADA	**85**	**35951**	**100**	**36957**	118	42512	201	50074	223	58447	**216**	**97984**
ADAAA	**45**	**52693**	**64**	**43094**	**108**	**39260**	110	100583	**193**	**94386**	217	104757
ADADA	**45**	**52693**	**63**	**42948**	**108**	**39260**	**108**	**101366**	199	93582	217	104757
DAAAA	100	38776	140	41511	**174**	**36299**	**285**	**49261**	**247**	**55735**	**244**	**83948**
DDAAA	50	54341	66	42576	108	42100	112	104275	195	88679	216	102737
DDADA	50	54341	66	42576	108	42100	112	103216	195	88679	216	102737
Average	70	45653	93	41126	127	40042	177	76242	215	74224	223	97357

V. RESULTS' DISCUSSION

The best obtained results for the duration and cost optimization has been compiled into the Table V. The table shows that SLS priority rule became the most effective in duration optimization for the most analysed project instances. Only for D4 instance the other method turned out to be more robust (TD). While first of proposed complex heuristic (SA) became not sufficient for duration optimization, the RP heuristic resulted well, being mostly equally efficient as SLS.

TABLE V
THE BEST OBTAINED DURATION AND COST INDICATORS.

ID	DO			CO		
	Method	D	C	Method	D	C
D1	SLS-A	45	47530	SA-ADAA	121	30140
D2	SLS-D	63	43221	SA-DDAA	131	26322
D3	SLS-D	74	40286	SA-DAAA	174	30163
D4	TD-D	107	101319	SA-DAAD	212	46132
D5	SLS-A	193	59627	SA-ADAD	329	51495
D6	SLS-A	216	75033	RS-A	254	51986

Beside the duration optimization, the project performance cost optimization has been also examined for each project instance. The cost optimization results are also presented in the Table V. Obtained results are quite interesting: SA methods became the best cost–optimization–oriented. An interesting fact is that for each best SA parameter configuration, the third criterion that reflects resource standard rate, was always set to ascending (A). Results given for RS were comparably good as those obtained by SA (see Tab. II), thus both SA and RS method became the best cost–optimization method for each project instance.

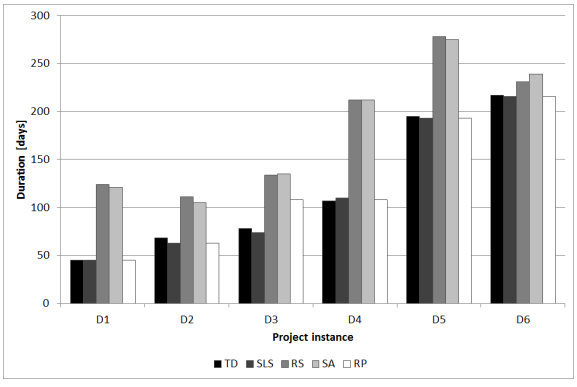


Fig. 3. Summary duration optimization results of examined methods.

The summary of duration and cost optimization for the best parameter configuration has been presented in Fig. 3 (duration optimization) and in Fig. 4 (cost optimization). In these figures

the issue of time/cost trade-off is clearly presented – reducing the project cost makes the performance cost larger and vice versa.

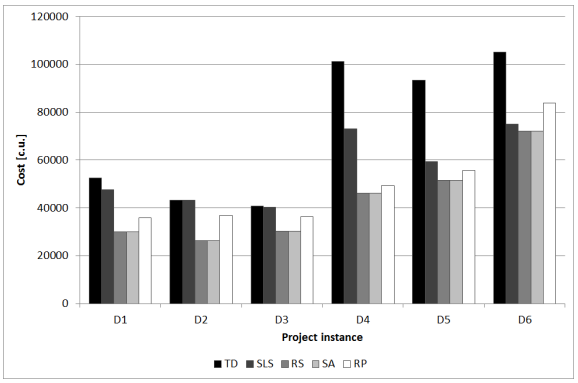


Fig. 4. Summary cost optimization results of examined methods.

Let's analyse the differences between the best duration and cost optimization. While the best project duration for D1 is 45 days, such a schedule would cost 47530 [c.u.], the cheapest project schedule could save about 17000 [c.u.] (37%), but would enlarge the project for 169%. For D2 project instance those indicators are 39% of reducing cost and enlarging the project duration by 108%. For D3 project instance 135% of project duration enlargement could cause saving 25% of budget. An interesting example is D4, where reducing the budget by the half, enlarges the duration about two times. D5 project instance turned out to be the least prone for cost optimization, where only 14% of budget could be saved, with the respect of enlarging the project duration by 70%. The last project instance – D6 – has the smallest potential in duration optimization - only 18% of duration was able to be reduced. Consequently, the potential of performance cost optimization for this instance is also reduced.

VI. Conclusions and Further Work

Before making some more general assumptions and conclusions, an important issue has to be reminded. The best optimization results, both time– or cost– oriented influence on the opposite project property. In other words, best time optimization could cause generate the schedule with unacceptable performance cost. Analogously, reducing the cost could be the reason of enormous enlargement of project duration. As it was presented in tables with detailed results.

The main drawback that can be stated for proposed SPR is that they are much less flexible in obtaining the final solution than other methods, like metaheuristics, especially those population–based. E.g. for EA the set of possible solutions is investigated during the EA *runtime*. Thus they can be compared and the *best compromise* can be found - the solution that is cheaper enough, but with acceptable cost. For SPR we do not have the possibility to find such a medium solution.

On the other hand the computing time of above mentioned methods is negligibly small, using such methods in EA could enlarge its processing times. The decision would have to be made, whether potential increase of robustness and flexibility is worth of enlarging processing time.

Obtained results provide also a conclusion that there is no need to make the priority rules too much complicated. TD, SLS and RS are operating in one or maximum two sorted lists (tasks and / or resources), while in the SA one more sorted list has to be made, based on the skills pool. It enlarges the computing complexity, but does not provides better results.

Furthermore, if cost–oriented scheduling is taken into account, the best results have been obtained using the SA heuristic. However, the results are the same like obtained for RS priority rule. It could lead to conclusion, that also cost–oriented optimization could be made with the usage of simpler priority rules.

Performed experiments showed that using SPR give possibility to get sub-optimal solution of proposed problem. Hence, in a further work proposed methods would be applied to EA as a local search (mutation operator) or as an initial population generator method.

As the fitness function in EA could be weighted by their duration and cost component, specified heuristic could be chosen depending to the weight settings of evaluation function in EA. Directed selection of local search or initial population creation method could enhance the final optimization results.

References

[1] Blazewicz J., Lenstra J.K., Rinnooy Kan A.H.G.; Scheduling subject to resource constraints: Classification and complexity, Discrete Applied Mathematics (5), pp. 11–24, 1983.
[2] Boctor F. F.; Heuristics for scheduling projects with resource restrictions and several resource–duration modes, International Journal of Production Research (31/11), pp. 2547–2558, 1993.
[3] Browning T. R., Yassine A. A.; Resource–constrained multi–project scheduling: Priority rule performance revisited, International Journal of Production and Economics (126), pp. 212–228, 2010.
[4] Brucker P., Drexl A., Mohring R., Neumann K., Pesch E.; Resource–constrained project scheduling: Notation, classification, models, and methods, European Journal of Operational Research (112), pp. 3–41, 1998.
[5] Buddhakulsomsiri J., Kim D., S.; Priority rule–based heuristic for multi-mode resource–constrained project scheduling problems with resource vacations and activity splitting, European Journal of Operational Research (178), pp. 374–390, 2007.
[6] Chen Z., Chyu C.; An Evolutionary Algorithm with Multi–Local Search for the Resource-Constrained Project Scheduling Problem, Intelligent Information Management (2), pp. 220–226, 2010.
[7] Hartmann S.; A competitive genetic algorithm for resource–constrained project scheduling, Naval Research Logistics (45), pp. 733–750, 1998.
[8] Hindi K. S., Yang H., Fleszar K.; An Evolutionary Algorithm for Resource–Constrained Project Scheduling, IEEE Transactions on evolutionary computation (6), pp. 512–518, 2002.
[9] Kolisch R., Sprecher A., PSPLIB - A project scheduling problem library, European Journal of Operational Research (96), pp. 205–216, 1996.
[10] Kolisch R.; Efficient priority rules for the resource–constrained project scheduling problem, Journal of Operations Management (14), pp. 179–192, 1996.
[11] Kolisch R., Serial and parallel resource-constrained project scheduling methods revisited: Theory and computation, European Journal of Operational Research (90), pp. 320–333, 1996.
[12] Kolisch R., Hartmann S., Experimental evaluation of state-of-the-art heuristics for the resource-constrained project scheduling problem, European Journal of Operational Research (127), pp. 394–407, 2000.
[13] Merkle D., Mittendorf M., Schmeck H.; Ant Colony Optimization for Resource–Constrained Project Scheduling, IEEE Transactions on Evolutionary Computation (6/4), pp. 333–346, 2002.

[14] Mendes J. J. M., Goncalves J. F., Resende M. G. C.; A random key based genetic algorithm for the resource constrained project scheduling problem, Computers & Operations Research (36), pp. 92–109, 2009.

[15] Lova A., Tormos P., Barber F., Multi-Mode Resource Constrained Project Scheduling: Scheduling Schemes, Priority Rules and Mode Selection Rules, Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, (10), pp. 69–86, 2006.

[16] Santos M., Tereso A. P.; On the multi-mode, multi-skill resource constrained project scheduling problem - computational results, Soft Computing in Industrial Applications, Advances in Intelligent and Soft Computing (96), pp. 239–248, 2011.

[17] Valls V., Ballestin F., Quintanilla S.; A hybrid genetic algorithm for the resource–constrained project scheduling problem, European Journal of Operational Research (185), pp. 495–508, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 167–174

Object Tracking and Video Event Recognition with Fuzzy Semantic Petri Nets

Piotr Szwed* and Mateusz Komorkiewicz*
*AGH University of Science and Technology

Email: {pszwed,komorkie}@agh.edu.pl

***Abstract*—Automated recognition of video events is an important research area in computer vision having many potential applications, e.g. intelligent video surveillance systems or video indexing engines. In this paper we describe components of an event recognition system building up a full processing chain from low-level features extraction to high-level semantic information on detected events. It is comprised of three components: object detection and tracking algorithms, a fuzzy ontology and Fuzzy Semantic Petri Nets (FSPN), a formalism that can be used to specify events and to reason on their occurrence. FSPN are Petri nets coupled with an underlying fuzzy ontology. The ontology stores assertions (facts) concerning object classification and detected relations being an abstraction of the information originating from object tracking algorithms. Fuzzy predicates querying the ontology are used in Petri net transitions guards. Places in FSPN represent scenario steps. Tokens carry information on objects participating in an event and have weights expressing likelihood of an event's step occurrence. Introduced fuzziness allow to cope with imprecise information delivered by image analysis algorithms. We describe the architecture of video event recognition system and show examples of successfully recognized events.**

***Index Terms*—video events, surveillance, fuzzy Petri Nets, fuzzy ontology**

I. Introduction

RECOGNITION of video events is an important research area in computer vision. Developed methods may have many potential applications: intelligent video surveillance, video indexing engines and various systems in which human-computer interactions are based on interpretation of video content.

Automated video event recognition comprise several tasks including detection of objects, intelligent tracking, recognition of compound events or activities and finally reasoning about occurrences of high-level events. Each of them may involve various problems to solve, e.g. how to distinguish real objects from such visual phenomena as shadows or reflexes, how to merge objects that have split into multiple segments, how to maintain objects' identities in case of occlusion, what kind of information is required to describe scene and which formalism should be used to specify events and efficiently detect them. Solutions to these problems are never perfect, each processing step may produce noisy and uncertain data, moreover a mapping between elements of semantic event specifications and low level video features often incorporate vagueness.

This work was supported from the National Centre for Research and Development (NCBiR) under Grant 0128/R/t00/2010/12 and AGH UST under Grant No. 11.11.120.612

In this paper we describe components of a video event recognition system building up a full processing chain from low-level features extraction to high-level semantic information on detected events. A conceptual layout of the systems is shown in Fig. 1. Input video sequence is analyzed with object detection and tracking algorithms. A tracking information is then represented in form of assertions in a more abstract Fuzzy Ontology layer and finally video events are detected with Fuzzy Semantic Petri Nets (FSPN), a specific class of fuzzy Petri nets, which reference terms in an ontology.

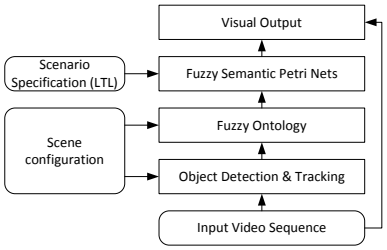


Fig. 1. Conceptual model of the event recognition system

Guards in FPNS are conjunctions of unary or binary predicates examining asserted class membership facts or object relations. Tokens in FSPN are tuples of selected scene objects participating in an event with associated weight factors. Such approach make event scenarios tolerant to classification errors, imprecise measurements and missed subevents or conditions. As transitions are fired, fuzzy weights obtained from guards evaluation are combined with token weights. In consequence, FSPN are not only capable of reasoning about scenario occurrences, but also about their likelihoods. This property is particularly important, as it allows to fine-tune the system operation by appropriately selected thresholds and filter output of less probable scenarios. Another important feature of the proposed approach is that several scenarios, starting at different time points and with different participating objects can be analyzed concurrently.

The paper is organized as follows: the next Section II reports known approaches to event specification and analysis. It is followed by Section III, in which an algorithm used to detect and track objects is briefly described; the next Section IV discusses the fuzzy ontology. FSPN are defined in Section V.

978-83-60810-53-8/$25.00

The detection system implementing the proposed approach is presented in Section VI and finally Section VII gives concluding remarks.

II. Related works

Recognition of video events has been intensively researched over last fifteen years. A large number of approaches is reported in recent surveys: [1] and [2]. Systems for video recognition usually have a layered architecture, e.g. [3], [4], in which lower level layers provide an *abstraction* of meaningful aspects of video sequences, whereas higher level layers are related to formalisms used for *event modeling* and algorithms that detect events based on formal specifications.

Probabilistic state-based methods use models comprised of states and transitions, in which transitions are attributed with probability factors learned from annotated video. During an analysis of an input video sequence a likelihood of a situation is computed. This group include methods based on neural networks [5], Hidden Markov Models, Dynamic Bayesian Networks [6] and stochastic Petri nets [7].

In grammar based methods complex activities are represented by production rules that generate strings of atomic actions. Hence, complex events can be recognized by language parsing techniques [8], [9]. In the review [2] a limitation of these methods as regards concurrent activities was indicated. The criticism seems to be founded in a case, where sequences of single actions are analyzed. However, in a more general setting, e.g. this provided by the Kripke structure [10], each string element is a set of low level events occurring in parallel and in consequence high-level concurrent events can be tracked.

Description based approaches specify events and scenarios using high level languages, either textual [11], or graphical as Situation Graph Trees [4], [12] and Petri nets [7], [13], [14]. The methods falling into this category are considered to be *semantic*, as specifications are prepared by experts, who give meaningful names to events, engaged objects, actions and conditions. Descriptions are often hierarchical: complex events can be expressed as graphs of subevents. Models can also include constraints and knowledge about scene objects, e.g. in [4] they are expressed as formulas of a Fuzzy Metric-Temporal Horn Logic. In some approaches scenarios and their ingredients: types of participating objects and relations are defined as ontologies [15], [16].

Petri Nets (PNs) are applied in the field of event detection in two modes [1]. In the first mode of *object PNs* tokens represent objects, places object states and transitions events of interest. Such approach was applied in surveillance of traffic [13] and people [17]. In the second mode of *plan PNs* places correspond to subevents building up a plan. Presence of a token in a place indicates that a particular event assigned to the place is occurring. The latter approach was applied to parking [18] and more recently people [14] surveillance.

The semantics of Petri nets proposed in this paper is closer to *plan PNs*, as tokens represent combination of objects participating in scenarios. There are, however, some salient differences. 1) In probabilistic PNs discussed in [14] in case of a conflict (e.g. two enabled transitions sharing input place with a single token) only one transition with a higher learned probability would fire, whereas in our model they both can be executed and produce two tokens with weights aggregating the weight of the input token and transition guards. This allows to reason concurrently about scenario alternatives. Moreover, a weak initial likelihood of a scenario branch can be amplified by future events. 2) In our approach all enabled transitions are executed in a single parallel step. Such behavior rather resemble reasoning with Fuzzy Cognitive Maps [19] than the most often utilized interleaving PN semantics. 3) Petri nets modeling scenarios are actually state machines. Their structure is sufficient to construct a Büchi automaton [20] representing a LTL formula.

There are a vast number of tracking algorithms and it would be hard to present all previous works in this field. However there exist two very interesting surveys which give an in-depth view of all methods [21] and [22]. Based on classification proposed in these surveys, the tracking algorithm used in this work can be assigned to a group, which detect objects by background modeling and subtraction. The final tracking is based on kernel tracking methods (region based tracking).

III. Object detection and tracking

The detection and tracking algorithm maintains a set of tracked objects O and updates it after an arrival of a new video frame. Each object has several attributes: a history of its bounding box position and size at current and N-1 previous frames, a unique object ID, information about object type (pedestrian, graffiti, group of objects etc.) and flags denoting object occlusion or information, that object can't be tracked and its position must be estimated.

For each i-th frame the set O is updated with a procedure comprised of the following steps:

A) A background is updated and foreground object segmentation is performed.
B) A set of segments S is extracted, labeled and tracked.
C) Segments S which are similar to objects from O are assigned to them.
D) All segments S which were not assigned to O are submitted to a classification process and detected objects are added to O.
E) Overlapping objects from O are merged.
F) Merged objects from O are split, if they have separate areas.
G) Positions of objects from O, that cannot be tracked by segments, are estimated.

These steps are explained in detail in the next paragraphs:

A. Foreground object segmentation

The method is based on background generation and foreground object detection described in [23]. It consists in creating a binary mask by background substraction and processing it in order to extract connected components (segments) and label them. A single segment can be a group of objects (e.g.

a crowd), single object (e.g. a pedestrian) or part of the object (e.g. a torso). Partitioning may be caused by failures of segmentation algorithm.

B. Segment tracking

Number of segments, their position and size may vary significantly from frame to frame. Segmentation errors can be caused by many factors e.g. camera noise, changes in lighting condition, shadows, occlusion by other objects etc. To make sure that only segments, which are correctly extracted will be used in the next processing stage, a simple tracking mechanism is applied that is based on checking bounding box positions of current and previous segments. If two bounding boxes overlap in two consecutive frames, segments are linked with the *history* relation describing evolution of segments in time. For further analysis only segments, which have clear history based on observation form N previous frames are used.

C. Segments to object assignment

In the next step all segments S detected in a frame i are assigned to objects from previous frame O. Let us introduce a function $h\colon O \times \mathbb{N} \to 2^S$ that defines a mapping between an object o at a frame i and a set of segments. The Algorithm 1 takes at input the sets of objects O and segments S, updates the function h and computes the set of assigned segments S_u.

Algorithm 1

procedure ASSIGN(O, S, i, S_u, h)
  for all $o \in O$ **do**
    for all $s \in S$ **do**
      if $d(o, s) \geq \epsilon$ **then**
        Add segment to an object:
        $h(o, i) \leftarrow h(o, i) \cup \{s\}$
        Add s to the set of used segment S_u:
        $S_u \leftarrow S_u \cup \{s\}$
      end if
    end for
  end for
end procedure

The function d calculates a normalized to $[0, 1]$ similarity between an object o and a segment s considering overlapping of segments in $h(o, i-1)$ and s (bounding boxes or image masks). The threshold ϵ is a small constant, e.g. 0.1.

After executing the procedure new positions and sizes of all objects $O_u = \{o \in O\colon h(o, i) \neq \emptyset\}$ are computed based on position and size of segments assigned to them.

D. Object detection and classification

All segments, which were not assigned to any object ($S_{NA} = S \setminus S_u$) are further analyzed by a set of classifiers. Simple geometrical rules are applied (perspective is compensated) e.g. an adult person should be higher than 160 cm and wider than 40 cm. If a segment with desired properties is detected, a new object is created o_{new} with a unique ID and added to object list $O \leftarrow O \cup \{o_{new}\}$. If an object is divided into several segments, a correct classification is not possible. In this case the system will not detect it on the current frame. As the video sequence is analyzed, it is very likely that in next few frames it will appear not split and the algorithm will be able to detect it. The benefit of this approach is that only very reliable objects are detected, what leads to smaller false detection rate.

E. Object merging

In the next processing step, the algorithm handles cases, when two previously tracked objects merge, e.g. if two pedestrians approach and finally their silhouettes overlap. As it is then impossible to track objects based on the information on segments position and history, the algorithm sets a special flag in the overlapping objects descriptors. From that time their positions are not computed based on the position of segments belonging to them. Instead, a prediction mechanism is used, which is based on information about previous sizes, movement direction and speed. Also a new temporal object is created (so called merged object). Its size and position is updated based on segments which previously belonged to merged objects. Thanks to this, the system is able not only to estimate the object position, but also to keep track of real area occupied by the estimated objects.

F. Object splitting

In the next step it is checked, if previously merged objects are split. Such situation occurs, e.g. when two previously joined pedestrians move away far enough to allow for total separation of segments belonging to each object. In this case, all segments are checked to test, if they resemble an object within a merged group. If appropriate correspondences can be found, the unique IDs are restored based on estimated positions of the original objects. Another scenario is also possible. A person may leave a luggage and start to move away (possible bomb planting scenario). In such case it is possible that an object, which has only one ID starts to split. To cover such situations, a maximum object size is checked and, if it exceeds the maximum allowed object size, the tracking object is removed and classification is rerun for all segments belonging to this object. The largest detected object is given the old ID, all other detected objects are given new unique IDs.

G. Position estimation

In the last step, for all objects, which are marked as lost or impossible to track, i.e. $h(o, i) = \emptyset$, a history of their positions on previous frames is analyzed to compute a mean velocities. Object's velocity is then used to estimate a new position. The object size is not estimated, the last known size is used instead. This estimation method is working well in most cases. It can also cause wrong results, if objects change speed or movement direction during the estimation. To overcome this types of errors, a guard mechanism was introduced. It is based on counting the number of pixels belonging to foreground objects within an estimated bounding box. If it drops below a fixed

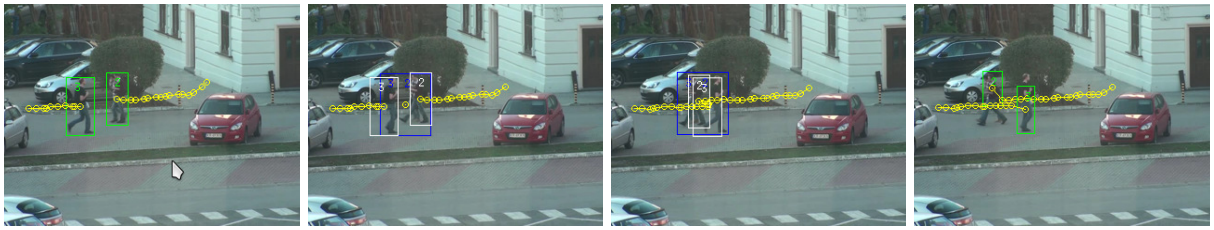


Fig. 2. Sample tracking sequence, green boxes - tracked pedestrians, white boxes - estimated pedestrians positions, blue boxes - group position, yellow circles - trajectories

threshold, it is a signal that the object position is not estimated correctly. In such case the estimated object is removed.

A sample tracking sequence of two passing pedestrians is presented in Figure 2.

IV. Fuzzy ontology

The fuzzy ontology constitute an intermediate layer between information on tracked objects and fuzzy Petri nets. Whereas objects within the tracking model are described with numeric values, like size, distance or speed, the ontology provide a kind of linguistic abstractions, e.g. *a small object*, *objects are close* or *a person is walking*.

There are several benefits of this approach:

- Scenario specifications can be prepared in a more general and meaningful manner, they can be decoupled from the code and implementation details can be hidden .
- It is much easier to customize a recognition system to specific needs and conditions, because the translation between numeric values and linguistic terms is accomplished in isolated and easy to identify functions
- Facts concerning classifications of objects and detected relations are materialized, hence they can be evaluated only once, what generally increases performance.

Ontologies are often described as unions of two layers: terminological (*TBox*) and assertional (*ABox*). The *TBox* defines concepts and types of relation including: taxonomic relations between concepts, *object properties* and *datatype properties*. The *ABox*, in turn, gathers facts about individuals and existent relations. In Description Logic, being a counterpart of ontology languages, concepts and relations can be expressed by means of unary and binary predicates, e.g.: $Person(x)$ – x is a member of the class $Person$, $isWalking(x)$ – a boolean datatype property $isWalking$ of an individual x or $isClose(x, y)$ – an object property between two individuals x and y.

For *fuzzy ontologies* and corresponding Fuzzy Description Logics the ontology relations are extended by adding weights being real numbers from $[0, 1]$. They can be used to express uncertainty, e.g. with respect to class membership or relation occurrence. Formalizations of fuzzy ontology languages including fuzzy classes, roles (object properties) and datatype properties can be found in [24] and [25].

In the case of a fuzzy ontology used with FSPN, its TBox is a stable part, whereas the ABox is updated for each frame. A crucial element of the described approach is that relations in the ABox are practically never fully evaluated. Only their subset that is requested from FSPN is calculated by making calls to plugged in functions (function objects in object-oriented implementation) called *evaluators*. They examine the tracking model and calculate fuzzy weights of predicates. In opposition to approach proposed in [25] evaluators are external entities beyond the ontology. In many cases they have a form of membership functions described by line segments, as in Fig. 3, but they can be also based on other features, as Jaccard metrics applied to object areas (Fig. 4). In this case a bounding box of a detected object is divided into a $n \times m$ grid and each cell is assigned with a probability density p_{ij}. The weight returned by the evaluator is calculated according to the formula: $w = \frac{1}{Z} \sum_{i=1}^{n} \sum_{j=1}^{m} h_{ij} p_{ij}$, where Z is a normalizing constant $Z = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}$ and $h_{ij} = 1$ if a cell (i, j) intersects with an object or 0 in other case. The selection of probability distribution is arbitrary and depends on the type of interaction, e.g. the grid in Fig. 4a was used to calculate intersection values for vertical objects, e.g. walls, whereas the grid in Fig. 4b for horizontal ones, e.g. forbidden zones on a floor.

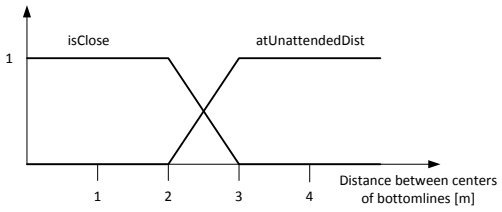
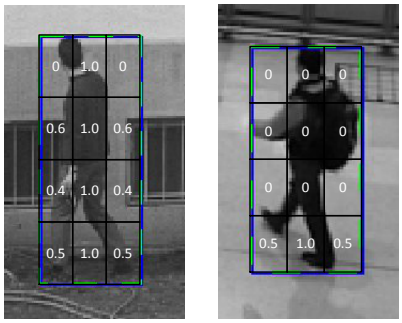


Fig. 3. Membership functions used by *evaluators*

V. Fuzzy Semantic Petri Nets

In this section we define Fuzzy Semantic Petri Nets and describe their behavior. FSPNs are placed at the top of video event recognition stack (Fig. 1 and are responsible for interpretation of low-level events and conditions represented as assertions in a sequence of ABoxes of the coupled fuzzy ontology.

It should be noted, that the presented semantics of FSPN is dedicated to a particular case of state machines, i.e. Petri nets, in which transitions link single places. Such restrictions



a. b.

Fig. 4. Evaluators based on Jaccard metrics.

stems from fact that we use Linear Temporal Logic (LTL) [26], [27] to specify video events. LTL specifications are then transformed to corresponding FSPN structures following the rules for translating them to Büchi automata [20].

Relations between FSPN and LTL are even deeper, an input sequence of ABoxes analyzed by a FSPN can be considered a specific kind of Kripke structure [10], which is defined as a sequence of states $s_0, s_1, s_2, \ldots, s_n, \ldots$ and a function that assigns sets of true propositions P_i to states s_i. In our case a state s_i corresponds to an i-th video frame and a set of propositions P_i to a set of assertions in an i-th ABox of the fuzzy ontology. Hence, detection of a video event can be considered a checking if a model (a sequence of ABoxes) satisfies an LTL formula that is translated to a FSPN.

In the presented approach we generalize and relax the acceptance requirements:

- Instead of Büchi automata, fuzzy Petri nets are used as a tool for scenario analysis. This allows to process concurrently scenarios, in which participate various combinations of objects.
- To manage uncertainty and inexactness of input data, fuzzy predicates returning values from $[0, 1]$ are used. These values are then combined with weights of tokens flowing through a net. Tokens, in turn, represent scenario occurrences. This enables monitoring scenario steps and reasoning about their likelihood.
- Sequences of accepted states strictly defined with LTL formulas can be interleaved with states not satisfying the specified conditions. In such case, the weight determining scenario satisfaction gradually decrease and, after passing a certain threshold, the scenario is rejected by token removal.

A. Definition of Fuzzy Semantic Petri Nets

Formal definition of Fuzzy Semantic Petri Nets is comprised of three concepts: Petri net structure, binding and fuzzy marking. We start with some auxiliary definitions. Unary predicate is defined as a pair (n, v_s) where, n is a predicate name and v_s is a variable name referring to a *subject* of the predicate. Binary predicate is a triple (n, v_s, v_o); the variable v_o is a predicate object. Set of all unary and binary predicates is denoted by $Preds$. By $Vars(p)$ we denote a set of variables appearing in the predicate p. Analogously, for a set $C \subseteq Preds$ we define $Vars(C)$, as $\bigcup_{p \in C} Vars(p)$.

Definition 1 (Petri net structure)**.** Petri net structure PN is a tuple $(P, T, F, Preds, G, L, H)$, where P is a set of places, T is a set of transitions, P and T are satisfying $P \cap T = \emptyset$ and $P \cup T \neq \emptyset$. $F \subseteq P \times T \cup T \times P$ is a set of arcs (flow relation), and $Preds$ is a set of unary and binary predicates. $G\colon T \to 2^{Preds}$ is a guard function that assigns sets of predicates to transitions. $L\colon P \to \mathbb{N} \cup \{0\}$ is a function assigning lower bound to a place; this value defines how long a token should stay in a place to be allowed to leave it. $H\colon P \to \mathbb{N} \cup \{\omega\}$ assigns upper bound to a place. The symbol ω represents infinity.

Following [28] the set of input places for a transition $t \in T$ is denoted as $\bullet t = \{p \in P\colon (p, t) \in F\}$ and the set of output places as $t\bullet = \{p \in P\colon (t, p) \in F\}$

Definition 2 (Binding)**.** Let V be set of variables and I a set of objects. *Binding* b is defined as a partial function from V to I. A variable v *is bound* for a binding b, iff $v \in \operatorname{dom} b$. A set of all bindings is denoted by B.

Let $p \in Preds$ a predicate and $b \in B$ be a binding. Predicate value for a binding $val\colon Preds \times B \to [0, 1]$ is a function that assigns value from the interval $[0, 1]$ to a pair (p, b), $p \in Preds$ and $b \in B$. If $Vars(p) \setminus \mathbf{dom}\, b \neq \emptyset$, then $val(p, b) = 0$.

Definition 3 (Fuzzy marking)**.** A set of fuzzy tokens FT is defined as $FT = B \times \mathbb{R} \times (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$. Components of a token tuple $(b, w, c, \tau) \in FT$ are the following: $b \in B$ denotes a binding, $w \in [0, 1]$ is a fuzzy weight, $c \geq 0$ is a counter storing information, how long a token rests in a place and τ is a time stamp. *Fuzzy marking* for a Petri net $PN = (P, T, F, Preds, G)$ is defined as a function that assigns sets of fuzzy tokens to places $FM : P \to 2^{FT}$.

B. Execution

The behavior of FPNs defined in previous section differs from the standard semantics for Petri nets, as they are not intended to focus on concurrency and conflicts, but to perform a kind of fuzzy reasoning and classification of sequences of events.

Single i-th step of execution of a fuzzy Petri Net is comprised of three basic stages:

1) *Firing enabled non-initial transitions and generating new tokens.* During this stage each token-transition pair (t, m), where $t \in T$ and $m = (b, w, c, \tau) \in FM(\bullet t)$ is analyzed. If a guard $G(t)$ references unbound variables, i.e. $Vars(G(t)) \setminus \mathbf{dom}\, b \neq \emptyset$, an attempt is made to create a new binding b' by grounding free variables with ontology individuals; in other case $b' = b$. Then, a weight of the guard is calculated: $w_g = \min\{val(p, b)\colon p \in G\}$ and aggregated with the old token weight: $w' = a\, w_g + (1 - a)w$. If w' is greater than a certain threshold (in experiments 0.2 value was used), a new token $m' = (b', w', c', t')$ is created and put

into the transition's output place $t\bullet$. The current iteration number i is assigned as token timestamp $\tau' = i$ and, if the transition t is a self-loop, the counter is updated: $c' = c + 1$. It should be mentioned, that only self-loop transitions can fire if token counter c does not belong to the interval $[L(\ bullett), H(\bullet t)]$ (see Definition 1).

2) *Removing old tokens.* A transition occurrence performed in the previous stage can be regarded as a triple (m, t, m'), where m is an input token, m' an output token and $t \in T$ a transition. Let C denote a set of such triples, and $w(m)$ a weight of a token. For each input token m a sum of weights of output tokens m' is calculated and subtracted from its weight: $w_{new}(m) = w(m) - \sum_{(m,t,m')\in C} w(m')$. If the value falls below a certain threshold, the token m is removed. Also in this step multiple tokens having the same binding and assigned to the same place are aggregated.
3) *Firing initial transitions.* Finally, new tokens are introduced into the net, by firing initial transitions (i.e. satisfying $\bullet t = \emptyset$). For each initial transition variables appearing in its guard are bound to objects, then the guard value is calculated and used as a weight of new tokens. To avoid analyzing scenarios with low likelihoods, a threshold preventing from creating tokens with small weights is defined. The mechanism is also protected against introducing tokens with a binding already present in the net.

C. Video event specification

We start preparing a specification of a video event by outlining a general scenario in form of Temporal Logic formula, then events appearing in the scenario are refined into conjunctions of low-level events or conditions expressed as predicates. The resulting LTL specification is used in two ways: (1) it is translated into FSPN and (2) predicates are included into TBox of the fuzzy ontology.

To give an example: a high-level video event, in which a person violates a forbidden zone can be expressed in LTL as a sequence of three medium-level events: $init \Rightarrow \Diamond move \Rightarrow \Box violate$, where $init$ defines conditions to start recognition, $move$ denotes a situation, when a person is moving towards a zone and $violate$ a situation, when the person enters the zone. During the refinement step the scenario is transformed into formula (1), which is further translated into FSPN shown in Fig. 5.

$$\begin{aligned} &Person(x) \wedge isWalking(x) \wedge atBorder(x) \wedge Zone(y) \\ &\Rightarrow \Diamond(isWalking(x) \wedge movesTowardsZone(x,y)_{\{8,\infty\}} \\ &\Rightarrow \Box(bottomInZone(x,y))_{\{4,\infty\}} \end{aligned} \quad (1)$$

Fig. 6 shows a FSPN defining a complex event, during which a person leaves unattended luggage. Its scenario (in a narrative form) with accompanied video material was published as a benchmark for PETS 2006 workshop [29]. The event is defined as a sequence of four simple steps: *init* – a still person appears, *separate* – the person puts a luggage on the floor and remains close to it, *leave* – distance between the person and luggage grows above a certain threshold (equal to 3 meters in PETS specification) and finally *remain* – the person disappears and the luggage remains alone.

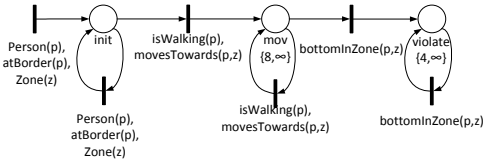


Fig. 5. Fuzzy Semantic Petri Net representing a scenario, in which a person violates a forbidden zone

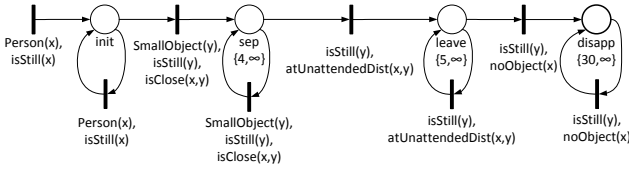


Fig. 6. Fuzzy Semantic Petri Net representing luggage left scenario

A more complex FSPN is presented in Fig. 7. It defines an event of graffiti painting on a wall decomposed into medium level events: a person moves towards a wall, then appears in front of the wall, optionally: a person is widening (what may indicate painting graffiti), then a new object emerges on a wall (but not inside a window) and remains still.

VI. Detection system

In this section we describe a prototype system allowing to test recognition of events based on specifications in FSPN. The system takes at input a video sequence with an XML file defining tracking information. For each frame a list of segments and identified objects is provided. The data originate from tracking algorithms described in Section III. The architecture of the prototype scenario detection system is presented in Fig. 8. Main components are: the *Fuzzy ontology*, a set of *Evaluators* and the *Fuzzy Semantic Petri Net* execution engine. The system is also equipped with GUI providing visual output shown in Fig. 9.

The control flow during a single iteration was marked in Fig. 8 with numbers in circles.

1) After a new frame appears, asserted relations between objects are removed from the ontology, then newly identified objects are added as individuals.
2) In the next step all enabled transitions in concurrently analyzed Petri nets are fired. Preparation of transitions requires calculations of guards and in some cases extensions of bindings.
3) In order to obtain weights of predicates appearing in guards, appropriate queries are made to ontology. If a weight for a predicate was evaluated earlier, it is immediately returned.
4) In other case an *evaluator* assigned to the predicate is called, and returned value is asserted in the ontology as a weight of corresponding fuzzy relation.

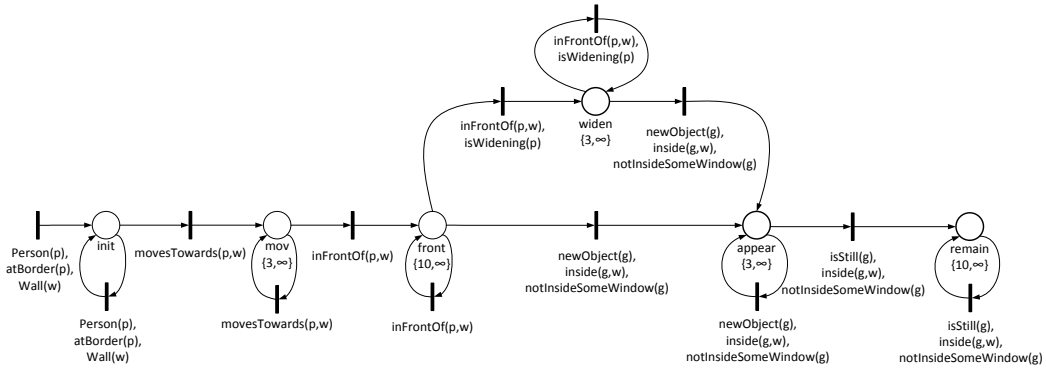


Fig. 7. FSPN representing graffiti painting scenario

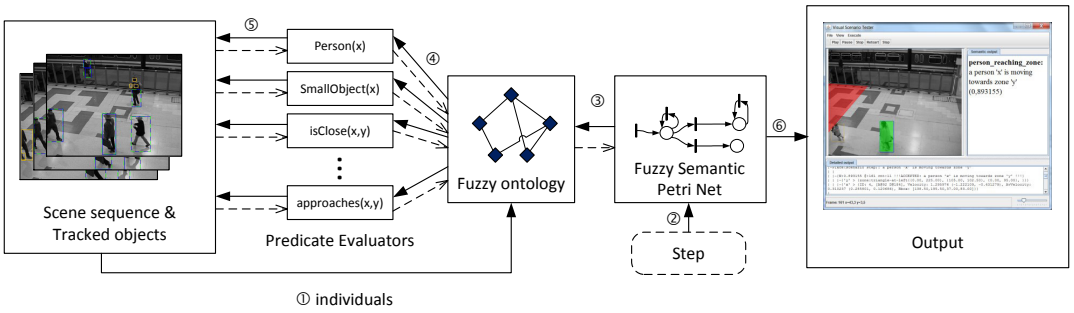


Fig. 8. Architecture of the detection system

5) Evaluators examine the tracking information. As a tracking history covering a number of past frames is kept, evaluators are capable of returning temporal properties, e.g. $newObject(x)$ - an object is considered *new* if it has recently appeared.
6) After the net state is updated, a reached marking is analyzed. If a token stays in a selected place long enough (observed places are defined in FSPN specification) its presence is reported as an important scenario step or a final stage.

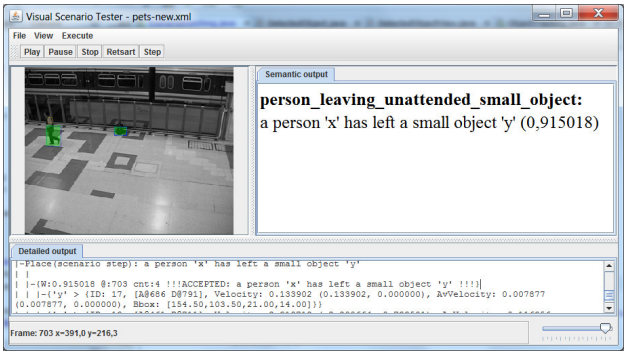


Fig. 9. Visual scenario tester. The displayed video frame and messages correspond to the place *leave* ($isStill(y) \wedge atUnattendedDist(x,y)$) of the FSPN specifying luggage left scenario.

The software is entirely written in Java. Its performance is quite good: for three concurrently analyzed scenarios and a scene with a few tracked objects, a single reasoning iteration, during which the ontology is updated, evaluators are called and multiple transitions in Petri nets are fired, is executed within 0.1ms to 1.6ms (average 0.45 ms) on a Pentium i7 2.2 GHz machine.

The system was successfully tested to recognize a number of events, including these specified in Fig. 5, Fig. 6 and Fig. 7, returning in each case high likelihood value close to 1.0.

Fig. 10 presents visual output corresponding to the steps of the graffiti painting scenario specified by FSPN in Fig. 7. Filled bounding boxes mark objects included into the binding of tokens that reached places (scenario steps), for which semantic messages are displayed.

VII. Conclusions

In this paper we describe components of video events recognition system building a full processing chain: from objects detection and tracking, through transforming tracking information into more abstract representation of Fuzzy Ontology, to reasoning on events occurrences with Fuzzy Semantic Petri Nets.

An advantage of FSPN is their capability of detecting concurrently occurring events, in which participate various combinations of objects, analyze scenario alternatives and their likelihoods. Petri nets state (marking) gives general overview of the situation, of *what's going on*. A presence of a token in a

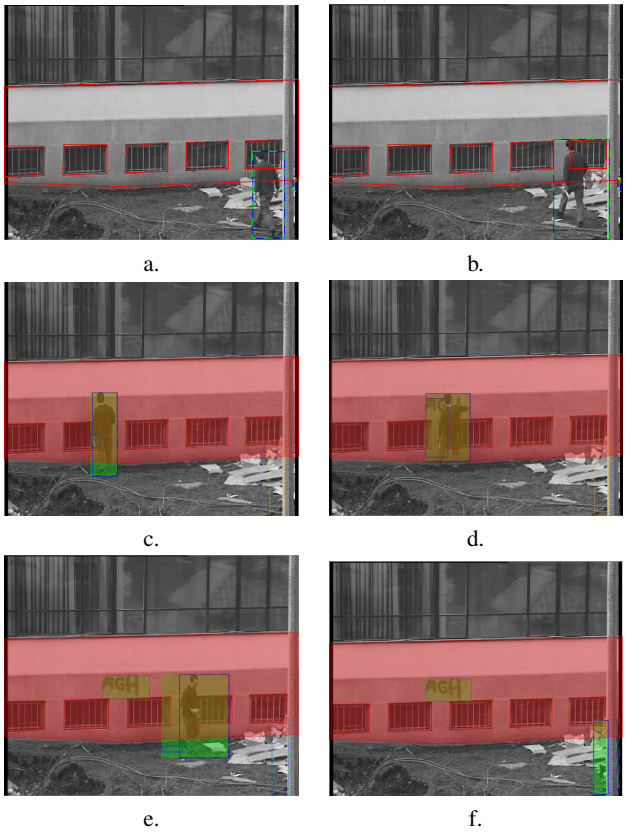


Fig. 10. Recognized steps of a graffiti painting event: a – *init*, b – *move*, c – *front*, d – *widen*, e – *appear*, f – *remain* (scenario completion)

place can be reported as *semantic output*, e.g. to a surveillance system operator.

The proposed scenario detection system is a generic framework, that can be adapted to specific needs by: 1) defining an ontology including classes of objects and relations of interest; 2) implementing evaluators, i.e. functions responsible for calculating values of fuzzy predicates, and plugging them into the framework; 3) configuring scene objects (their types must be defined in the ontology) 4) writing a scenario using in formulas entities (classes and relations) from the ontology.

REFERENCES

[1] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 39, no. 5, pp. 489–504, Sept. 2009.

[2] J. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys (CSUR)*, vol. 43, no. 3, p. 16, 2011.

[3] F. Brémond, M. Thonnat, and M. Zúniga, "Video-understanding framework for automatic behavior recognition," *Behavior Research Methods*, vol. 38, no. 3, pp. 416–426, 2006.

[4] D. Munch, J. Jsselmuiden, M. Arens, and R. Stiefelhagen, "High-level situation recognition using fuzzy metric temporal logic, case studies in surveillance and smart environments," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov. 2011, pp. 882–889.

[5] M. Barnard, J.-M. Odobez, and S. Bengio, "Multi-modal audio-visual event recognition for football analysis," in *Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on*, Sept. 2003, pp. 469–478.

[6] J. Aggarwal and S. Park, "Human motion: modeling and recognition of actions and interactions," in *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, Sept. 2004, pp. 640–647.

[7] G. Lavee, M. Rudzsky, E. Rivlin, and A. Borzin, "Video event modeling and recognition in generalized stochastic Petri nets," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 1, pp. 102–118, Jan. 2010.

[8] S.-W. Joo and R. Chellappa, "Attribute grammar-based event recognition and anomaly detection," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, June 2006, pp. 107–107.

[9] G. Guerra-Filho and Y. Aloimonos, "A language for human action," *Computer*, vol. 40, no. 5, pp. 42–51, May 2007.

[10] S. Kripke, "Semantical considerations on modal logic," *Acta philosophica fennica*, vol. 16, no. 1963, pp. 83–94, 1963.

[11] V.-T. Vu, F. Bremond, and M. Thonnat, "Automatic video interpretation: A novel algorithm for temporal scenario recognition," in *International Joint Conference on Artificial Intelligence*, vol. 18. Lawrence Erlbaum Associates Ltd, 2003, pp. 1295–1302.

[12] H.-H. Nagel, "Steps toward a cognitive vision system," *AI Magazine*, vol. 25, no. 2, p. 31, 2004.

[13] N. Ghanem, D. DeMenthon, D. Doermann, and L. Davis, "Representation and recognition of events in surveillance video using Petri nets," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, June 2004, pp. 112–112.

[14] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V. S. Subrahmanian, P. Turaga, and O. Udrea, "A constrained probabilistic Petri net framework for human activity detection in video," *Multimedia, IEEE Transactions on*, vol. 10, no. 8, pp. 1429–1443, Dec. 2008.

[15] F. Bremond, N. Maillot, M. Thonnat, V.-T. Vu *et al.*, "Ontologies for video events. Research report number 51895," INRIA Sophia-Antipolis, Tech. Rep., 2004.

[16] U. Akdemir, P. Turaga, and R. Chellappa, "An ontology based approach for activity recognition from video," in *Proceedings of the 16th ACM international conference on Multimedia*. ACM, 2008, pp. 709–712.

[17] A. Borzin, E. Rivlin, and M. Rudzsky, "Surveillance event interpretation using generalized stochastic Petri nets," in *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS'07. Eighth International Workshop on*. IEEE, 2007, pp. 4–4.

[18] C. Castel, L. Chaudron, and C. Tessier, "What is going on? a high level interpretation of sequences of images," 1996.

[19] J. Aguilar, "A Survey about Fuzzy Cognitive Maps Papers (Invited Paper)," *International Journal*, vol. 3, no. 2, pp. 27–33, 2005.

[20] J. R. Büchi, "On a Decision Method in Restricted Second-Order Arithmetic," in *International Congress on Logic, Methodology, and Philosophy of Science*. Stanford University Press, 1962, pp. 1–11.

[21] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, Dec. 2006.

[22] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *The Visual Computer*, pp. 1–27, 2012. [Online]. Available: http://dx.doi.org/10.1007/s00371-012-0752-6

[23] T. Kryjak and M. Gorgoń, "Real-time implementation of moving object detection in video surveillance systems using FPGA," *Computer Science Journal, Wydawnictwa AGH*, vol. 12, pp. 149–162, 2011.

[24] S. Calegari and D. Ciucci, "Fuzzy ontology, Fuzzy Description Logics and fuzzy-OWL," *Applications of Fuzzy Sets Theory*, vol. 4578/2007, no. D, pp. 118–126, 2007.

[25] T. Lukasiewicz and U. Straccia, "Managing uncertainty and vagueness in description logics for the Semantic Web," *Web Semantics Science Services and Agents on the World Wide Web*, vol. 6, no. 4, pp. 291–308, 2008.

[26] Z. Manna and A. Pnueli, "Temporal logic," in *The Temporal Logic of Reactive and Concurrent Systems*. Springer New York, 1992, pp. 179–273.

[27] M. Fisher, *An Introduction to Practical Formal Methods Using Temporal Logic*. Wiley, 2011.

[28] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

[29] PETS 2006, "Ninth IEEE international workshop on performance evaluation of tracking and surveillance - benchmark data," 2006, [Online; accessed 04-March-2013]. [Online]. Available: http://www.cvg.rdg.ac.uk/PETS2006/data.html

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 175–178

Collective Belief Revision in Linear Algebra

Satoshi Tojo
School of Information Science
JAIST
Asahidai 1-1, Nomi, Ishikawa 923–1292, Japan
Email: tojo@jaist.ac.jp

***Abstract*—Although the logic of belief update has mainly concerned a belief state of one agent thus far, the real world settings require us to implement simultaneous belief changes. Here, however, we need to manage so many indices: agent names, time stamps, and the difference of information. In this paper, we introduce the notation of vectors and matrices for the simultaneous informing action. By this, we show that a matrix can represent a public announcement and/or a consecutive message passing, with the time of the change of belief states properly. A collective belief state multiplied by a communication matrix, including matrices of accessibility in Kripke semantics, becomes a hypercuboid.**

I. Who Knows What at Which Time?

THE authors have tackled legal reasoning system thus far [12], [13], in which one of the main issues is to properly represent *who knows what at which time*. In Fig. 1, we show the informing actions of three agents: judge, lawyer, and the suspected. At the initial stage, the judge sentenced the suspected to be guilty, and at the same time the lawyer pleaded innocent to the judge with a new witness. The judge changed his mind and he was going to sentence the defendant to be innocent, but at the same time the maladjusted defendant insulted the judge in the court, and which badly impressed jury members ...

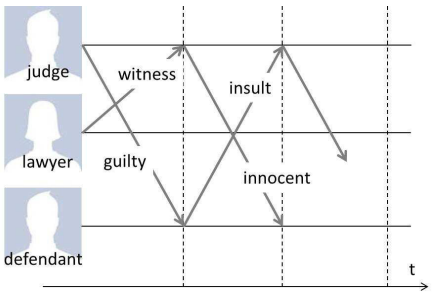


Fig. 1. Informing Agents on Time-axis

Multi-agents communication includes such many factors as agent ID, many messages, and time. In this paper, we introduce vectors and matrices to represent those agents' informing actions collectively, to clarify the complicated relations of those many indices.

II. Preliminaries

First we show the simplest prescription for an informing action; in a similar way to FIPA/ACL [3], we place the precondition in the upper deck and the result in the lower deck.[1]

$$\frac{B_i\varphi}{B_j\varphi}[c_{ij}^{\varphi}] \tag{1}$$

That is, when agent i believes φ ($B_i\varphi$) and there is a communication from i to j (c_{ij}^{φ}), agent j comes to believe φ ($B_j\varphi$). At this stage, there are several issues we need to consider:

- The problem of belief revision; the recipient of information may not believe what was informed of, or he/she may need to change some of what he/she has believed.
- The resultant state should include nested belief states, i.e., both of the sender and the recipient should recognize that the information is shared between them as $B_iB_j\varphi$ and $B_jB_i\varphi$. In addition, if those agents are quite introspective, each of them also possesses $B_iB_i\varphi$ and $\neg B_i\neg\varphi$.[2]

Incidentally, a Kripke frame is such $\mathfrak{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ that $\mathcal{W}$ is a set of possible worlds, $\mathcal{R}$ is the accessibility of belief modal operator B , and $\mathcal{V}$ gives valuation to each φ. Dynamic Epistemic Logic (DEL) [2] presents a change of belief state, restricting accessibility to possible worlds, as:

$$\mathfrak{M}, w \models [\varphi!]\psi \iff \mathfrak{M}^{\varphi!}, w \models \psi. \tag{2}$$

where $R^{\varphi!}$ in $\mathfrak{M}^{\varphi!}$ is:

$$R^{\varphi!}(w) = R(w) \cap \{w' \in W \mid \mathfrak{M}, w' \models \varphi\}.$$

On the contrary, Public Announcement Logic (PAL) [11] masks those contradicting possible worlds, as follows.

$$\mathfrak{M}, w \models [!\varphi]\psi \iff \mathfrak{M}^{[\![\varphi]\!]}, w \models \psi.$$

where in $\mathfrak{M}^{[\![\varphi]\!]}$ let $W_{\uparrow\varphi}$ be the worlds in which φ holds and

$$\mathfrak{M}, w \models \varphi \rightarrow w \in V_{W_{\uparrow\varphi}}(\psi).$$

Note that the significance of these methods is to make formula $[!\varphi]B\varphi$ valid since φ holds in all the accessible possible worlds.

Among various trials to represent agent communication formally [1], [9], [10], Yamada [14] showed a command from i to j as $[!_{(i,j)}\chi]$ where χ is the contents of the command. Kobayashi and Tojo [6], [7] generalized this notion to an informing action, representing the dynamic operator as $[inf_{ij}^{\varphi}]$.

[1] In this paper, we disregard U (uncertain) and U_{if} (uncertain if) operators for simplicity.

[2] In general, belief modality is often implemented with $KD45$, including $B_i\varphi \rightarrow B_iB_i\varphi$ (Axiom 4) and $B_i\varphi \rightarrow \neg B_i\neg\varphi$ (Axiom D), while knowledge modality requires $KT5$ including $K_i\varphi \rightarrow \varphi$ (Axiom T).

978-83-60810-53-8/$25.00

As for linear algebraic representation of belief, Fusaoka [4] used a matrix to show probability of knowledge source. Also, as we have mentioned, Liau [8] represented the network of accessibility in matrix. We combined these works, though we avoid probabilistic point of view and restricted the elements to truth values.

III. Informing Action

The belief modality B_i^t represents the belief state of agent i at time t. As to information φ, the belief states of multiple agents are written collectively in a vector as follows.

$$\begin{pmatrix} B_1^t\varphi \\ B_2^t\varphi \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix}$$

where $1/0$ are the truth values. Therefore, our specification of belief revision of agent j, with regard to (1), becomes:

$$B_j^{t+1}\varphi = B_j^t\varphi \vee (c_{ij}^{\varphi} \wedge B_i^t\varphi). \tag{3}$$

Here, c_{ij}^{φ} represents the informing action from j to i.

We define the addition and the multiplication of linear algebra as the logical 'or' and the logical 'and', respectively, as follows.

$\wedge$	1	0
1	1	0
0	0	0

$\vee$	1	0
1	1	1
0	1	0

Then, we can generalize (3) to be:

$$B_j^{t+1}\varphi = \bigvee_i (c_{ij}^{\varphi} \wedge B_i^t\varphi), \tag{4}$$

and the dynamic operator becomes such an $n \times n$ matrix that

$$(c_{ij}^{\varphi}) = \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

where n is the number of agents. Its (i, j)-element represents the truth value of c_{ij}^{φ}, that is the existence of informing action from j to i. We assume that diagonal elements c_{ii} $(i = 1, \cdots, n)$ are always true to maintain his/her original knowledge as (3). Now, the collective belief revision becomes:

$$\begin{pmatrix} B_1^{t+1}\varphi \\ B_2^{t+1}\varphi \\ \vdots \end{pmatrix} = (c_{ij}^{\varphi}) \begin{pmatrix} B_1^t\varphi \\ B_2^t\varphi \\ \vdots \end{pmatrix}.$$

Example 1: Let

$$(c_{ij}^{\varphi}) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} B_1^t\varphi \\ B_2^t\varphi \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

Then,

$$\begin{pmatrix} B_1^{t+1}\varphi \\ B_2^{t+1}\varphi \end{pmatrix} = (c_{ij}^{\varphi}) \begin{pmatrix} B_1^t\varphi \\ B_2^t\varphi \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Namely, by the informing action from agent 2 to 1, as is $(1, 2)$-element of the matrix, agent 1 comes to know φ. ∎

In the following examples, we highlight our attention with the boxed truth values.

Example 2: Suppose the following three kinds of communication matrices:

$$C_1 = \begin{pmatrix} 1 & \boxed{1} & 0 \\ \boxed{1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ C_2 = \begin{pmatrix} 1 & \boxed{1} & 0 \\ 0 & 1 & 0 \\ 0 & \boxed{1} & 1 \end{pmatrix},\ C_3 = \begin{pmatrix} 1 & \boxed{1} & \boxed{1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If there happened a reciprocal and simultaneous communication as to the same information, the matrix becomes symmetric (C_1), that is, agent 2 tells some information to agent 1 (c_{12}) and at the same time so does agent 1 to agent 2 (c_{21}). An agent can announce some information publicly, in which case a certain column is filled with all 1's (the second column of C_2). If multiple different agents tell the same information to a certain agent, then there appear multiple 1's in the same row; if agent 2 and 3 inform the same content to 1, then the situation becomes C_3. ∎

IV. Transitive Communication

Let us consider to connect two communications. Now, we introduce a vector representation for the collective belief state of multiple agents at time t:

$$\boldsymbol{B}^t\varphi = \begin{pmatrix} B_1^t\varphi \\ B_2^t\varphi \\ \vdots \end{pmatrix}.$$

For the time being, we may omit φ unless we need to mention it explicitly. Two consecutive informing actions can be written in the following matrix multiplication.

$$\begin{pmatrix} B_1 \\ B_2 \\ \vdots \end{pmatrix}^{t+2} = \begin{pmatrix} c'_{11} & c'_{12} & \cdots \\ c'_{21} & c'_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & \cdots \\ c_{21} & c_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \end{pmatrix}^t.$$

A. Associativity

First, we need to prove that communication matrices are associative. Let X and Y be communication matrices and $\boldsymbol{B}^t$ be a collective belief state.

$$(XY)\boldsymbol{B}^t = X(Y\boldsymbol{B}^t).$$

As $XY = \bigvee_k (x_{ik} \wedge y_{kj})$,

$$(XY)\boldsymbol{B}^t = \bigvee_l (\bigvee_k (x_{lk} \wedge y_{kl}) \wedge B_l^t) = \bigvee_l \bigvee_k (x_{lk} \wedge y_{kl} \wedge B_l^t).$$

On the other hand, since $Y\boldsymbol{B}^t = \bigvee_l (y_{kl} \wedge b_l)$,

$$X(Y\boldsymbol{B}^t) = \bigvee_k (x_{lk} \wedge \bigvee_l (y_{kl}) \wedge B_l^t) = \bigvee_k \bigvee_l (x_{lk} \wedge y_{kl} \wedge B_l^t).$$

As the both results meet, Q.E.D.

B. Repetitive communication

Let us consider the case that the same communication matrix is employed repeatedly.

$$\boldsymbol{B}^{t+n} = (c_{ij})^n \boldsymbol{B}^t.$$

Example 3: Suppose

$$C = (c_{ij}) = \begin{pmatrix} 1 & \boxed{1} & 0 \\ 0 & 1 & \boxed{1} \\ 0 & 0 & 1 \end{pmatrix},$$

that is, c_{12} and c_{23} are true, besides self-informing. Then,

$$C^2 = \begin{pmatrix} 1 & 1 & \boxed{1} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Namely, agent 3 is reachable from agent 1 in two steps. ∎

Now let $\boldsymbol{B}^t$ be the initial belief state of the community, and let us consider the following sequence.

$$\boldsymbol{B}^{t+1} = C\boldsymbol{B}^t,\ \boldsymbol{B}^{t+2} = C\boldsymbol{B}^{t+1},\ \boldsymbol{B}^{t+3} = C\boldsymbol{B}^{t+2},\ \cdots.$$

Note that the number of 1's in the matrix increases monotonously, since $c_{ii} = 1$ and once an agent believes the proposition (s)he keeps it in his/her recognition. Let $\boldsymbol{B}^*$ be the fixed point such that $\boldsymbol{B}^* = C\boldsymbol{B}^*$. If $\boldsymbol{B}^* = \boldsymbol{B}^{t+k}$, C^k is the transitive closure of the communication graph.

C. Anti-commutativity

As is the case in usual matrix multiplication, communication matrices are not commutative.

Example 4:

$$\begin{pmatrix} 1 & 0 & \boxed{1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 \\ \boxed{1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are communication matrices from agent 3 to 1 (left), and that from 1 to 2 (right), respectively. If agent 3 first believes φ, as

$$\begin{pmatrix} 1 & 0 & 0 \\ \boxed{1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & \boxed{1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ \boxed{1} \end{pmatrix} = \begin{pmatrix} \boxed{1} \\ \boxed{1} \\ 1 \end{pmatrix},$$

agent 2 comes to believe φ. But, when agent 1 does not believe φ, as

$$\begin{pmatrix} 1 & 0 & \boxed{1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ \boxed{1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ \boxed{1} \end{pmatrix} = \begin{pmatrix} \boxed{1} \\ 0 \\ 1 \end{pmatrix}$$

the first informing action, that is from 1 to 2, results in vain, and thus 2 still remains ignorant of φ. ∎

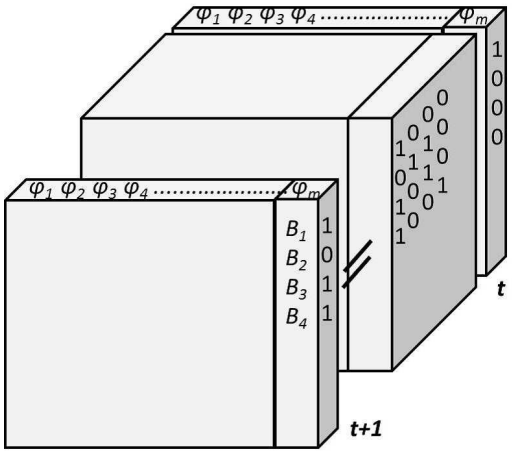


Fig. 2. Multiple Propositions in Informing Action

V. Multiple Information Tensor

Thus far, we have restricted our concern to single information passing. However, we can represent the message passing of multiple information $\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_m$ as a tensor in Fig. 2.

The flat matrix in the front in Fig. 2 represents the resultant state of informing action. The i-th column is the belief states of the agents as to φ_i and the j-th row is what agent j believes, at time $t+1$. In the similar way, the flat matrix in the behind is that of belief states at time t. The in-between $n \times n \times m$-cuboid is a simultaneous communication, where n is the number of agents and m is the number of information. In order to clarify the relation of elements, we place the contravariant elements as superscripts and the covariant ones as subscripts as a tensor:

$$(B_j^{\varphi})^{t+1} = (c_j^{i,\varphi})(B_i^{\varphi})^t.$$

In Fig. 2, we only have shown the atomic propositions. We can add such composite propositions as $\varphi_1 \vee \varphi_2$ and $\varphi_1 \wedge \varphi_2$ simply in the figure, as these truth values are composed by φ_i's.

VI. Model Updating

A kripke frame for multiple agents is such $\mathfrak{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{R}_1, \cdots, \mathcal{R}_n, \mathcal{V} \rangle$ that $\mathcal{A}$ is a set of agents and $\mathcal{R}_i$ is the accessibility of belief modal operator B_i.

A belief state of an agent can be represented also in matrix, when we render (i, j)-element as the accessibility from possible world i to j of Kripke semantics [8]. For example,[3]

$$\begin{array}{c} \\ w_1 \\ w_2 \\ w_3 \end{array}\begin{array}{c} \begin{array}{ccc} w_1 & w_2 & w_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \tag{5}$$

represents $\mathcal{R} = \{w_1 R w_1\ w_1 R w_2,\ w_2 R w_2,\ w_2 R w_3,\ w_3 R w_3\}$ in $\mathcal{W} = \{w_1, w_2, w_3\}$. In this matrix representation, the configuration of truth values directly shows the axioms of modality.

[3]We employ square brackets for the accessibility to distinguish it from the communication matrix.

For example, if the matrix is symmetric, it satisfies axiom of symmetricity (Axiom B). If the diagonal elements are all 1's, it is reflexive (Axiom T), If there is at least one 1 in each row, it satisfies seriality (Axiom D).

A belief change becomes a change in matrix. For example, let $\mathcal{V}(\varphi) = \{w_2, w_3\}$; then matrix (5) cannot satisfy $B_i^t\varphi$ as $\mathfrak{M}, w_1 \not\models B_i^t\varphi$ (for $w_1 R w_1$, $\mathfrak{M}, w_1 \not\models \varphi$). Here, we consider DEL style belief update (2), that is, to cut some of accessibility for an agent to come to believe an informed proposition; namely some 1's in the accessibility matrix at time t are replaced by 0's at $t+1$. In the case of (5), if $(1,1)$-element becomes 0, then $B_i^{t+1}\varphi$.

Since the accessibility with the valuation maps a belief state of a given agent to a truth value:

$$(\mathcal{R}_i, \mathcal{V})\colon B_i^t\varphi \to 1/0,$$

we identify such accessibility matrices with truth values in the following example.

Example 5: Suppose $\mathcal{V}(\varphi) = \{w_2, w_3\}$. A belief vector at time t is:[4]

$$\boldsymbol{B}^t\varphi = (\begin{bmatrix}1&0&0\\1&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&1&0\\1&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&0&0\\1&1&0\\1&0&1\end{bmatrix}, \cdots).$$

After a public announcement of φ from some agent, the revised belief vector becomes:

$$\boldsymbol{B}^{t+1}\varphi = (\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}0&1&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}, \cdots). \blacksquare$$

Now, the multiple belief states of agents at time t becomes $(n \times m \times k^2)$-hypercuboid, where n is the number of agents, m is the number of information, and k is the number of possible worlds.[5]

VII. Discussion

In the belief revision of multiple agents, as there are so many indices appears, we have represented them in linear logic. As we have shown, the collective belief state and the communication matrix include more then three indices, those matrices become hypercuboids.

We have shown that the consecutive informing action can be realized by a product of matrices, including the n-th power, with associativity and anti-communtativity. This discussion naturally leads us to the inversion of matrix, as

$$\boldsymbol{B}^t = (c_{ij})^{-1}\boldsymbol{B}^{t+1},$$

being the inversion regarded as belief contraction [5]. However, some communication matrices do not have their inversions, as their rank fails to be n. Furthermore, we cannot give proper semantics for value '-1', or '0' may appear on the diagonal elements in the inversion matrix. We need to recognize that how we can *redo* the revision is difficult problem, especially in Kripke semantics.

In this paper, we evaluated formulae in the same possible world even though time shifts. We are now to intend that we regard the world itself as a temporal state, as:

$$\mathfrak{M}, t \models [c_{ij}^{\varphi}]\psi \iff \mathfrak{M}^{c_{ij}^{\varphi}}, t+1 \models \psi.$$

Now, let us get back to the two issues: belief revision and nested modalities. We could not implement simultaneous arrivals of two different information to one agent, i.e.,

$$B_j^{t+1} := B_j^t \dot{+} \{\varphi\} \dot{+} \{\psi\},$$

regarding B_j^t as a belief set and '$\dot{+}$' as revision operator. Note that how we can revise the belief of each agent is an independent topic from our formalism in this paper, and depends on the preference of revision. Also, it is difficult to send such propositions including modality as $B_j\varphi_2$, which results in the nested belief state. In addition, sending a negative formula also affects how we can revise the accessibility; these would be our common research topics in the community of belief update logic in future.

[4] We show belief vectors in row vectors here just for visibility, although they are intrinsically column vectors.

[5] If we fix the number of atomic propositions to m, then the number of maximal possible worlds becomes $k = 2^m$.

References

[1] Baltag, A., Moss, L. S., Solecki, S. : The Logic of Public Announcements, Common Knowledge, and Private Suspicions. Technical Report TR534, Department of Computer Science(CSCI), Indiana University (1999)

[2] van Ditmarsch, H., van der Hoek, W., and Kooi, B.: *Dynamic Epistemic Logic*, Springer (2008)

[3] Foundation for Intelligent Physical Agents(FIPA), Communicative act library specification, http://www.fipa.org (2000)

[4] Fusaoka,A., Nakamura, K., Sato, M.: On A Linear Framework for Belief Dynamics in Multi-agent Environment, in Inoue, K., Satoh, K., Toni, F. (eds.), Computational Logic in Multi-Agent Systems, CLIMA VII, LNAI4371, Springer Verlag, pp.41–59 (2007)

[5] Gärdenfors, P. : Belief revision, Cambridge University Press (1992)

[6] Hagiwara, S., Kobayashi, M. and Tojo, S. : Belief Updating by Communication Channel, in Inoue, K., Satoh, K., Toni, F. (eds.), Computational Logic in Multi-Agent Systems, CLIMA VII, LNAI4371, Springer Verlag, pp.211–225 (2007)

[7] Kobayashi, K. and Tojo, S. : Agent Communicability in Belief Update Logic, In *Proc. Declarative Agent Languages and Technology* (DALT), pp.207–221 (2008)

[8] Liau, C-G.: Matrix Representation of Belief States: an Algebraic Semantics for Belief Logics. Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems, vol.12, no.5, pp.613–633 (2004)

[9] Meyer, Ch. J-J. : Dynamic logic for reasoning about actions and agents, Spring Carnival of Philosophical Logic, Uppsala University (2001)

[10] Parikh, R., Pacuit, E., Cogan, E. : The logic of knowledge based obligation, In *Proc. Declarative Agent Languages and Technologies*(DALT), pp. 53–60 (2006).

[11] Plaza, J. A. : Logics of public communications, in Emrich, M. L., Pfeifer, M. S., Hadzikadic, M., and Ras, Z. W. (eds.), *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pp. 201–216 (1989)

[12] Satoh, K. and Tojo, S.: *Disjunction of Causes and Disjunctive Cause: a Solution to the Paradox of Conditio Sine Qua Non using Minimal Abduction*, The 19th International Conference on Legal Knowledge and Information Systems (JURIX) (2006)

[13] Hagiwara, S. and Tojo, S. JURIX, volume 152 of Frontiers in Artificial Intelligence and Applications, page 111-120. IOS Press, (2006)

[14] Yamada, T. : Acts of commanding and changing obligations, in Inoue, K., Satoh, K., Toni, F. (eds.), Computational Logic in Multi-Agent Systems, 7th International Workshop, CLIMA VII, Hakodate, Japan, May 2006, Revised Selected and Invited Papers, Lecture Notes in Artificial Intelligence, Band 4371, Springer Verlag, pp.1–19 (2007)

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 179–182

Medical Decision Support System Architecture for Diagnosis of Down's Syndrome

Hubert Wojtowicz*, Jolanta Wojtowicz*, Wojciech Koziol* and Wieslaw Wajs†
*University of Rzeszow
Faculty of Mathematics and Nature
Institute of Computer Science
Email: hubert.wojtowicz@gmail.com
† AGH University of Science and Technology
Faculty of Electrical Engineering
Institute of Automatics
Email: wwa@ia.agh.edu.pl

***Abstract*—The paper presents the development of a new system that is used to solve the problem of the recognition of the dermatoglyphic pattern and the understanding of the classification process of the symptoms of Down's syndrome. The method used in the system for diagnosing Down's syndrome in infants is based on the combination of text knowledge found in the scientific literature describing Down's syndrome with the knowledge obtained from the analysis of dermatoglyphic indices characteristic of Down's syndrome with the use of digital pattern recognition techniques. The scientific goal is to design a classifier system that realizes automatic medical diagnosis through the application of an expert system designed on the basis of knowledge included in the scientific text descriptions of the Down's syndrome. One other aim is the application of the pattern recognition algorithms to the analysis of indices present in the images of dermatoglyphic patterns. This approach, similar to the approach used by anthropologists, is realized by the system through the juxtaposition of the knowledge described in the form of expert system rules and the information provided by the appropriate digital equipment, and on the basis of this juxtaposition an arbitrary classification of the investigated patterns is performed.**

I. Introduction

THE present state of knowledge in the discipline of medical pattern recognition allows the extraction of image features using computer methods and data processing algorithms. The results obtained using these methods are presented to an expert who affirms in an arbitral way his understanding of the pattern presented to him as an after-effect of the occurrence of a pathological process and describes it using terms established in the scientific medical literature. In the present state of science and technology development the number of available patterns and the amount of information included in the scientific literature is constantly increasing. It can be assumed with certainty that both of these elements will have a tendency to increase in a hard to predict tempo. The community of medical experts copes with the high increase in information by dividing medical knowledge into many new specializations. The main and basic source of professional knowledge is the knowledge written in the form of natural language sentences. The attempt described in the project is to combine knowledge extracted from scientific medical literature with features extracted and classified by pattern recognition algorithms in the process of analyzing digital images of infants' dermatoglyphic patterns. In its basic form this approach is based on building a hybrid decision support system that comprises an expert system module inferring the occurrence of a genetic disorder on the basis of the results passed from pattern recognition modules that analyze the dermatoglyphic images automatically without the participation of a human expert. The success of this approach may lead to the application of a computer technique to achieve near complete automation of screening tests for the detection of Down's syndrome in infants.

II. The Aim of the Work

The tasks of the recognition and understanding of dermatoglyphic patterns, whose results form the basis for inferring the occurrence of a genetic disorder in infants, are complex issues. The classification of the patterns and the diagnosing of the presence of a genetic disorder on the basis of recognitions is carried out by professional anthropologist. The service of automatic pattern recognition and clinical decision support designed in the form of a telemedical system can perform specialized screening tests of dermatoglyphic data delivered from medical centers that do not employ anthropologists. The data is collected in a non-invasive manner using touch scanners or specialized cameras and then sent to a distant server that is running the telemedical system via the Internet. The uploaded data is then subject to a detailed analysis whose aim is the extraction of features from the collection of images on the basis of which the classification of the case is carried out. It should be emphasized that it is possible to implement and use the designed system with the use of the existing IT infrastructure of local hospitals.

The design of the system involves the following information processing scheme:

1. The analysis of texts containing specialized field knowledge in the form of natural language sentences, arithmetic expressions and arithmetic-logic relations leads to the formulation of conditions which are basis for the conclusion.

978-83-60810-53-8/$25.00

2. The synthesis of partial information contained in the digitally stored image leads to the generation of features that represent characteristic image patterns.

As a result of the proposed scheme application a set of rules for the expert system is obtained, on the basis of which the probability of the occurrence of a genetic disorder in infants is determined. The calculation of the values of premises for the expert system is realized through the determination of the classes of the analyzed patterns of particular dermatoglyphic areas and through the assessment of other features of dermatoglyphs. The values of the conclusions of the expert system that was built on the basis of a text analysis allow for the qualification of the medical case to the group of healthy infants or to the group of infants with a genetic disorder. Undertaking the research topic results from the difficulty in the direct access to the screening tests that are carried out by an anthropology specialist. The proposed decision support system allowing for remote access to these tests overcomes limitations that result from the shortage of employment of specialists in small medical centers. Therefore it brings a new substantial value into social life and helps in the task of improving the accessibility to specialized medical services. It is assumed that the application of the system will improve the effectiveness of the treatment, i.e. the number of complications due to improper treatment of infants with certain genetic defects will decrease and thus the costs of the treatment and the length of hospital care will also decrease. The authors believe that the system will prove particularly useful as a support system for doctors working in small hospitals which do not employ anthropologists and as an automatic system for screening tests of infants performed on a large scale. The device for the analysis of dermatoglyphs will consist of an average desktop computer with a touch scanner or a digital camera attached to it. Depending on the computing capabilities of the workstation, it can serve as an independent unit performing diagnosis or as a terminal used for the acquisition of and for sending data over the Internet and for displaying the results of the analysis.

III. Design of the Medical Decision Support System

The project of the decision support system is of interdisciplinary character as it combines the achievements of modern sciences in the fields of computational intelligence, digital image processing, pattern recognition and the design of expert systems containing medical knowledge. The design proposed by the author assumes the modular architecture of the diagnostic system. Based on the design of dermatoglyphic nomogram [4] the system consists of four main modules. One may distinguish three modules that realize the pattern recognition of medical images. Another superior module in the form of an expert system generates diagnosis on the basis of recognition results that come from pattern recognition modules.

The role of the first of the pattern recognition modules is the classification of fingerprints. Fingerprint classification is one of the fundamental tasks of dermatoglyphic analysis. Several

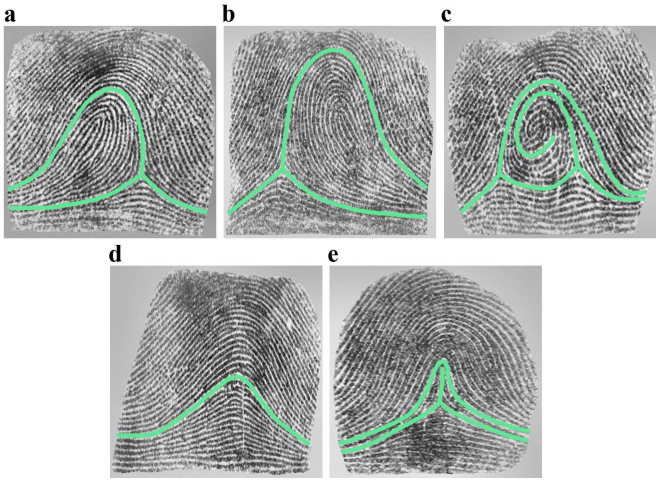


Fig. 1. Classification scheme of fingerprints: (a) left loop (LL); (b) right loop (RL); (c) whorl (W); (d) plain arch (A); (e) tented arch (TA). (The green lines traced on the prints are the type lines, which define the unique skeletons of the patterns.)

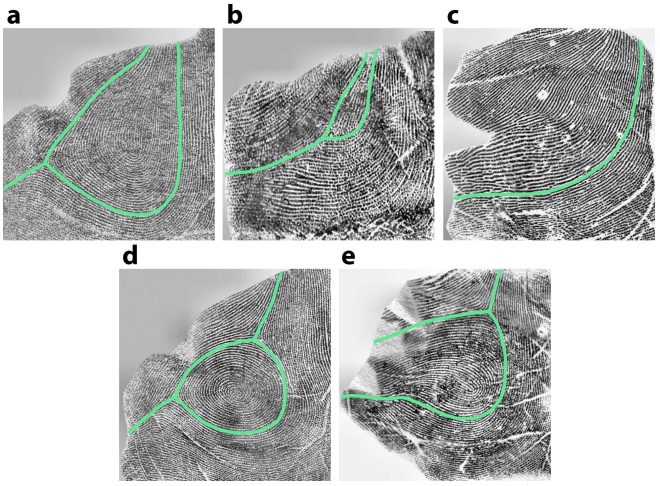


Fig. 2. Classification scheme of the hallucal area of the sole prints: (a) large distal loop (LDL), (b) small distal loop (SDL), (c) tibial arch (TA), (d) whorl (W), (e) tibial loop (TL).

methods of fingerprint classification have been proposed in scientific literature and are used for a variety of applications relating to the analysis of fingerprints [2]. The classification method used in the dermatoglyphic analysis is called Henry's classification method after the name of the originator of this method. It distinguishes between the following five classes of fingerprint patterns: left loop (LL), right loop (RL), whorl (W), plain arch (A) and tented arch (TA) Fig. 1.

The second module performs the classification of patterns of the hallucal area of the sole. A classification scheme of hallucal area patterns includes the following classes: large distal loop (LDL), small distal loop (SDL), tibial arch (TA), whorl (W) and tibial loop (TL) Fig. 2.

Both pattern recognition modules employ the image pro-

cessing algorithms for the segmentation of the background and contrast enhancement of the analyzed images. An important element of enhancement of the local ridge structures of dermatoglyphs is the application of contextual image filtration, which is realized by STFT algorithm [1]. In the classification process, the features extracted from the enhanced images represent the unique properties of the patterns contained within the images. The features used by the classification modules are local ridge flow directions of dermatoglyphic patterns. The classification is accomplished by an ensemble of SVM classifiers that make use of the RBF kernel functions in the learning process and that are trained with the use of the one vs. one voting scheme. Both classification modules recognize the patterns of appropriate dermatoglyphic areas with 90% accuracy ratio.

The third pattern recognition module is used to determine the ATD angle of the right palm print Fig. 3. The value of this angle is determined by the location of digital triradii A and D and the axial triradius t. A reliable and accurate identification of the location of the characteristic points is a complex issue, therefore the algorithm devised for finding these points uses two independent local image descriptors calculated in different ways. The first of these descriptors is an improved variant of the Poincare index [7] that is determined from the ridge flow directions map calculated by the algorithm based on the image pyramid decomposition and PCA [3]. The second of the determined descriptors is a local coherence map calculated from the image texture. For each point of the image containing the analyzed pattern, the values of eigenvectors which create the coherence map are calculated. Eigenvectors are calculated from the confusion matrix that contains the values determined for each image pixel with the use of multiplication of a local image segment centered in pixel (i,j) and the combination of two dimensional Gauss-Hermite moments [7]. The information in the form of the Poincare index and the information in the form of a coherence map are compared for all pixels of the image that contains the dermatoglyphic pattern. The points of the image in which the values of both of these descriptors simultaneously indicate the occurrence of a characteristic point are considered to be the true characteristic points.

The fourth module implemented as an expert system is superior to the modules that carry out the pattern recognition tasks. On the basis of the recognition results that come from pattern recognition modules this module carries out an automatic diagnosis which determines the qualification of the infant to a group of healthy children or to a group of children with Down's syndrome. The basis of the design of an expert system is a dermatoglyphic nomogram. The dermatogram was created on the basis of statistical test results. The statistical test allowed for selecting, from the group of all known dermatoglyphic features that indicate the presence of a genetic disorder, the four most significant features on the basis of which a credible diagnosing of the likelihood of the presence of Down's syndrome in newborns is possible [4].

The premises in the rules of an expert system are the recognized types of patterns of the dermatoglyphic areas that

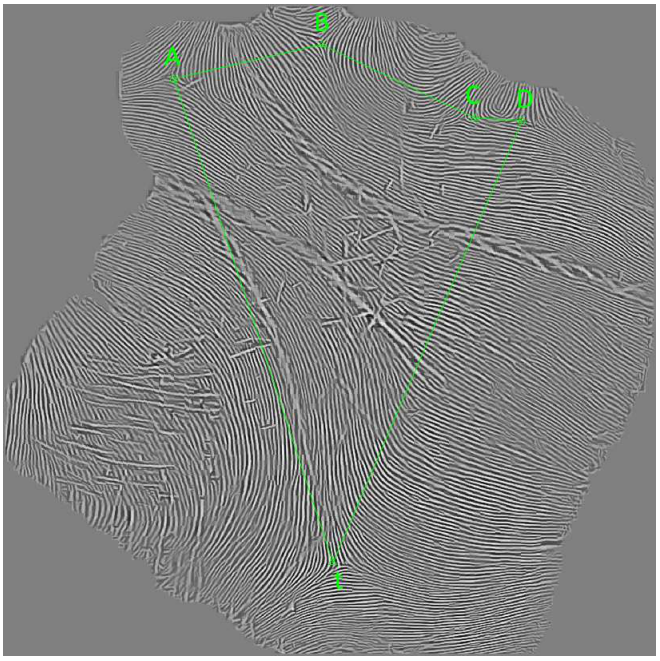


Fig. 3. Singular points of the palm print located with the use of a two stage algorithm that use the improved Poincare index and Gauss-Hermite moments.

contain the four significant features which are: pattern types of index fingers of the left and right hand, pattern type of the hallucal area of the right sole, value of the ATD angle of the right palm. The conclusions of the expert system are diagnostic values determined for diagnostic criterions of the dermatogram that result from the types of dermatoglyphic patterns and from the value of the ATD angle. A combination of these diagnostic criteria values determines the value of the dermatogram diagnostic score and the fact of Down's syndrome occurrence in the newly-born. For the dermatogram 125 possible combinations of patterns exist, which corresponds to the same number of expert system rules.

Table I presents a set of premises and diagnosis results for the rules that correspond to the combination of the recognition of the left and right hand index fingerprints patterns UL - UL (denoting the left loop pattern type on the index finger of the left hand and the right loop pattern type on the index finger of the right hand) and all the possible combinations of the hallucal area of the sole pattern types and the ranges of the ATD angle values.

IV. Results

In the course of the research carried out, components of the system have been designed and implemented. The system will allow for an automatic diagnosis of the occurrence of genetic disorders in infants on the basis of sets of dermatoglyphic images. The system has a modular design. The implementation of the system required the application of numerous computer technologies. The modules responsible for carrying out image processing, feature extraction and pattern recognition tasks

TABLE I
A PARTIAL SET OF EXPERT SYSTEM RULES FOR THE EXAMPLE COMBINATION OF DERMATOGLYPHIC PATTERNS

Combination	Right Index Finger	Left Index Finger	Right Hallucal Area	Right ATD Angle	Diagostic Line Index
1	UL	UL	LDL	⟨15;28⟩	Normal
	UL	UL	LDL	(28;78⟩	NN
	UL	UL	LDL	(78;120⟩	Down
2	UL	UL	Other	⟨15;37⟩	Normal
	UL	UL	Other	(37;88⟩	NN
	UL	UL	Other	(88;120⟩	Down
3	UL	UL	SDL	-	Normal
	UL	UL	SDL	⟨15;58⟩	NN
	UL	UL	SDL	(58;120⟩	Down
4	UL	UL	TbA	-	Normal
	UL	UL	TbA	⟨15;31⟩	NN
	UL	UL	TbA	(31;120⟩	Down
5	UL	UL	W or FL	⟨15;34⟩	Normal
	UL	UL	W or FL	(34;85⟩	NN
	UL	UL	W or FL	(85;120⟩	Down

have been implemented with the use of the Matlab and C languages. The database containing dermatoglyphs was designed and implemented with the PostgreSQL database management system. In the database, collections of images are stored along with their descriptions. The diagnostic module in the form of an expert system has been implemented in the Prolog language. The analysis of the collection of images realized by pattern recognition modules allows for the classification of fingerprint patterns [5], for the classification of the hallucal area of the sole patterns [6] and the determination of the value of the ATD angle of the palm print. The outcomes of the classifications and image features calculations are passed to the set of rules of the expert system which on their basis calculates the value of the diagnostic score that determines to which of the three groups the diagnosed infant belongs: healthy infants, infants with Down's syndrome, or infants for whom the value of the diagnostic index does not give a clear answer as to the occurrence of a genetic disorder. The client application implemented in the project is equipped with typical useful capabilities such as enabling users to search and view data. It also provides scientific capabilities allowing for the following: sending requests to the decision support server to perform a dermatoglyphic analysis, the visualization of analysis outcomes in a numerical form and the visualization of the diagnosis results, generated by the explanation facilities of the expert system module in the form of a text description.

V. SUMMARY

The paper presents the architecture of the medical decision support system for the diagnosis of Down's syndrome in infants on the basis of collections of images. The results of the research on the system have reached a high level of advancement. From the technical standpoint, the modules responsible for the pattern classification of the fingerprint and hallucal area of the sole impressions were accomplished. A procedure for the calculation of the ATD angle of the palm has been proposed and implemented. A decision module that determines the diagnostic index value on the basis of the results passed from pattern recognition and image parameters analysis modules has also been implemented. The intention of the authors is an implementation of the system in a form that is useful for public institutions (hospitals, health care institutions, etc.) and for medical universities that use the advisory nature of the diagnostic system to assist in the education of medical students. A client application will be installed in dedicated terminals as a tool that supports the process of infants' diagnosis. A stand-alone application designed for personal computers, having the functionality of the above-mentioned application, is also planned. This software will serve the public as a free to use assistance for those interested in performing a stand-alone diagnosis. After a full implementation of the system, its further development is assumed depending on the degree of extension of the dermatoglyphic database used as a knowledge base. The users of the system (medical doctors, other medical personnel) will be encouraged to upload the data collected in their daily work to the database and thus improve its efficiency and thereby contribute to the development of the system.

REFERENCES

[1] S. Chikkerur, A. N. Cartwright and V. Govindaraju, "Fingerprint image enhancement using STFT analysis," *Pattern Recognition,* vol. 40, pp. 198–211, Elsevier (2007).

[2] H. Cummins and C. Midlo, "Fingerprints, palms and soles - Introduction to Dermatoglyphics," Dover Publications Inc., New York (1961).

[3] X. G. Feng and P. Milanfar, "Multiscale principal components analysis for image local orientation estimation," *Proc. of the 36th Asilomar Conf. on Signals, Systems and Computers,* vol. 1, pp. 478–482, (2002).

[4] T. E. Reed, D. S. Borgaonkar, P. M. Conneally, P. Yu, W. E. Nance and J. C. Christian, "Dermatoglyphic nomogram for the diagnosis of Down's syndrome," *The Journal of Pediatrics,* vol. 77, no. 6, pp. 1024–1032, (1970).

[5] H. Wojtowicz and W. Wajs, "Intelligent Information System for Interpretation of Dermatoglyphic Patterns of Down's Syndrome in Infants," *Proc. of 4th Asian Conference on Intelligent Information and Database Systems (ACIIDS),* Springer - Verlag, LNAI vol. 7197, pp. 284–293, (2012).

[6] H. Wojtowicz and W. Wajs, "Classification of Plantar Dermatoglyphic Patterns for the Diagnosis of Down's Syndrome," *Proc. of 5th Asian Conference on Intelligent Information and Database Systems (ACIIDS),* Springer - Verlag, LNAI vol. 7803, pp. 295–304, (2013).

[7] Y. Yin and D. Weng, "A New Robust Method of Singular Point Detection from Fingerprint," *Proc. of Int. Symposium on Information Science and Engieering,* IEEE, (2008).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 183–190

An Investment Strategy for the Stock Exchange Using Neural Networks

Antoni Wysocki and Maciej Ławryńczuk
Institute of Control and Computation Engineering, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland, tel. +48 22 234-76-73
Email: A.T.Wysocki@stud.elka.pw.edu.pl

***Abstract*—This paper describes a neural system which helps to make the current investment decisions. Some well known financial indicators usually considered by investors are inputs of the system. The basic problem is to select appropriately the indicators which would give the best predictor. Two methods are used and compared: the combination method and the correlation method. To analyze the problem daily quotations of companies included in the Warsaw Stock Exchange Index (WIG20) are used.**

Keywords: Stock exchange, prediction, nonlinear modeling, neural networks, soft computing.

I. Introduction

NEURAL networks [5] are universal approximators [6]. It means that a network with at least one hidden layer is capable of approximating any nonlinear smooth function to an arbitrary accuracy (provided that the number of hidden units is sufficient enough). Multilayer perceptron neural networks are most common. As the practical experience indicates, they can be successfully used in numerous application, including e.g.: pattern recognition [1], [12], numerical methods [2], biomedical engineering [7], optimization [9], fault diagnosis [8], robotics [11], load forecasting in a power system [13] and control algorithms [14], [15]. Neural network can be also used in financial forecasting [4], [16]. Usually, a stock exchange time-series model is trained the role of which is to predict the future price of shares. Although such a "black-box" approach is commonly used in system identification but it is completely different from the classical technical analysis methods [3], [10] popular in banks and in the financial community. In this work neural networks are used for developing an investment strategy. Unlike "black-box" approaches some technical analysis indicators are used as inputs of the system. Investors often consider those indicators before making decisions because they have intuitive interpretation. Because there are various indicators used in the technical analysis, the basic problem to solve is to select the indicators which would give the best predictor.

II. Description of the problem

One of the very interesting investment strategies is based on stock quotes. In some periods the stock quotes stabilize for some time and share price fluctuations are very small. Such a time period is named stagnation. It is not difficult to find such periods on the price chart because the value of the shares should be analyzed from the last 4 weeks (20 market days). From observations of the stock quotes it can be observed that they appear immediately after a period of stagnation followed by a sharp change in the value of shares (either upward or downward). The investment strategies frequently used by investors use the popular technical analysis indicators. The "buy", "sell" and "hold" decisions are made taking into account the values and behavior of selected indicators in periods of stagnation points of interest.

In this paper the investment strategy shortly described above leads to developing a neural network which would be able to serve as the decision support systems. The neural network uses the information represented by the classical technical indicators well understand by the investors. The objective of the neural network is to answer the question whether or not the investor should invest. To analyze the problem daily quotations of the last 10 years of the biggest 20 companies included in the Warsaw Stock Exchange Index (WIG20) are used. This choice was made due to the fact that the stock market of WIG20 is the most liquid, it allows for greater investment in less time.

For determination of important points associated with finding periods of stagnation of stock prices a time window of variable width of 20 days is used, which measures changes in the average value of the shares. It is assumed that the interesting point is when the change in value of the shares is greater than the average change in the time window, while the previous two trading points are characterized by below-average volatility of the window.

Fig. 1 shows the interesting points on the background of the graph of a company's share price. It can be observed that in the periods prior to the occurrence of these points there are periods of stagnation of stock prices, and in the near future from the important points share price rises or falls rapidly.

III. The neural predictor

The set of all points of interest is divided in half: the first part is the training data set for the neural network, the second part is the test data set.

From an analysis of all the points of interest it can be concluded that 93% of those points actually lie before significant changes of stock prices. The interesting points that lie before increases in the shares of companies described as positive points, but those points, followed by the decline in stock prices described as negative points. The analysis of points of interest is shown in Table I.

978-83-60810-53-8/$25.00

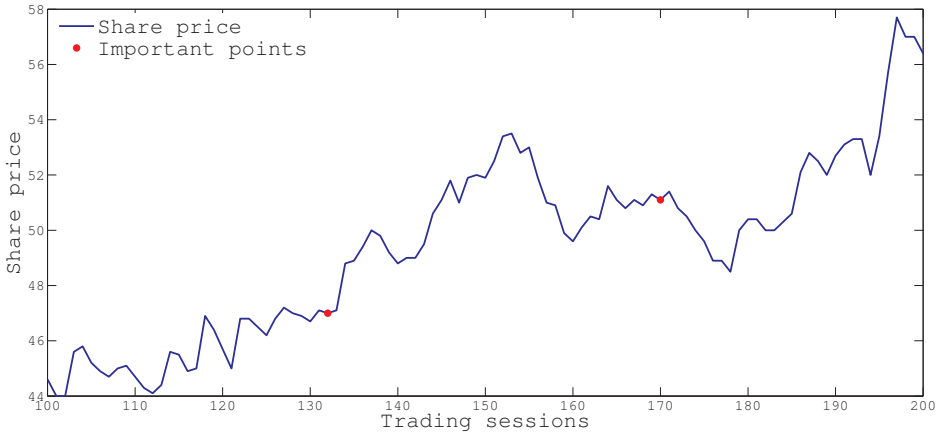


Fig. 1. Interesting points on the background of the graph of a company's share price

TABLE I
ANALYSIS OF A SET OF POINTS OF INTEREST

Data set	Positive points	Sum of all points	Good investments (%)
All data	1187	2251	52.73%
Training data set	582	1126	51.69%
Testing data set	605	1125	53.78%

While analyzing data from a set of points of interest one can identify a strategy that conventionally is called "the trivial strategy". It is based on the fact that if the investor finds an interesting point, he or she invests. The percentage of all positive points in the set of all points of interest is 52.73%. It means that if the investor uses any signal of an interesting situations to invest in the stock market, in 52.73% of such situations he or she would be successful.

It can be argued that "the trivial strategy" is "buy" and "hold". However, the art of investing in the stock market, which is characterized by quickly transfer money between investments so that the whole time money is working in an optimal way.

The task of the intelligent investor and the proposed neural decision support system is to analyze the points of interest, reject the situations that lead to losses, and recognize those that generate profit. In order to analyze the problem considered in the article it is necessary to answer the following questions:

1) Do the indicators proposed by the investors carry the necessary information that allows the choice of the positive points?
2) Are all the indicators proposed by the investors required to make the correct decision?
3) Is it possible to construct a computerized system of investment, which would give the results much better than the trivial strategy?

IV. CREATING A NEURAL NETWORK

A multi-layer perceptron neural network with nonlinear (tanh) hidden neurons and one linear output neuron is used for modeling. Due to the fact that the study is based on finding the correct classification of investment situation, in the output layer of neural network one linear neuron is used. When the network output gives a positive value, the investment situation is recognized as opportunity and the system suggests to invest. Whet the network output gives a negative value, the system suggests to omit the investment. The linear output neuron is chosen, because it can show general suggestion of investing or omitting the investment, but also magnitude of this signal can tell the investor if this investment situation is clear or it is difficult to take a decision. When the system gives a positive value, but this value is close to 0, the investor should be aware of the increased risk of this particular investment. On the other hand, when there is a big output value, it may suggest that the investment situation is clear and the risk is small. The risks is understood as the number of conflicting signals from the analyzed indicators. If a large majority of indicators generates the same signal, the risk of making a wrong decision is small.

The following assumptions are made:

1) All the indicators suggested by the investors are chosen as network inputs.
2) Neural network with different number of hidden nodes are trained and compared.
3) During training the network classification error is calculated for the training data set, the test data sets are used to finally choose the structure of the neural system.

The neural network are trained by means of the Levenberg-Marquardt algorithm. Initial experiments indicate that after 200 iterations of the training procedure the classification error stabilizes and further training does not bring better results. Thus, in all further experiments, the number of training iterations is 200.

Classification accuracy (percentage of hits in a good investment) for different number of hidden nodes is shown in Fig. 2. The network with 15 hidden neurons is finally chosen because it gives classification results better than the results obtained by means of all other networks (the test data set is taken into account). Smaller networks have low approximation abilities whereas bigger networks are overparameterized, they have too many parameters (weights).

The answers to the questions put above (in the previous section) require the study of the influence of the neural network inputs on the classification accuracy. Thus, the above questions are equivalent to the following:

1) Do the neural network input signals allow the classification and selection of positive points?
2) What is the number of inputs that allows the best classification and selection of positive points?
3) Does the best possible network achieves better results than the trivial strategy?

A. Technical analysis indicators used in the investment strategy

When one has a set of interesting points, the next step is to calculate the value of technical analysis indicators for the points. As recommended by investors from the Warsaw Stock Exchange, seven indicators are taken into account: the slow stochastic oscillator %K %D, the Moving Average Convergence-Divergence (MACD), the Commodity Channel Index (CCI), the Relative Strength Index (RSI), the three backward linear regression values for the 5, 10 and 15 days.

1) The slow stochastic oscillator %K %D (the 1^{st} *input):* Stochastic oscillator [3] is commonly used by traders. It tracks the relationship of each closing price to the recent high-low range. The stochastic oscillator consists of two lines: a fast line called %K and a slow line called %D. The first step is to calculate the %K from this equation:

$$\%\text{K} = \frac{C_{tod} - L_n}{H_n - L_n}$$

where C_{tod} – today's close, L_n – the lowest point for the selected number of days, H_n – the highest point for the selected number of days, n – the number of days. The second step is to obtain %D. It is done by smoothing %K over a three-day period:

$$\%\text{D} = \frac{\text{3-day sum of } (C_{tod} - L_n)}{\text{3-day sum of } (H_n - L_n)} \cdot 100$$

Fast stochastic oscillator is very sensitive to the returns on the market, but it leads to many erroneous signals. Many investors use the slow stochastic oscillator, which is less sensitive. The value of %D of fast stochastic oscillator becomes the %K of slow stochastic oscillator and is smoothed by repeating step 2 to obtain the value of %D of slow stochastic oscillator. An example slow stochastic oscillator is demonstrated in Fig. 3. The stochastic lines help identify top and bottom areas when they move above or below their reference lines. The stochastic oscillator gives its best signals when it diverges from prices.

2) Moving Average Convergence-Divergence (the 2^{nd} *input):* The MACD index [3] consists of three Exponential Moving Averages (EMAs). It appears on the charts as two lines whose crossovers give trading signals. The original MACD indicator consists of two lines: a solid line (called the MACD line) and a dashed line (called the signal line). The MACD line is made up of two Exponential Moving Averages. It responds to changes in prices quickly. The signal line is made up of the MACD line smoothed with another Exponential Moving Average. It responds to changes in prices more slowly. Fig. 4 demonstrates an example MACD index. To create MACD one has to:

1) calculate a 12-day EMA of closing prices,
2) calculate a 26-day EMA of closing prices,
3) subtract the 26-day EMA from the 12-day EMA, and plot their difference as a solid line (it is the fast MACD line),
4) calculate a 9-day EMA of the fast line and plot the result as a dashed line (it is the slow Signal line).

When the fast MACD line crosses above the slow signal line, it gives the "buy signal". When the fast line crosses below the slow line, it gives the "sell signal".

3) Commodity Channel Index (the 3^{rd} *input):* The CCI [10] is an oscillator originally developed by Donald Lambert. Since its introduction the indicator has grown in popularity and it is now a very common tool for traders to identify cyclical trends. The CCI was developed to determine overbought and oversold levels. It is done by measuring the relation between price and a moving average (MA), or, more specifically, normal deviations from that average. The value of the CCI is calculated from

$$\text{CCI} = \frac{1}{0.015} \cdot \frac{p_t - \text{SMA}(p_t)}{\sigma(p_t)}$$

where p_t – typical price (average of the maximum price, minimum price and closing price), SMA – the arithmetic moving average, σ – the absolute deviation. Fig. 5 demonstrates an example CCI index.

Possible "sell" signals are:

- the CCI crosses above 100 and has started to curve downwards,
- there is bearish divergence between the CCI and the actual price movement, characterized by downward movement in the CCI while the price of the asset continues to move higher or moves sideways.

Possible "buy" signals are:

- the CCI crosses below -100 and has started to curve upwards,
- there is a bullish divergence between the CCI and the actual price movement, characterized by upward movement in the CCI while the price of the asset continues to move downward or sideways.

4) Relative Strength Index (the 4^{th} *input):* The RSI index [3] measures the strength of any trading vehicle by monitoring

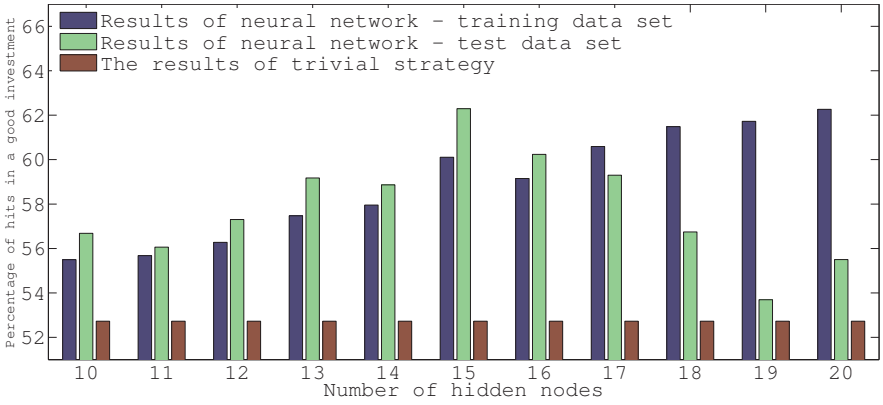


Fig. 2. Testing different structure of neural network

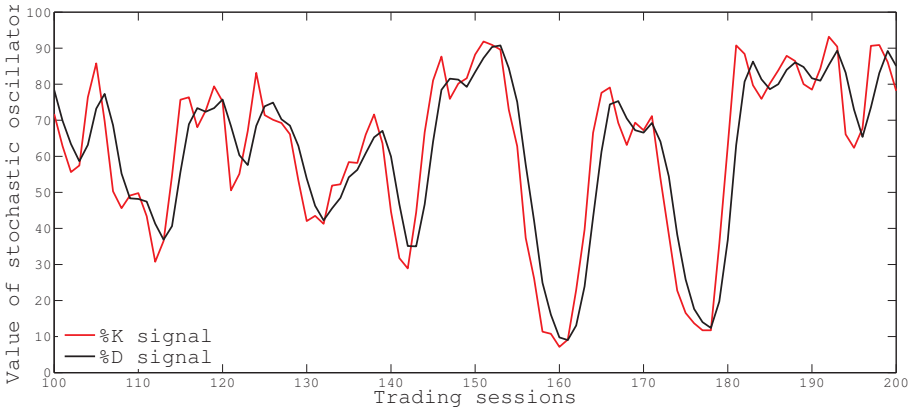


Fig. 3. The example slow stochastic oscillator

changes in its closing prices. It is a leading indicator, it is never a laggard. Its value is

$$\mathrm{RSI} = 100 - \frac{100}{1 - \mathrm{RS}}$$

where

$$\mathrm{RS} = \frac{\text{average of net UP closing changes for n days}}{\text{average of net DOWN closing changes for n days}}$$

The RSI fluctuates between 0 and 100. When RSI reaches a peak and turns down, it identifies a top. When RSI falls and then turns up, it identifies a bottom. Fig. 6 demonstrates an example RSI index.

Divergences between RSI and prices give the strongest "buy" and "sell" signals. They show when the trend is weak and ready to reverse. Horizontal reference lines must cut across the highest peaks and the lowest valleys of RSI. They are often drawn at 30 and 70.

Bullish divergences give "buy" signals. They occur if prices fall to a new low but RSI makes a more shallow bottom than during its previous decline. One buys as soon as RSI turns up from its second bottom. "Buy" signals are strong if the first RSI bottom is below its lower reference line and the second bottom is above the line.

5) Backward linear regressions (the 5^{th}*,* 6^{th} *and* 7^{th} *inputs):* The last information that investors take into account are backward linear regressions. Based on trading behavior in the recent past, investors are trying to guess the future behavior of the market. Three backward linear regressions are taken into account: 5, 10 and 15 days. Investors are interested whether these regressions indicate an increasing or a decreasing trend. The relationships between these trends suggest further course of trading stocks. Fig. 7 shows the example trends of the linear regressions of the corresponding points of interest.

One of the basic rules used by investors is to invest in accordance with the medium-term trend when the linear regressions of 15 and 10 days reference the same trend. Fig. 7 shows two interesting points, one of which meets the above condition. Linear regressions 15 and 10 days for the second point shows the different trends, so there is a signal to omit the investment.

6) Network output – 21 days forward linear regression: All the indicators describing the selected points of interest are enough for investors to make a decision to proceed with the investment or its omission. In the case of neural network modeling the given signal is used which indicates whether the situation related to the point, in fact, is an important investment

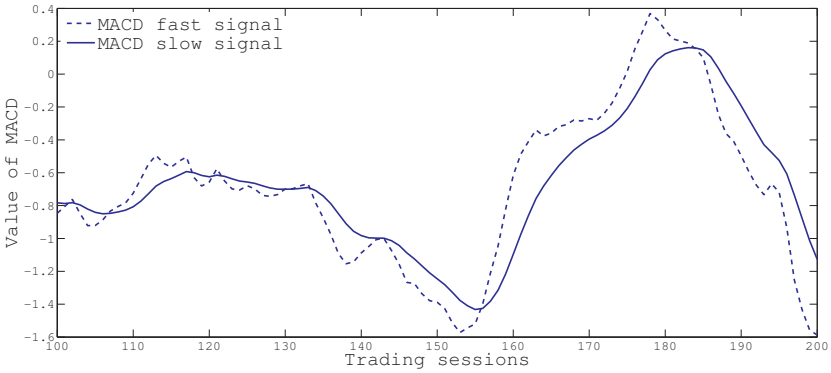


Fig. 4. The example Moving Average Convergence-Divergence (MACD) index

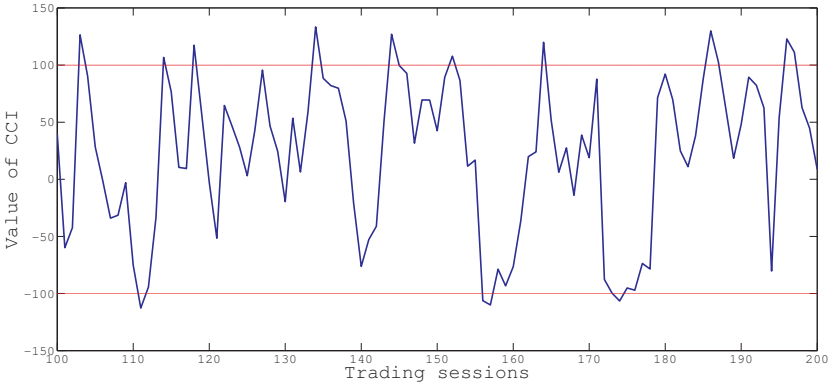


Fig. 5. The example Commodity Channel Index (CCI)

situation (it is possible to make money on it) or not. The investment horizon is also determined by investors and its value is 21 days. The linear regression is used to check if at the relevant time horizon value of the shares is in an uptrend or not. An example of the 21 days forward linear regression is shown in Fig. 8.

V. The Choice of Inputs of the Neural Predictor

The basic problem is therefore the choice of inputs of the neural network and determination how many inputs and how their combination allows to achieve the best results. Two approaches which make it possible to choose the inputs of neural networks are discussed. Those methods were chosen to represent the exemplary approach to solve the problem, but there is a whole class of different approaches that can give different results. In the first method all possible combinations of inputs are considered, neural models are trained, compared, and finally the best model is selected which produces the best results. The second method is based on the elimination of inputs approach, which are most correlated with the rest of the inputs. The numerical results are given only for the test data set, the training data set is used only for training.

A. The combination method

Table II shows the results of the conducted experiments. It is interesting that the best results are achieved with 5 inputs of neural network, which included the stochastic oscillator %K %D, oscillator MACD, RSI index and backward regressions 5 and 10 days. The best result of the neural network for the 5 inputs is 67.42% accurate decisions on positive points. When the neural network suggests that analyzed point of interest is the negative point, the investment decision is the omission of buying the shares. Otherwise, the neural network over 67% of the time classified as a positive point of interest properly. The result is much better than the result of the trivial strategy.

In the case of neural networks operating with all the inputs indicated by the investor, the result is clearly worse but still higher than the result of a trivial strategy. Therefore one can conclude that the two indices that analyze the investors, in fact, do not bring valuable information and even obscure the decision-making situation. It is likely that the CCI indicators and backward linear regression for 15 days carry conflicting information from other indicators, which makes the classifier more was wrong, when all indicators are taken as inputs.

Comparison of the results of the neural network and trivial strategy is shown in Fig. 9. It is worth noting that the neural networks of the worst matched inputs achieve results far worse than trivial strategy and is approximately 40%. It is an experimental proof that a key consideration is the choice of set of inputs, and that from the proposed inputs by investors there are combination which gives very bad results.

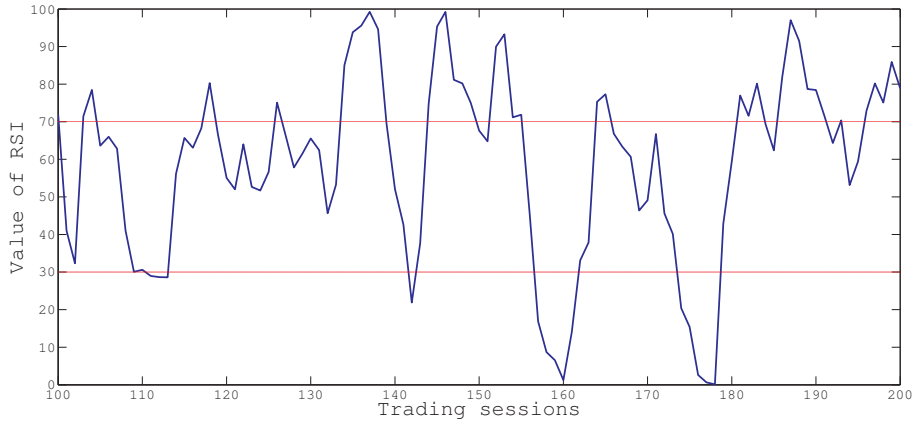


Fig. 6. The example Relative Strength Index (RSI)

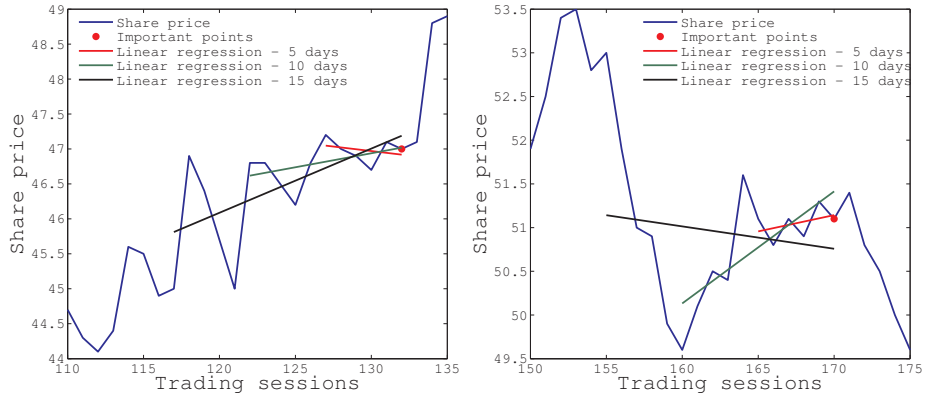


Fig. 7. The backward linear regression for 5, 10 and 15 days

TABLE II
THE RESULTS OF THE NEURAL NETWORK FOR THE BEST COMBINATION OF INPUTS

Number of inputs	Best inputs	Positive points found	Good investments (%)
1	4	312	60.00%
2	2 7	388	60.95%
3	2 4 6	377	63.33%
4	2 5 6 7	368	63.57%
5	1 2 4 5 6	300	67.42%
6	1 2 3 4 5 6	329	65.93%
7	1 2 3 4 5 6 7	377	62.29%

B. The correlation method

Although in the combination method the best result can be found, as many as 1270 neural networks are learned, which takes a lot of time.

The correlation method of selecting the inputs of the neural network consists of eliminating inputs that are most correlated with other inputs. This method is based on the belief that inputs that are correlated with each other can be removed, and the other inputs take over the role of the removed inputs. Following this rule it is sufficient to examine the correlation between all inputs, and then eliminate the ones that carry the most common information with other inputs. The optimal set of inputs for each number of inputs is given in Table III. One may conclude that the achieved results are not optimal. It should be noted, however, that the best combination of inputs determined by this method differs a little from the optimum. The process of elimination gives the best result of the neural network with 5 inputs, which corresponds to the best global result. Selection of inputs is also very similar, because only one indicator is only different.

Computing effort of the second method is much smaller than searching all possible combinations of inputs of neural networks. Fig. 10 shows the results of this method against the results of a trivial strategy. Referring to the results of

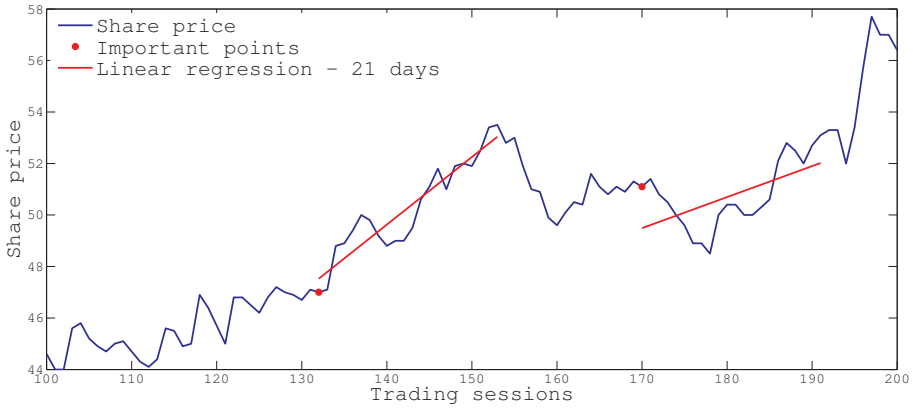


Fig. 8. The output of neural network – 21 days forward linear regression

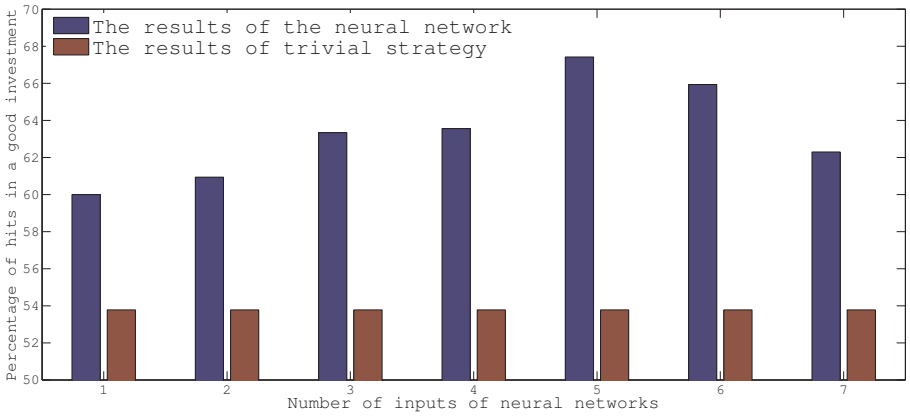


Fig. 9. The percentage of good results depending on the number of inputs of the neural network

TABLE III
THE RESULTS OF NEURAL NETWORK THE LEAST CORRELATED INPUTS COMBINATIONS

Number of inputs	Best inputs	Positive points found	Good investments (%)
2	1 4	338	60.14%
3	1 4 6	303	60.00%
4	1 3 4 6	357	60.61%
5	1 3 4 5 6	279	62.48%
6	1 3 4 5 6 7	373	59.39%
7	1 2 3 4 5 6 7	377	62.29%

the strategy proposed by the investors one must concede that the results obtained by the method of elimination of most correlated inputs are a little better from the results when all indicators are used, as suggested by the investors.

C. Comparison of methods

To compare the two methods used for the selection of neural network inputs it is necessary to look into the results from two perspectives. First, the combination method, of course, gives the best results. The method of eliminating most correlated inputs also brings good results, but it fails to find solutions much better than the approach suggested of investors. Secondly, considering the computational burden necessary to carry out all experiments, one can see the undoubted advantage of the second approach. Both methods brought similar results, so it can be concluded that the impact on the result of the trend predictor is associated with the selection of uncorrelated inputs that carry a consistent information. Fig. 11 shows a comparison of the results of both methods.

VI. SUMMARY

This paper describes the development of the neural investment strategy system. In order to determine the best set of the model inputs the two methods are compared: the neural networks for all possible combinations of inputs can be evaluated or the elimination method can be used. The first

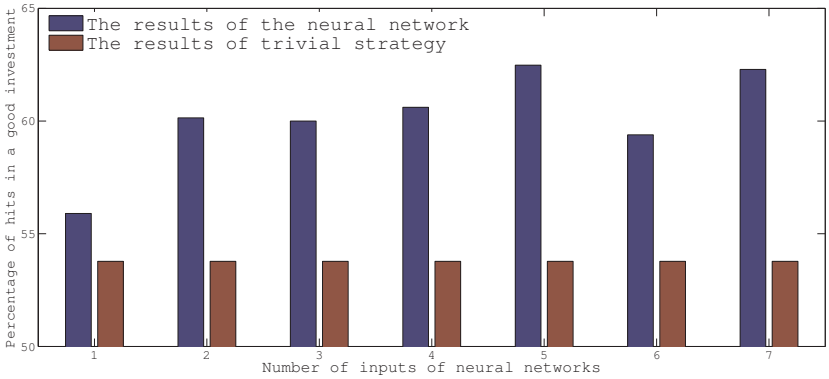


Fig. 10. The percentage of good results in correlation method

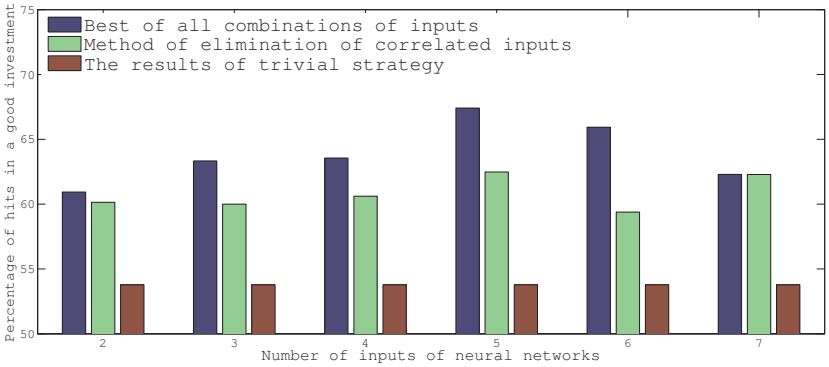


Fig. 11. Comparison of methods for the selection of inputs of neural network

method makes it possible to find the best model, but is very time consuming since as many as 1270 neural networks must be trained. The second method is much more effective and practically gives the neural model of the same accuracy.

It is necessary to emphasize that the discussed investment strategy is based on both the economic knowledge and intuition. The neural predictors are not entirely "black-box" time-series models, but technical analysis indicators are used as the input signals. The obtained experiments shows that the selected indicators make it possible to obtain 62% of investment efficiency. When the number of indicators is reduced, the neural system is able to realize investment decisions of 67% success rate, which indicates that the redundant indicators could give wrong signals to investors.

Acknowledgement. The work presented in this paper was supported by Polish national budget funds for science.

REFERENCES

[1] Bishop, C. M.: Neural networks for pattern recognition. Oxford University Press. Oxford. (1995)
[2] Cichocki, A.: Neural networks for singular value decomposition. Electronic Letters 28, 784–7860 (1992)
[3] Elder, A.: Trading for a Living: Psychology, Trading Tactics, Money Management. John Wiley & Sons. New York. (1993)
[4] Gately, E.: Neural networks for financial forecasting. Wiley. New York (1996)
[5] Haykin, S.: Neural networks – a comprehensive foundation. John Wiley & Sons. New York. (1993)
[6] Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366 (1989)
[7] Hudson, D. L., Cohen, M. E.: Neural networks and artificial intelligence for biomedical engineering. IEEE Press series on Biomedical Engineering (1999)
[8] Korbicz, J., Kościelny J. M., Kowalczuk, Z., Cholewa, W. (editors): Fault diagnosis: models, artificial intelkligence, applications. Springer, Berlin (2004)
[9] Liu, S., Wang, J.: A simplified dual neural network for quadratic programming with its KWTA application. IEEE Transactions on Neural Networks 17, 1500–1510 (2006)
[10] Murphy J.: Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York. (1999)
[11] Ortega, J. G., Camacho, E. F.: Mobile robot navigation in a partially structured static environment, using neural predictive control. Control Engineering Practice 4, 1669–1679 (1996)
[12] Ripley, B. D.: Pattern recognition and neural networks. Cambridge University Press. Cambridge. (1996)
[13] Siwek, K. Osowski, S., Szupiluk, R.: Ensemble neural network approach for accurate load forecasting in a power system. International Journal of Applied Mathematics and Computer Science 19, 303–315 (2009)
[14] Tatjewski, P.: Advanced control of industrial processes, Structures and algorithms. Springer, London (2007)
[15] Tatjewski, P. Ławryńczuk, M.: Soft computing in model-based predictive control. International Journal of Applied Mathematics and Computer Science 16, 101–120 (2006)
[16] Wong, B. K., Selvi, Y.: Neural network applications in finance: a review and analysis of literature (1990–1996). Information and Management 34, 129–139 (1998)

International Workshop on Artificial Intelligence in Medical Applications

THE workshop on Artificial Intelligence in Medical Applications – AIMA'2013 - provides an interdisciplinary forum for researchers and developers to present and discuss latest advances in research work as well as prototyped or fielded systems of applications of Artificial Intelligence in the wide and heterogenious field of medicine, health care and surgery. The workshop covers the whole range of theoretical and practical aspects, technologies and systems based on Artificial Intelligence in the medical domain and aims to bring together specialists for exchanging ideas and promote fruitful discussions.

The topics of interest include, but are not limited to:

- Artificial Intelligence Techniques in Health Sciences
- Knowledge Management of Medical Data
- Data Mining and Knowledge Discovery in Medicine
- Health Care Information Systems
- Clinical Information Systems
- Agent Oriented Techniques in Medicine
- Medical Image Processing and Techniques
- Medical Expert Systems
- Diagnoses and Therapy Support Systems
- Biomedical Applications
- Applications of AI in Health Care and Surgery Systems
- Machine Learning-based Medical Systems
- Medical Data- and Knowledge Bases
- Neural Networks in Medicine
- Ontology and Medical Information
- Social Aspects of AI in Medicine
- Medical Signal and Image Processing and Techniques
- Ambient Intelligence and Pervasive Computing in Medicine and Health Care

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 193–198

Automatic computer aided segmentation for liver and hepatic lesions using hybrid segmentations techniques

Ahmed M. Anter*, Ahmad Taher Azar†, Aboul Ella Hassanien ‡, Nashwa El-Bendary§, Mohamed Abu ElSoud ¶
*Faculty of Computers and Information, Computer Science Department, Mansoura University, Egypt.
Email: sw_anter@yahoo.com
†Faculty of Computers and Information, Benha University, Egypt,
Scientific Research Group in Egypt (SRGE), Egypt. Email: ahmad_t_azar@ieee.org
‡Faculty of Computers and Information, Computer Science Department - Cairo University,
Scientific Research Group in Egypt (SRGE) Email: aboitcairo@gmail.com
§Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
Scientific Research Group in Egypt (SRGE), Egypt
¶Faculty of Computers and Information, Computer Science Department, Mansoura University, Egypt

***Abstract*—Liver cancer is one of the major death factors in the world. Transplantation and tumor resection are two main therapies in common clinical practice. Both tasks need image assisted planning and quantitative evaluations. An efficient and effective automatic liver segmentation is required for corresponding quantitative evaluations. Computed Tomography (CT) is highly accurate for liver cancer diagnosis. Manual identification of hepatic lesions done by trained physicians is a time-consuming task and can be subjective depending on the skill, expertise and experience of the physician. Computer aided segmentation of CT images would thus be a great step forward to scientific advancement for medical purposes. The sophisticated hybrid system was proposed in this paper which is capable to segment liver from abdominal CT and detect hepatic lesions automatically. The proposed system based on two different datasets and experimental results show that the proposed system robust, fastest and effectively detect the presence of lesions in the liver, count the distinctly identifiable lesions and compute the area of liver affected as tumors lesion, and provided good quality results, which could segment liver and extract lesions from abdominal CT in less than 0.15 s/slice.**

I. Introduction

THE LIVER cancer is one of the most common internal malignancies worldwide and also one of the leading death causes. Early detection and accurate staging of liver cancer is an important issue in practical radiology. Liver lesions are a wound or injury to body tissues. It is the area of tissue that caused damage because a wounding or disease. Liver lesions refer to those abnormal tissues that are found in the liver. In a CT scan these can be identified by a difference in pixel intensity from that of the liver. Manual segmentation of this CT scans are tedious and prohibitively time-consuming for a clinical setting. Automatic segmentation on the other hand, is a very challenging task, due to various factors, such as liver stretch over 150 slices in a CT image, indefinite shape of the lesions and low intensity contrast between lesions and similar to those of nearby tissues. The irregularity in the liver shape and size between the patients and the similarity with other organs of almost same intensity make automatic liver segmentation difficult [1, 2].

Several studies have developed various algorithms that can be categorized on the degree of automation (fully, semi or interactive) and in two approaches: region-based or contour-based. Region-based segmentation is commonly based on intensity of neighbour pixels. While contour-based segmentation includes geometrical or statistical active shape model. Each of these approaches has its advantages and disadvantages in terms of applicability, suitability, performance, and computational cost [3,4].

Particularly, no one who did not consider above characteristics of the abdominal CT image can meet desirable results on liver segmentation. In addition, the traditional method of getting volume of the liver is to perform a by-hand 2D segmentation of parallel cross-sectional CT slices and to multiply all voxels of the stacked slices by their size while the procedure is often time consuming and non-systematic [5].

Therefore, to address the above mentioned problems, we present fully automatic liver segmentation and detection algorithms in abdominal CT images based on a hybrid approach using an adaptive threshold, morphological operators and Connected Component Labelling algorithm (CCL) to segment liver parenchyma from abdominal CT and Watershed algorithm coupled with Region Growing algorithm to extract lesions from liver parenchyma. The hybrid system is proposed to improve the segmentation performance and time consuming compared with the conventional process. The proposed approach starts with search for suitable abdominal liver CT image from DICOM file, and then this suitable image is passed to a filter to enhance and remove noise, and finally passing to segmentation algorithm to segment the whole liver then passed to hybrid segmentation system to extract hepatic lesions.

978-83-60810-53-8/$25.00

The process of segmentation is done in two phases. The first phase aimed to segment liver parenchyma from abdominal CT, this phase consist of three steps. The first step is to convert CT image into binary image using adaptive threshold that examine the intensity values of the local neighbourhood of each pixel. The second step is to apply multi-scale morphological operators to filter tissues nearby liver, to preserve the liver structure and remove the fragments of other organs. The third step is a CCL algorithm to remove small objects and false positive regions. The second phase aimed to segment and extract lesions from liver parenchyma which is segmented in first phase, in this phase integration between segmentation algorithms was applied to boost and increase the efficiency of segmentation behaviour. Marker-controller watershed algorithm was applied to cluster liver and define ROI, after liver clustered using watershed, an adaptive region growing is integrated with watershed algorithm to increase the performance and accuracy of segmentation. This system was tested on two different datasets. Good results were obtained in terms of quality and less processing time of the segmentation operation.

The reminder of this paper is ordered as follows. Section II discusses the previous work on liver segmentation. Details of the proposed methods and datasets are given in Section III. The proposed system is presented in Section IV. Section V shows the experimental results and analysis. Finally, Conclusion and future work are discussed in Section VI.

II. Previous Work

Several researchers have focused their attention on the use of threshold to segment liver. Massoptier and Casciaro [6] used adaptive thresholding to detect livers and refined the segmentations by graph cut. Campadelli et al. [7] detected livers by using heart segmentation information and then used adaptive thresholding and morphology as an alternative to graph cut to segment livers. Masumoto et al. [8] utilized conventional thresholding in multi-phase images to delineate the liver. Rusko et al. [9] mainly used region-growing with various pre-processing and post-processing steps to segment liver. Extracting regions of interest (ROIs) requires a sharpening filter to stress the regions edges [10]. Kumar and Moni [11] proposed their thresholding and morphological operator based algorithm to segment liver from abdominal CT image slices. Susomboon et al. [12] proposed a hybrid approach consisted of intensity based partition, region-based texture classification, connected component analysis and thresholding for liver segmentation. Massieh et al. [13] proposed an automatic region growing method that incorporates fuzzy c-means clustering algorithm to find the threshold value and modified region growing algorithm to find seed point automatically. However, their approach is very time consuming. In contrast with active shape models, Seghers et al. [14] incorporated both local intensity and local shape models for liver segmentation. Wan Nural and Hans Burkhardt [15] used Integration of Morphology and Graph-based Techniques for liver segmentation. Abdalla et al. [16], proposed new combined approach level set and watershed approach for CT liver segmentation to separate the liver from other organs and obtained overall accuracy of 92%. Shweta and Sumit [17] proposed level set segmentation technique using Variational Level Set Formulation techniques without initialization with various filtering methods. It was found that maximum filter provided the best results on the samples of the segmentation of CT images.

Jeongjin et al. [18] applied two steps of seeded region growing onto level-set speed images to define liver region. Ruchaneewan etal. [19] used intensity-based partition and region-based texture to segment liver. Abdalla et al. [10] Proposed for segment and isolate the liver region of interest using a region growing segmentation approach, and achieved highest performance for contrast stretching filter.

III. Martials and Methods

CT scanning is a diagnostic imaging procedure that uses X-rays in order to present cross-sectional images ("slices") of the body. The proposed system will be work on two different datasets: First dataset has divided into seven categories depends on the tumour type of benign (Cyst (CY), Hemangioma (HG), Hepatic adenoma (HA), and Focal nodular hyperplasia (FNH)) or malignant (hepatocellular carcinoma (HCC), Cholangiocarcinoma (CC), and Metastases (MS)), each of these categories have more than fifteen patients, each patient has more than one hundred slices, and each patient has more than one phases of CT scan (arterial, delayed, portal venous, non-contrast), also this dataset has a report diagnosis for each patient. All images are in JPEG Format selected from DICOM file and have Dimensions 630 x 630 with horizontal and vertical resolution of 72 DPI and bit depth 24 bit. All CT images captured from Radiopaedia web site [19].

For the second dataset, the data is acquired on a GE Discovery ST with the breathing trace provided by a Varian RPM system, and processed by a Varian 4D workstation. Information found in series description DICOM tag [0008,103E], T=0% is end-inhale and T=50% end-exhale. Livers and liver tumours CT images are manually segmented by five expert radiologists. These datasets are provided by Dr. Kevin Cleary at the Imaging Science and Information Systems (ISIS) Center from the Georgetown University Medical Centre [20].

A. Pre-Processing

The main objective of image pre-processing is to enhance, smoothness, remove noise that caused by defects of CT scanner, improve quality and emphasizes certain features of an image so that it makes segmentation or recognition easier and more effective. Filtering is a key pre-processing technique used for various effects including contrast stretching, sharpening and smoothing. In this paper, the effective filtering techniques were evaluated and analysed to modify, smooth the image and to enhance the efficiency of proposed algorithm. Pre-processing of Liver CT's are typically aimed at either improvement of the overall visibility of features or enhancement of a specific sign of malignancy also morphological operators

based algorithm is sensitive to noise, for these reasons pre-processing and filters is very important for liver images.

B. Liver segmentation methods

Segmentation of the liver and hepatic lesions from abdominal CT image is difficult. Therefore, a system is developed to extract liver and lesions automatically with sophisticated hybrid technique. To achieve the segmentation process, the following methods was proposed:

1) Adaptive thresholding Technique

Global thresholding, local adaptive thresholding are used to separate the desirable foreground image objects from the background based on the difference in pixel intensities of each region. Global thresholding uses a fixed threshold for all pixels in the image and therefore works only if the intensity histogram of the input image contains neatly separated peaks corresponding to the desired subject(s) and background(s). Hence, it cannot deal with images containing, a strong illumination gradient. Local adaptive thresholding, on the other hand, selects an individual threshold for each pixel based on the range of intensity values in its mean of local neighbourhood. This allows for thresholding of an image whose global intensity histogram doesn't contain distinctive peaks. Adaptive thresholding is more sophisticated and accommodate changing lighting conditions in the image. This approach is used for finding the local threshold to statistically examine the intensity values of the local neighbourhood of each pixel. This method is simple, fast and less computationally intensive and produces good results for CT liver images.

2) Morphological Operator-based Algorithm

A morphological processing is an obvious choice to refine the segmentation. The main idea of morphological operators is to detect the object forms or shapes from the images based on a set of pre-defined structuring elements. Usually a set of structuring elements (SE) is based on the prior knowledge, and then some morphological operators apply structuring elements to images [21]. Dilation and erosion are the two main morphological processing. Dilation expands objects by a structuring element, filling holes, and connecting disjoint regions. Erosion deletes the small region by a structuring element. Morphological operations based algorithm has several advantages. First it does not need any specific initialization, which makes it possible to design the fully-automatic algorithms. Second it focuses less on the structure of the object of interest. Therefore, it can work well on the liver whose structure varies between different persons.

3) Connected Component Labeling algorithm(CCL)

CCL works by scanning a binary image pixel by pixel (from top to bottom and left to right) in order to identify connected pixel regions [22]. The result of applying an adaptive threshold is a collection of different regions, applying morphological operators to preserve the liver structure and remove the fragments of other organs, but still some regions not interested to be liver will be removed by post-processing approach CCL. The largest region extracted by using CCL, it is used to label the separate regions in CT, yielding a new labelled image. In general, this algorithm is useful to find non-connected objects in images.

C. Lesions segmentation methods

In this phase we used two different methods to extract lesions. Watershed algorithm as edge-based image Segmentation and Region Growing (RG) algorithm as region-based image Segmentation.

1) Watershed Algorithm

Separating lesions from liver image is one of the more difficult processing operations. The watershed transform is often applied to this problem. Watershed image segmentation can be regarded as an image in three dimensions (two spatial coordinates versus intensity). We will use three types of point which "minimum", "catchment basin", and "watershed line" to express a topographic interpretation. Watershed algorithm has an advantage that it is fast speed. While disadvantages of this algorithm is over-segmentation results, to solve this problem used marker-controlled for watershed Segmentation.

The watershed marker finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low. Segmentation using the watershed marker works better if you can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:

1. Use a smoothing filter to pre-process the original image, then the action can minimize the large numbers of small spatial details.

2. Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.

3. Compute foreground markers. These are connected blobs of pixels within each of the objects.

4. Compute background markers. These are pixels that are not part of any object.

5. Modify the segmentation function so that it only has minima at the foreground and background marker locations.

6. Compute the watershed transform of the modified segmentation function.

2) Region Growing Algorithm

The region growing (RG) algorithm is one of the simplest region-based segmentation methods. It performs a segmentation of an image with examine the neighboring pixels of a set of points, known as seed points, and determine whether the pixels could be classified to the cluster of seed point or not [22].The advantages of this algorithm is simplest, can correctly separate the regions of same properties, give good shape matching of its results. The algorithm procedure is as follows.

Step1. Start with a number of clusters and seed points which have been identified from watershed algorithm, cluster called C_1, C_2,...., C_n. And the positions of initial seed points is set as P_1, P_2,....., P_n.

Step2. To compute the difference of pixel value of the initial seed point pi and its neighboring points, if the difference is

smaller than the threshold criterion that define, the neighboring point could be classified into C_i, where i = 1, 2,....,n.

Step3. Recompute the boundary of C_i and set those boundary points as new seed points pi (s). In addition, the mean pixel values of C_i have to be recomputed, respectively.

Step4. Repeat Step 2 and 3 until all pixels in image have been allocated to a suitable cluster.

The mean drawback of RG is initial seed-points. The initial seed-points problem means the different sets of initial seed points cause different segmentation results. This problem reduces the stability of segmentation results from the same image. Furthermore, how many seed points should be initially decided is an important issue because various images have individually suitable segmentation number. These problems will be handled in this paper by integrated RG with watershed algorithm.

IV. Proposed System

The proposed fully automatic technique and methods to segment liver structure and lesions from abdominal CT divided into two phases liver structure segmentation and lesions segmentation. The first phase of liver parenchyma segmentation from abdominal CT is comprised of five fundamental building steps as seen in Figure 1. The first step searches for suitable slices in CT DICOM file because liver intensity distribution is different between slices. Liver parenchyma is the largest abdominal object in middle slices. These slices are suitable for segmentation and give high accuracy.

Pre-processing step: In this step pre-processing algorithm is used before the segmentation phase to enhance contrast, remove noise and emphasize certain features that affect segmentation algorithms and morphology operators.

Adaptive threshold step: In this step CT image is converted into binary image using adaptive threshold method that examines the intensity values of the local neighbourhood of each pixel.

Morphological Operators step: After the CT image is converted into binary image using adaptive threshold, morphological operators will be applied to filter tissues nearby liver, to preserve the liver structure and remove the fragments of other organs.

Connected Component Labeling phase: CCL algorithm is used to remove small objects, false positive regions and focused on liver parenchyma.

The second phase of liver lesions segmentation and extraction aimed to integrate between watershed algorithm and RG to boost and increase the performance of segmentation. Watershed used to segment liver into different regions and solving the problem of over-segmentation using watershed marker. RG used to improve watershed segmentation using the clusters and centroid point for each cluster in watershed as seed point for RG.

V. Experimental Results

Hybrid system was used to segment liver structure from abdominal CT. The reason to do these hybrid methods was that each method as such has problems, which the other method does not have. The proposed hybrid system divided into two phases. In the first phase suitable abdominal CT slice image of a patient with liver lesions was selected from DICOM file. Liver parenchyma is the largest abdominal object in middle slices as shown in Figure 2(a). Pre-processing median filter is used to enhance, remove noise and emphasize certain features that affect segmentation algorithms and morphology operators with 3x3 window as shown in Figure 2(b). After pre-processing stage, the hybrid segmentation based on adaptive threshold algorithm is applied on enhanced abdominal CT as shown in Figure 2(c). The adaptive threshold used the mean of the local intensity distribution to decide whether a pixel belongs to an organ of interest based on its neighboring features. Output is a binary image with the mean of local threshold. This mean of the local area is not suitable as a threshold, because the range of intensity values within a local neighborhood is very small and their mean is close to the value of the center pixel. The quality of adaptive threshold was improved by using static coefficient factor for all slices to increase performance. This method is simple, fast, less computationally intensive and produces good results for all slices.

After applying adaptive threshold, the morphological erosion and dilation operator with the shape and size of structuring element (SE) was used to shrink and remove small regions and extract liver from abdominal CT as shown in Figure 2(d). The experimental results show that the best shape

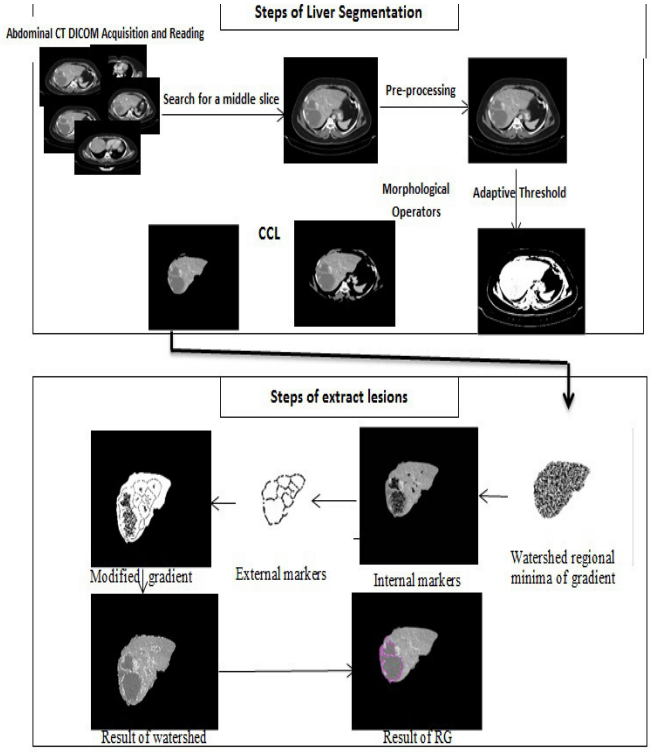


Fig. 1. Architecture of the proposed automatic liver segmentation approach

TABLE I
COMPARISON WITH EXISTING WORK ON LIVER SEGMENTATION

Author	Yearn	Patients	Slices	accuracy
Jeongjin et al. [18]	2007	20	—	0.70
Ruchaneewan et al.[12]	2007	20	30	0.86
Toshiyuki et al. [23]	2008	28	159	0.89
Abdalla et al. [10]	2012	—	26	0.84
Abdalla Z. et al. [16]	2012	4	27	0.92
Proposed	2013	112	860	0.93

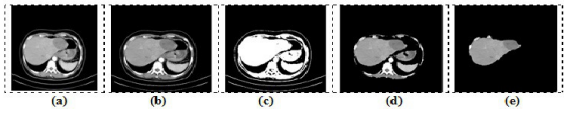


Fig. 2. Results liver segmentation on different patient's slices, a) The original suitable Slice, b) pre-processing median filter, c) Automatic adaptive threshold for each slice, d) Operators Dilation and erosion Morphology, e) The final results for ROI selected by CCL

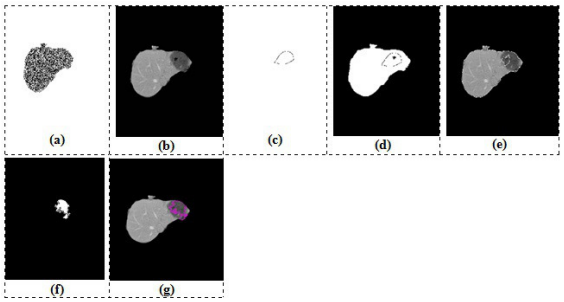


Fig. 3. Results of lesions segmentation, a) Watershed over-segmentation result, b) Internal Markers, c) External Markers, d) Modified watershed, e) Output watershed segmentation, f) Automatic seeded point for RG, g) Final result from RG segmentation

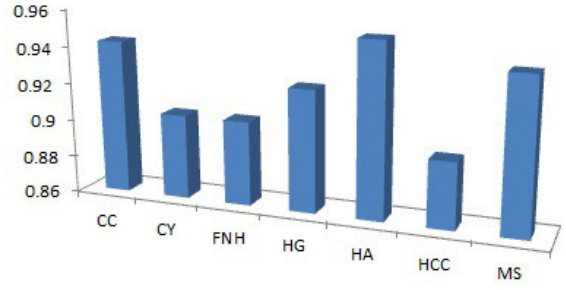


Fig. 4. Segmentation accuracy of livers with hepatic lesions

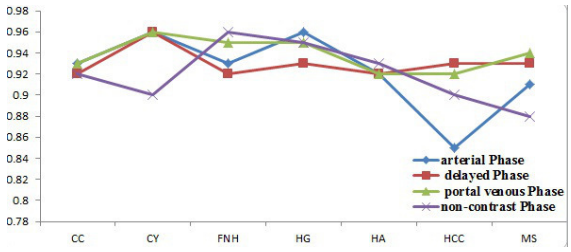


Fig. 5. Segmentation accuracy of abdominal CT phases for hepatic lesions

is diamond with SE size value 4. The shape and size of SE is decided after analyzing many Liver CT's. After applying morphological operators, post-processing CCL is applied on adaptive threshold with 8-connected objects to search for the largest connected region, remove false positive regions and focus on the ROI as shown in Figure 2(e).

In the second phase, after liver parenchyma segmentation passed to watershed to segment liver into distinct regions. Watershed approach combines the edge detection and region growing approaches, producing more stable results and connected boundaries. The main idea in watershed approach is to check whether one point belongs to one minimal, then it merges the point to it. Otherwise, the pixel is considered a boundary between the two minimal. This is done in a binary image using the morphological dilation. A huge number of potential minima of small objects in an image lead to over-segmenting problem. A smoothing filter should be used to eliminate that huge number. The simulation result of this algorithm has an advantage that it is fast speed. At the same time, it has a critical over-segmented problem, to solve this problem we use internal and external marker control to segment objects with closed contours, expressing the boundaries as ridges. Then pass this segmented liver to RG. The RG algorithm is used to improve watershed segmentation using the centroid point for each segmented region from watershed as seed point for RG. The region growing method had basically three problems, the rugged border, seeded point very difficult to assign it automatically for ROI and the leakage problem, but when combined with watershed gives accurate results as shown in Figure 3. The performance and accuracy of the proposed system was evaluated by Similarity Index technique (SI) between automated segmented images and manual segmented images.

The proposed hybrid system applied on 112 patients from different datasets with different hepatic lesions. The overall accuracy obtained is 0.93 for livers segmentation, and the proposed system applied on 860 abdominal CT slices achieved overall accuracy result 0.90. The proposed approach also gives acceptable accuracy of livers segmented with hepatic lesions are 0.94, 0.91, 0.91, 0.93, 0.95, 0.90, and 0.94 for CC, CY, FNH, HG, HA, HCC, and MS respectively as shown in Figure 4. In Figure 5, the proposed system also measures the segmentation accuracy of abdominal CT phases (arterial, delayed, portal venous, and non-contrast) for hepatic lesions. The better segmentation accuracy was achieved in portal venous phase, while arterial phase gives non accurate results in Hepatocelleour Carcinoma. This is because the nature of livers tissue which are ambiguous in this phase.

Comparing the results of proposed system with other previous work on CT liver segmentation from abdominal CT as shown in Table 1. The proposed approach is fast, precise, robust and provides good quality results and there is no meaningful loss of information, which could segment liver and extract lesions from abdominal CT in less than 0.15 s/slice

when implemented in Matlab on PC Intel Core I5, 2.5GHz with 4GBytes of RAM.

VI. Conclusion and Future Work

The presented system for segmentation and liver lesions extraction is able to reliably segment and extract the lesions in the used patient database. Liver segmentation is a complicated process which consists of many steps for segmentation process. Integration between edge-based segmentation and region-based segmentation methods give more reliable and trust segmentation. The experimental results show that it is a robust proposed algorithm and obtained 93% of good extraction for liver from abdominal CT.

In conclusion, our results suggest that ensemble segmentation is effective in segmentation of livers and liver lesions, boosting and increasing the performance of weak segmentation processes.

In future work, we plan to assess the performance using a large dataset to evaluate generalization performance of the algorithm that includes a number of parameters in the feature measurement process, which means it might sensitive to size and characteristics of lesions.

References

[1] L. Seong-Jae, J. Yong-Yeon, H. Yo-Sung, "Automatic liver segmentation for volume measurement in CT Images ", Elsevier, J. Vis. Commun. Image R. 17 860–875, 2006.

[2] K. Suzuki, R. Kohlbrenner, M. L. Epstein, A. M. Obajuluwa, J. Xu, and M. Hori, "Computer-aided measurement of liver volumes in CT by means of geodesic active contour segmentation coupled with level-set algorithms", Med Phys. 37(5): 2159–2166, 26 April 2010.

[3] H. F. Amir, A. Z. Reza, H. Masatoshi, S. Yoshinobu,"A knowledge-based technique for liver segmentation in CT data", Computerized Medical Imaging and Graphics, vol. 33, 8, pp. 567–58, Dec. 2009.

[4] A. Militzer, T. Hager, F. Jager, C. Tietjen, J. Hornegger, "Automatic detection and segmentation of focal liver lesions in contrast enhanced CT images", Proc. Int. Conf. on pattern recognition, vol. 10, pp 2524–2527, 2009.

[5] T. Saitoh, Y. Tamura, T. Kaneko, "Automatic segmentation of liver region based on extracted blood vessels". Syst Comput Jpn, 35 (5), pp. 1–10, 2004.

[6] L. Massoptier, S. Casciaro, "Fully automatic liver segmentation through graph-cut technique", Proceedings of the 29th Annual International Conference, IEEE EMBS, 2007.

[7] C. Paola, C. Elena, L. Gabriele, "Automatic liver segmentation from abdominal CT scans", IEEE Computer Society Washington, DC, USA, PP. 731–736, 2007.

[8] J. Masumoto, M. Hori, Y. Sato, T. Murakami, T. Johkoh, H. Nakamura, "Automated liver segmentation using multislice CT images", Syst Comput Jpn, 34 (9), pp. 71–82, 2003.

[9] R. Laszlo, B. Gyorgy, F. Marta, "Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images", Medical Image Analysis, Science Direct, Vol. 13, 6, PP. 871–882, Dec., 2009.

[10] M. Abdalla, H. Hesham, I. G. Neven, H. Aboul Ella, S. Gerald, "Evaluating the Effects of Image Filters in CT Liver CAD System", In proceeding of: IEEE-EMBS International Conference on Biomedical and Health Informatics, The Chinese University of Hong Kong, Hong Kong, 2012.

[11] S.S. Kumar, R.S. Moni, "Diagnosis of Liver Tumor from CT Images Using Fast Discrete Curvelet Transform", Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications, IJCA, CASCT, 2010.

[12] S. Ruchaneewan, S.R. Daniela, and F. Jacob, "A Hybrid Approach for Liver Segmentation", Intelligent Multimedia Processing Laboratory, 3D Segmentation in The Clinic: A Grand Challenge, pp. 151–160, 2007.

[13] A. massieh, N. Hadhoud, M. Amin, "A novel fully automatic technique for liver tumor segmentation from CT scans with knowledge-based constraints", Intelligent Systems Design and Applications 10th International Conference on, vol., no.,pp.1253–1258, 2010.

[14] S. Dieter, S. Pieter, L. Yves, H. Jeroen, L. Dirk, M. Frederik, and S. Paul, "Landmark based liver segmentation using local shape and local intensity models", 3D Segmentation in The Clinic: A Grand Challenge, pp. 135–142, 2007.

[15] W. Y. WanNural, B. Hans, "Integration of Morphology and Graph-based Techniques for Fully Automatic Liver Segmentation", Majlesi Journal of Electrical Engineering,Vol. 4, No. 3, Sept. 2010.

[16] Z. Abdalla, I. G. Neveen, H. Aboul Ella, and A. H. Hesham, "Level set-based CT liver image segmentation with watershed and artificial neural networks", HIS, IEEE, pp. 96–102, 2012.

[17] G. Shweta, K. Sumit, "Variational Level Set Formulation and Filtering Techniques on CT Images". International Journal of Engineering Science and Technology (IJEST), Vol. 4, No.07 July, 2012.

[18] L. Jeongjin, K. Namkug, L. Ho, B. Joon, J. Hyung, M. Yong, S. Yeong, K. Soo-Hong, "Efficient liver segmentation using a level-set method with optimal detection of the initial liver boundary from level-set speed images", Elsevier,computer methods and programs in biomedicine 88, 26–38, 2007.

[19] http://radiopaedia.org/search?q=CT&scope=all,18.3.20131:44PM

[20] http://insight-journal.org/midas/collection/view/38,18.3.20131:44PM

[21] L. S. Jae, J. Y. Yeon, H. Y. Sung, "Automatic liver segmentation for volume measurement in CT Images", Elsevier, J. Vis. Commun. Image 860–875, R. 17, 2006.

[22] M. A. ElSoud , A. M. Anter , "Automatic mammogram segmentation and computer aided diagnoses for breast tissue density according to BIRADS dictionary", Int. J. Computer Aided Engineering and Technology, Vol. 4, No. 2, pp.165–180, 2012.

[23] O. Toshiyuki, S. Ryuji, S. Yoshinobu, H. Masatoshi, Y. Keita, N. Masahiko, C. Yen-Wei, N. Hironobu, and T. Shinichi, "Automated Segmentation of the Liver from 3D CT Images Using Probabilistic Atlas and Multi-level Statistical Shape Model", Springer, N. Ayache, S. Ourselin, A. Maeder (Eds.): MICCAI, Part I, LNCS 4791, pp. 86–93, 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 199–205

An Improved Ant Colony System for Retinal Blood Vessel Segmentation

Ahmed Hamza Asad[1,*], Ahmad Taher Azar[2,*], Mohamed Mostafa M. Fouad[3,*]
Aboul Ella Hassanien[4,*]
[1]Department of Computer Sciences and Information, ISSR, Cairo University, Egypt
[2]Faculty of Computers and Information,Benha University, Egypt
[3]Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
[4]Faculty of Computers and Information, Cairo University, Egypt
*Scientifc Research Group in Egypt (SRGE)
http://www.egyptscience.net

***Abstract*—The diabetic retinopathy disease spreads diabetes on the retina vessels thus they lose blood supply that causes blindness in short time, so early detection of diabetes prevents blindness in more than 50% of cases. The early detection can be achieved by automatic segmentation of retinal blood vessels in retinal images which is two-class classification problem. This paper proposes two improvements in previous approach uses ant colony system for automatic segmentation of retinal blood vessels. The first improvement is done by adding new discriminant feature to the features pool used in classification. The second improvement is done by applying new heuristic function based on probability theory in the ant colony system instead of the old that based on Euclidean distance used before. The results of improvements are promising when applying the improved approach on STARE database of retinal images.**

I. Introduction

THE DIABETIC retinopathy is the most common cause of blindness worldwide since the blindness occurs as a result of retina death due to loss of blood supply by the widely spread diabetes on it [1]. The blindness brings significant costs both to individual and society. There are two types of diabetic retinopathy; the first type is the non-proliferative diabetic retinopathy where the capillaries of retina swell and interfere with normal vision. The second type is the proliferative diabetic retinopathy where the capillaries of retina shut down. In both types, the diabetic retinopathy usually leads to retina revascularization [2]. So the segmentation of blood vessels in retinal images is an important step in treatment of diabetic retinopathy. Also there are many other diseases are often diagnosed based on their changes on reflectivity, bifurcations and tortuosity of retinal blood vessels such as hypertension [3]. Retinal blood vessels segmentation is also the core stage in automated registration of two retinal blood vessels images of a certain patient to follow and diagnose his disease progress at different times [4]. The retinal blood vessels segmentation is a classification problem where each pixel in the field of view of retinal image is classified as vessel-like or non-vessel. The manual segmentation of retinal blood vessels is a long and tedious task which also requires training and skill. So, for the last two decades, the automated retinal blood vessels segmentation attracts a lot of research in the medical image processing area since it's the critical component of circulatory blood vessel analysis systems [5]. The Reliable automated retinal vessel extraction encounters several challenges [6]: "(1) The blood vessels have a wide range of widths from very large (15 pixels) to very small (3 pixels) and diffident bifurcations. (2) Various structures appear in retinal image, including the optic disc, fovea, exudates and pigment epithelium changes, which severely disrupt the automatic vessel extraction. (3) The narrow vessels with various local surroundings may appear as some elongated and disjoint spots, which are usually lost. (4) The vessels intensity contrast is weak and variant and the small vessels especially are overwhelmed by Gaussian-like noises".

This paper proposes two improvements of the previous approach [7] used for automatic segmentation of blood vessels in retinal images based on the ant colony system (ACS) [8]. The improvements are performed in two ways, first adding new discriminant feature to the features pool to be consisted of fifteen features and second applying new heuristic function in ACS. The features pool consists of features that are simple, fast in computation, needn't to be computed at multiple scales or orientations and highly discriminate between vessels and background in retinal images [9]. These features are based on gray-level of the green image of retina, gray-level of computed vessels-enhanced image and Hu moment-invariants [10]. Since the large number of computed features increases the classification complexity, time and reduces its accuracy, so feature selection is an essential step for successful classification because it removes irrelevant features and achieves less complex, more accurate and faster classification. In this paper, the correlation based feature selection heuristic (CFS) [11] is used and it reduced the features set from fifteen to the best four features set. The performance of this improved approach is evaluated on a publicly available STARE database of retinal images [12] for scientific research in terms of the sensitivity, specificity and accuracy. Thus it's the first paper that tests ACS performance on STARE database. The rest of this paper is organized as follows: Section II surveys the previous popular related work. Section III presents scientific background on

978-83-60810-53-8/$25.00

the used features, CFS and ACS. Section IV presents the proposed approach and its improvements. In section V, there are experimental results. Finally in Section VI, conclusions and directions for future research are presented.

II. Related Work

The automatic segmentation methods of retinal blood vessels are categorized into supervised and unsupervised. In this section, short survey of popular retinal blood vessels segmentation methods from two categories is presented. This short survey shows how these methods were developed in the last two decades. For the supervised methods, they depend on pixel classification into vessel class or non vessel class using a classifier previously trained on manually-labelled samples by ophthalmologists from two classes. So these methods depend on pre-classified data which mayn't be available in real life applications. Also there are significant differences between ophthalmologists themselves in delineation of blood vessels. The training makes these methods give better performance than unsupervised methods especially in healthy images. Staal et al. [13] used KNN-classifier with 27-D feature vector based on ridges information. Their method depends on extracting ridges in the image, forming line elements from ridges, assigning each pixel to nearest line to partition image into patches and computing features vector of each pixel based on its line and patch attributes. Then the feature vector reduced to those result in best class separability by sequential forward selection algorithm. Ricci and Pefetti [14] used 3-D feature vector consists of the inverted gray-level of green color plus the two maximum responses of two orthogonal line detectors rotated in twelve angles and their classifier was the support vector machine (SVM). Marin et al. [9] used 7-D feature vector consists of five features encode gray-level variation between pixel and its surroundings plus other two features based on Hu moment-invariants and their classifier was the neural network (NN). Fraz et al. [15] used 9-D feature vector consists of the inverted gray-level of green color, the sum of gradient orientation maps at three scales, sum of tophat transform responses in eight directions using linear structure element, the two maximum responses of two orthogonal line detectors rotated in twelve angles and the four maximum responses of Gabor filter rotated in ten angles at four scales. They used an ensemble classifier from two-hundred bagged and boosted decision trees.

For the unsupervised methods, they are classified into methods based on mathematical morphology, vessel tracking, matched filter, bio-inspired algorithms and active contour. For the methods based on mathematical morphology, they utilize the fact that retinal vessels have morphology of connected piecewise linear segments. The top-hat morphological transformation is widely used in blood vessels segmentation since it estimates the background of retinal image using morphological opening operation and the retinal vessels are better enhanced when subtracting this estimated background from original image. The advantages of mathematical morphology are the speed and noise resistance but its drawback is that it doesn't exploit the known shape of retinal vessel cross-section. Miri and Mahloojifar [16] used the fast discrete Curvelet transform (FDCT) for contrast enhancement and multi-structure morphological transformation for detection of retinal vessels edges. The false positive detections are pruned by morphological opening by reconstruction and length filtering. Fraz et al. [17] extracted the centerlines of retinal vessels using first-order derivative of Gaussian (FODOG) filter rotated in four orientations to detect retinal vessels in all directions. Then the shape and orientation maps of the retinal vessels are produced by applying morphological top-hat transform with linear structuring element at eight directions to emphasis vessels in all possible orientations followed by morphological bit plane slicing of gray-level image. The final vessels tree is reconstructed using detected centrelines and maps of shape and orientation.

For the methods based on vessel tracking, they work at single retinal vessel rather than entire retinal vasculature. Starting with initial set of pixels as seeds that are selected automatically or manually labelled, the trace of a vessel is done based on pixels local information by selecting the next candidate pixel in the retinal vessel from set of pixels that are closed to current pixel under consideration. The main advantage of these methods is providing information about single vessel such as its width, connectivity and branching. Their drawback is the need for good initial set of pixels to trace all retinal vessels or they may be missed especially at bifurcations and crossings. Kelvin et al. [18] initially determine sparse seed points along the vessel boundary and found the optimal contours connecting these points using Dijkstras algorithm. After that, they used cost function incorporates Frangis multiscale vesselness measure, vessel direction consistency, the edge evidence and the spatial and radius smoothness constraints into conventional Livewire framework to efficiently compute optimal vessels medial axes. Delibasis et al. [19] initialized the seeds pixels for vessel tracking using a multiscale vesselness filter and picked a random non-zero pixel as a seed. They utilize a parametric model that exploits the geometric properties of retinal vessel composed of a "stripe" and they defined a measure of match (MoM) which quantifies the similarity between the model and the given image. The vessel tracking is done by identifying the best matching strip with the vessel by using the seed point, strip orientation, strip width and the MoM. This method actively seeks vessel bifurcation, without user intervention.

For the methods based on matched filter, the matched filter is 2-D kernel convolved with the image to search for three features of retinal vessel in the image at unknown position and orientation. These features should be considered when designing a matched filter; (1) intensity profile of cross-section of a retinal vessel is approximated by Gaussian curve so the matched filter has Gaussian profile. (2) The retinal vessel has little curvature so it can be approximated by piecewise linear segments. (3) The retinal vessel diameter decrease as it moves outward from the optic disk. The kernel is rotated in multiple orientations to detect the all vessels in all directions so it takes more computation overhead. The response of matched filter is

high with retinal vessels that have the same standard deviation of Gaussian function modelled by matched filter so it may miss retinal vessels that have different profiles. The illumination variation in background and presence of pathologies increases false positive detections resulted by the matched filter. Al-Rawi et al. [20] applied exhaustive search based optimization on DRIVE retinal images database [13] to find optimal values of matched filter parameters such as size, standard deviation and threshold. Zhang et al [21] extended the matched filter by using two kernels; one based on Gaussian and another based on first derivative of Gaussian (FODOG) to filter out false positive detections resulted by matched filter such as non vessel edges which has high responses to both kernels while vessels has high responses only to the basic Gaussian-profiled matched filter.

For the methods based on bio-inspired algorithms, little work was done used ACS. Cinsdikici and Aydn [22] fused the results of ACS and matched filter using OR operator to construct the final segmentation result. Hooshyar and Khayati [23] used fuzzy ACS where the maximum eigenvalue of Hessian matrix at multiples scales and the maximum response of Gabor filter at multiple orientations are the features of each pixel. Asad et al. [7] used ACS standalone where the features of each pixel are simple because they are based on gray-level variations between it and its surroundings in green color of retinal image and the known Hu moment-invariants. These features don't need to be computed at multiple scales or orientations, so they are fast on computation.

For methods based on active contour, the active contour (snake) is initialized curve moves on the image under internal forces by the curve itself and external forces by the image data. Both types of forces are defined so that the snake fits to a desired feature in the image. The external forces are defined by a human user or supervising process. The active contour are used in edge detection, shape recognition and object tracking. The advantage of active contour for example in shape recognition is that it's independent and self-adapting so it conforms to any shape but its drawback is that it needs to be initialized close to the search target to avoid local minima. Espona et al. [24] initially mapped vessel creases using the multilevel set extrinsic curvature based on structure tensor. Next tracing of the optic nerve circumference is performed to initialize an active contour at the intersection of vessel creases with this circumference. The active contour is deformed influenced by external forces of vessel creases. Al-Diri et al. [25] initially used tramline filter to segment some pixels of likely vessel centerline then the segment growing algorithm initialized multiple vessels segments from these segmented pixels using pairs of active contours. The twins active contours conformed to vessels edges while the bifurcations and crossings are extended using the junction resolution algorithm.

III. Background

A. Features

The features pool consists of the gray-level of green channel of RGB retinal image (*green*), group of five features based on the green gray-level (*f1, f2, f3, f4, f5*), group of eight features based on Hu moment-invariants (*Hu1, Hu2, Hu3, Hu4, Hu5, Hu6, Hu7, Hu8*) and the gray-level of the vessels-enhanced image (*Ive*) [7]. Most of the vessels segmentation approaches extract and use the green color image of RGB retinal image for further processing since it has the best contrast between vessels and background so it's taken as feature. The five gray-level based features group is presented by Marin et al. [9] and its features describe the gray-level variation between vessel pixel and its surrounding. The Hu moment-invariants are best shape descriptors which are invariant to translation, scale and rotation change. So they are used by the second group of eight features to describe vessels have variant widths and angles. The vessels-enhanced image [9] is better enhancing blood vessels while removing the bright retinal structures as optic disc and exudates, so it's used for computing the group of eight Hu moment invariants based features and its gray-level (*Ive*) is the new added feature to the previous features pool as first improvement. These features are simple, better discriminate between vessel and non-vessel classes and needn't be computed at multiple scales or orientations. The features computation is more detailed in [7].

B. Correlation Based Feature Selection Heuristic

It's a heuristic approach for evaluating the worth or merit of a subset of features [11]. The main premise behind this selection method is that the features that are most effective for classification are those that are most highly correlated with the classes (intensifiers and dissipaters), and at the same time are least correlated with other features. The method is therefore used to choose a subset of features that best represent these qualities. The best individual feature based on the following merit metric:

$$M_s = \frac{k r_{cf}}{\sqrt{k + k(k-1) r_{ff}}} \tag{1}$$

where M_s is the heuristic merit of a features subset S containing k features, r_{cf} and r_{ff} are the average feature-class correlation and the average feature-feature inter-correlation respectively. The numerator gives an indication of how predictive a group of features are; the denominator of how much redundancy there's among them.

C. Ant Colony System

The ACS as meta-heuristic searching algorithm was first proposed by Dorigo et al. [8] for solving the travelling sales man problem (TSP). The ACS is based on simulating the foraging behaviour of real ants in nature. In nature when real ants are searching for foods, multiple ants are going out in random paths. As the ant is moving, it deposits a chemical substance which is called pheromone on its moving path for guiding other subsequent ants to its path. As the time goes,

the pheromone is evaporating. So as the path is shorter as its pheromone concentration remains more time and more other ants are attracted to it. Thus the shortest path is the only one which attracts other ants. ACS is more detailed in [7].

IV. PROPOSED APPROACH

After features selection process by CFS heuristic and determining its recommended features set for STARE databases *(f2, f5, Hu1, Ive)*, the proposed approach is applied to retinal images. It consists of four phases; preprocessing, features computation, ACS based segmentation and post-processing. In preprocessing phase, the green channel of RGB retinal image is extracted and its contrast is better enhanced by covering the whole range of intensity [0, 255] since it's a feature*(green)* and it's used in further processing. After that, the central light reflex which runs down the central length of some vessels is removed. In features computation phase, the varying illumination in background is corrected by computing the homogenized background image for better discriminating vessels from background and computing the gray-level based features *(f2, f5)*. Finally, the vessels-enhanced image is computed since its gray-level *(Ive)* is used for computing Hu moment-invariants based feature *(Hu1)* as well as it's selected feature by CFS. In The ACS based segmentation phase, each pixel is classified as vessel or background depending on its pheromone level τ and heuristic function η value. The old heuristic function η was the Euclidean distance between the target pixel and both centers of vessels and background classes in the feature space:

$$\eta = \frac{Euclidean\,distance\,to\,background\,class\,center}{Euclidean\,distance\,to\,vessels\,class\,center} \qquad (2)$$

The class center consists of averages of all selected features by CFS over all training pixels of the database. From the final pheromone map image τ, the vessels are segmented from background by thresholding. In post processing phase, linking of disjoint pixels is performed by setting pixel to 1 if it's surrounded at least by four neighbouring pixels of 1; otherwise it's sat to 0. All small regions have area less than 20 are filtered out. Finally, a median filter of size 3*3 eliminates all remaining isolated noisy pixels. Algorithm (I) shows the proposed approach which is more detailed in [7].

As second improvement, new heuristic function based on probability theory is applied to enhance the performance of proposed approach. The heuristic function is also computed from all training pixels of the database. So for probability calculation, the values of features in the feature space are transformed to specific interval and rounded to the nearest integers inside this interval. The following equation defines the new heuristic function η^*:

$$\eta^* = \frac{P(v/Vessels)}{P(v/Background)} \qquad (3)$$

where *P(v/Vessels)* and *P(v/Background)*are the likelihood of feature value *v* for vessels and background classes respectively.

Algorithm 1 PROPOSED APPROACH

```
/* Preprocessing Phase*/
Extraction of the green channel of retinal image
Linear transformation of its intensity to cover the whole intensity range [0, 255]
Remove of the central light reflex from it.
Computation of its homogenized-background image.
Computation of its vessels-enhanced image.
/* Features Computation Phase*/
for each pixel in the homogenized-background image do
    Compute its gray-level based features (f2,f5)
end for
for each pixel in the vessels-enhanced image do
    Compute its Hu moment-invariants based feature (Hu1)
end for
/* ACS based segmentation */
Initialize ACS parameters and the pheromone map image
Compute the heuristic function based on Eq.3 for each pixel in the image
Apply ACS operation
Threshold the resulted ACS pheromone map image to segment vessels from background
/* Post-processing */
for each pixel in the thresholded image do
    if surrounded by at least four neighboring pixels of 1 then
        Setting it to 1
    else
        Setting it to 0
    end if
end for
for each region in the thresholded image do
    if area is less than 20 then
        Removing it by morphological opening using disk structure element
    end if
end for
Applying median filter of size 3*3 to remove all remaining isolated pixels
```

V. EXPERIMENTAL RESULTS

A. Material

The STARE database, originally collected by Hoover et al. [12], comprises 20 eye fundus color images (ten of them contain pathology) captured with a TopCon TRV-50 fundus camera at 35 degree field of view (FOV). The images were digitalized to 700*605 pixels, 8 bits per color channel and are available in PPM format. The database contains two sets of manual segmentations made by two different observers. The FOV binary mask for each image isn't available, so we create it by hand for each image using MATLAB function named "impoly" which creates an interactive draggable and resizeable polygon on the image to specify the FOV. Also the images aren't divided into separated train and test sets, so the training pixels are selected from the manual segmentations made by the first observer. The segmentation performance of the proposed approach is compared against the segmentations of the second observer as ground truth.

B. Samples Selection

The STARE training set contains 919308 vessels pixels and 5281987 non-vessels (background) inside FOV. These pixels

needed to select samples from them for computation of the best features set using CFS heuristic. Since the ratio of vessel pixels to background pixels in each image and overall images is 1:9 on average, the samples set consists of randomly selected 1000 vessels pixels and 9000 background pixels from each image; there are 20 images give 200000 total samples. The best features set consists of the most repeated features resulted from multiple runs of CFS.

C. Results

Three measures are calculated for evaluating the segmentation performance of recommended features set by CFS with ACS. The first measure is the sensitivity (SN) which is the ratio of well-classified vessels pixels. The second measure is the specificity (SP) which is the ratio of well-classified background pixels. The third measure is the accuracy (ACC) which is the ratio of well-classified vessel and background pixels. Tables I and II show the ACS performance values with the old η and new heuristic function η^* respectively. It's shown from both tables that new heuristic function η^* improves largely the average performance of the ACS especially in sensitivity which increased by about 10% while specificity increased by 1% and accuracy increased by 2%. In the normal images set, the least performance of ACS with the old heuristic function was on image 7 but it's improved when using the new heuristic function; in sensitivity from 0.6608 to 0.9289, in specificity from 0.9005 to 0.9269 and in accuracy from 0.8742 to 0.9271. Also in the abnormal images set the least performance of ACS with the old heuristic function was on image 6 but it's improved when using the new heuristic function; in sensitivity from 0.6472 to 0.8188, in specificity from 0.8883 to 0.9213 and in accuracy from 0.8670 to 0.9122.

TABLE I
PERFORMANCE OF ACS WITH OLD HEURISITC FUNCTION

Image Number	SN	SP	ACC
01	0.6863	0.9565	0.9272
02	0.7210	0.9035	0.8869
03	0.8656	0.8966	0.8941
04	0.7965	0.9043	0.8933
05	0.7453	0.9301	0.9074
06	0.6472	0.8883	0.8670
07	0.6608	0.9005	0.8742
08	0.7493	0.9056	0.8897
09	0.7696	0.9216	0.9053
10	0.6645	0.9012	0.8751
11	0.7329	0.9098	0.8925
12	0.7412	0.9214	0.9024
13	0.7932	0.9271	0.9108
14	0.7680	0.9319	0.9115
15	0.8397	0.9377	0.9262
16	0.8038	0.9358	0.9174
17	0.6472	0.9188	0.8855
18	0.7351	0.8458	0.8381
19	0.8980	0.8865	0.8872
20	0.8316	0.8759	0.8718
Average	**0.7549**	**0.9100**	**0.8932**

Fig.1 and Fig.2 compare between the average performances of ACS with the old η and new heuristic η^* functions with respect to normal and abnormal images in STARE database. It's shown from Fig.1 that ACS with the new heuristic function outperforms ACS with the old heuristic function with respect to normal images by about 13.5% in sensitivity, 1.5% in specificity and 2.7% in accuracy. Also in Fig.2 the ACS with new heuristic function outperforms the ACS with old heuristic function with respect to abnormal images by about 6% in sensitivity, 1% in specificity and 1.4% in accuracy.

TABLE II
PERFROMANCE OF ACS WITH NEW HEURISITC FUNCTION

Image Number	SN	SP	ACC
01	0.8071	0.9157	0.9039
02	0.7305	0.9077	0.8915
03	0.8762	0.8921	0.8908
04	0.7684	0.9098	0.8953
05	0.7631	0.9263	0.9063
06	0.8188	0.9213	0.9122
07	0.9289	0.9269	0.9271
08	0.9391	0.9211	0.9229
09	0.9085	0.9385	0.9353
10	0.8368	0.9197	0.9106
11	0.9124	0.9223	0.9213
12	0.9204	0.9384	0.9365
13	0.8700	0.9432	0.9343
14	0.8779	0.9431	0.9350
15	0.8590	0.9371	0.9280
16	0.7824	0.9437	0.9211
17	0.8610	0.9401	0.9304
18	0.8805	0.8912	0.8905
19	0.8831	0.8881	0.8878
20	0.8500	0.9022	0.8974
Average	**0.8537**	**0.9214**	**0.9139**

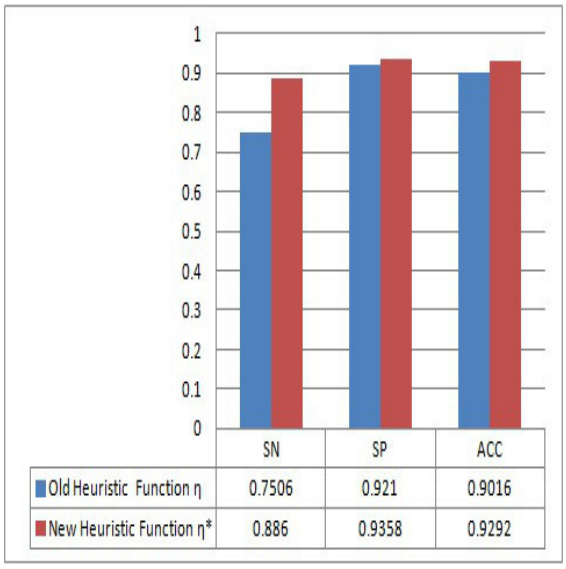


Fig. 1. ACS Performance on Normal Images in STARE

Fig.3 shows how ACS with the new heuristic function outperforms ACS with the old heuristic function in segmenting the small capillaries and bifurcations Fig.3-(d) with fewer false positives Fig.3-(h). So its better to use the new heuristic

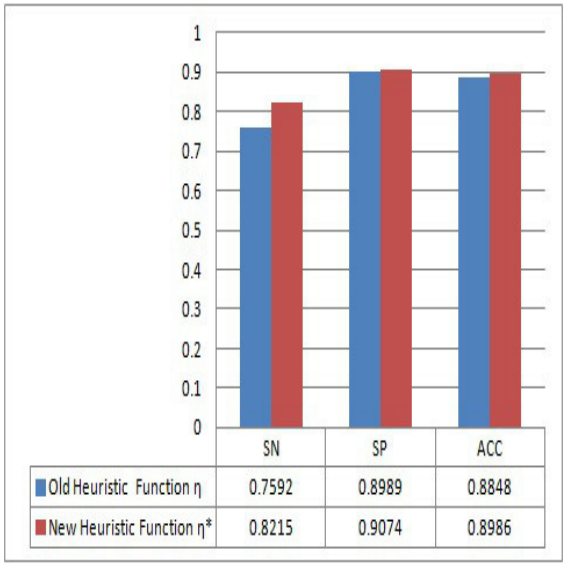


Fig. 2. ACS Performance on Abnormal Images in STARE

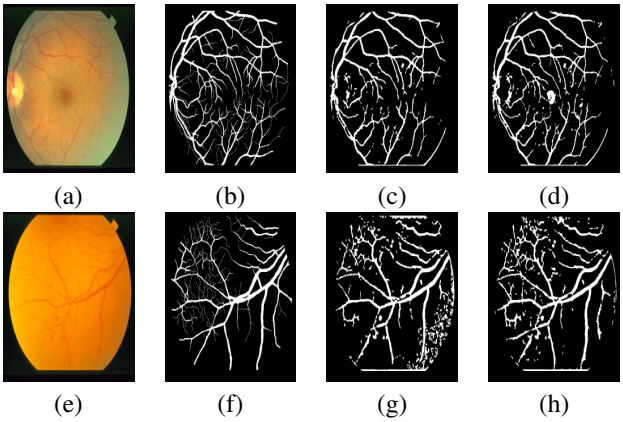


Fig. 3. ACS performance with old and new heuristic functions on normal and abnormal image: (a),(e) retinal image; (b),(f) manual segmentation by second observer; (c),(g) results of ACS with old heuristic function; (d),(h) results of ACS with new heuristic function

function that's based on probability theory instead of the old one that's based on Euclidean distance. Table III shows the performance comparison of the state of art methods as well as the old and improved approaches on SATRE database where the empty cells aren't reported by their authors. Both the old and improved approaches give the best largest sensitivity over the state of art methods. But their values of specificity are the lowest due to high false positives that affected negatively on their accuracies. Although this improved approach gives performance less than performances of state of art methods but its results are promising because they are obtained without learning and using complex features need to be computed at multiple scales and orientations. With respect to the time consumed by this improved approach, for single retinal image the selected four features computation takes on average forty two seconds while the ACS phase takes on average two minutes and forty five seconds on PC with Intel Core-i3 CPU at 2.53 GHz and 3 GB of RAM.

TABLE III
PERFORMANCE COMPARISON OF SATE OF ART METHODS

Method	SN	SP	ACC
Second Human Observer	0.8949	0.9390	0.9354
Staal et al. [13]	0.6970	0.9810	0.9516
Ricci and Pefetti [14]	-	-	0.9646
Marin et al. [9]	0.6944	0.9819	0.9526
Fraz et al. [15]	0.7548	0.9763	0.9534
Hoover et al. [12]	0.6751	0.9567	0.9267
Fraz et al. [17]	0.7311	0.9681	0.9442
Old Approach	**0.7549**	**0.9100**	**0.8932**
Improved Approach	**0.8537**	**0.9214**	**0.9139**

VI. CONCLUSION

The automated extraction of blood vessels in retinal images is an important step in early diagnoses of diabetic retinopathy and prevention of visual loss since diabetic retinopathy leads to retina revascularization. This paper presents an approach for automatic segmentation of blood vessels in retinal images using ACS based on features that are simple, fast in computation, needn't to be computed at multiple scales or orientations and highly discriminate between vessels and background in retinal images. The paper improves the features by adding new discriminant feature which is selected by CFS heuristic within the best features set. The paper also improves the performance of ACS using new heuristic function based on probability theory instead of the old one which is based on Euclidean distance. The paper is the first one that tested ACS performance on STARE database while the other two papers that also perform ACS-based segmentation of retinal images tested on DRIVE database. The improved approach sensitivity is the largest among the state of art methods. Because of simplicity of its used features, it can be more improved in future to give comparable performance to state of art methods especially by focusing on false positives reduction to increase its specificity and accuracy. Also in future it's needed to asses this improved approach performance on other databases of retinal images such as the DRIVE database.

REFERENCES

[1] K. Goatman, A. Charnley, L. Webster and S. Nussey, "Assessment of automated disease detection in diabetic retinopathy screening using two-field photography," PLoS. One, vol. 6, no.12, pp. 275-284, 2011.
[2] K. Verma, P. Deep and A.G. Ramakrishnan, "Detection and classification of diabetic retinopathy using retinal images," Annual IEEE India Conference (INDICON), pp. 1-6, 2011, DOI: 10.1109/INDCON.2011.6139346.
[3] M. Foracchia, E. Grisan and A. Ruggeri, "Extraction and quantitative description of vessel features in hypertensive retinopathy fundus images," In Book abstracts of 2nd international workshop on computer assisted fundus image analysis, 2011.
[4] V. Vijayakumari and N. Suriyanarayanan, "Survey on the detection methods of blood vessel in retinal images," Eur. J. Sci. Res., vol. 68, no.1, pp. 83-92, 2012.
[5] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen and S.A. Barman, "Blood vessel segmentation

methodologies in retinal images-a survey," Comput. Methods Programs Biomed., vol. 108, no.1, pp.407-433, October 2012, doi: 10.1016/j.cmpb.2012.03.009.

[6] X. You, Q. Peng, Y. Yuan, Y Cheung and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," Pattern Recogn., vol. 441, pp. 2314–2324, 2011.

[7] Ahmed. H. Asad, A. T. Azar, and A. E. Hassanien, "Integrated features based on gray-level and Hu moment invariants with ant colony system for retinal blood vessels segmentation," Int. J. Syst. Boil. Biomed. Tech., vol. 1, no. 4, pp. 61-74, 2012.

[8] M. Dorigo and L.M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," IEEE Trans. Evol. Comput. vol. 1, no. 1, pp. 53–66, 1997.

[9] D. Marin, A. Aquino, ME. Gegundez-Arias and JM. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using grey-level and moment invariants-based features," IEEE Trans. Med. Imaging, vol. 30, no. 1, pp. 146-158, 2011.

[10] M.K. Hu, "Visual pattern Recognition by Moment Invariants," IRE. Trans. Inform. Theory., vol. 8, no. 2, pp. 179–187, 1962.

[11] M.A. Hall, "Correlation-based feature felection for discrete and numeric class machine learning," ICML, pp. 359-366, 2000.

[12] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Trans. Med. Imaging , vol. 19, no. 3, pp. 203–210, Mar.2000.

[13] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," IEEE Trans. Med. Imaging, vol. 23, no. 4, pp. 501-509, 2004.

[14] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," IEEE Trans. Med. Imaging, vol. 26, no. 10, pp. 1357–1365, Oct. 2007.

[15] M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe, A.Basit, B. Uyyanonvara, A. R. Rudnicka, and C.G. Owen, "An ensemble classification-based approach applied to retinal blood vessel segmentation," IEEE Trans. Biomed. Eng., vol. 59, no. 9, pp. 1427–1435, Sep. 2012.

[16] M. S. Miri and A. Mahloojifar, "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction," IEEE Trans. Biomed. Eng., vol. 58, no. 5, pp. 1183–1192, May 2011.

[17] M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe,A.Basit, B. Uyyanonvara, A. R. Rudnicka, and C.G. Owen, "An approach to localize the retinal blood vessels using bit planes and centerline detection," Comput. Methods Programs Biomed., Sep. 2011.

[18] P. Kelvin, H. Ghassan and A. Rafeef, "Live-vessel: extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images", in Proc. 10th Int. Conf. Med. Image Computing and Computer-Assisted Intervention, Springer-Verlag, Brisbane, Australia, 2007.

[19] K.K. Delibasis, A.I. Kechriniotis, C. Tsonos and N. Assimakis, "Automatic model-based tracing algorithm for vessel segmentation and diameter estimation," Comput. Method. Programs. Biomed., vol. 100, pp. 108–122, 2010.

[20] M. Al-Rawi, M. Qutaishat, and M. Arrar, "An improved matched filter for blood vessel detection of digital retinal images," Comput. Biol. Med., vol. 37, pp. 262–267, 2007.

[21] B. Zhang, L. Zhang, L. Zhangb and F. Karray, "Retinal vessel extraction by matched filter with first-order derivative of Gaussian," Comput. Biol. Med, vol. 40 , pp. 438–445, 2010.

[22] M.G.Cinsdikici and D.Aydn, "Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm," Comput. Methods Programs Biomed. vol. 96, no. 2, pp. 85-95, 2009.

[23] S. Hooshyar and R. Khayati, "Retina vessel detection using fuzzy ant colony algorithm," in Proc Canadian Conf. Computer and Robot Vision (CRV), Ottawa, pp. 239-244, 2010.

[24] L. Espona, M.J. Carreira, M.G. Penedo, M. Ortega, "Retinal vessel tree segmentation using a deformable contour model, in Proc 19th international Conf. Pattern Recognition (ICPR), pp. 1–4, 2008.

[25] B. Al-Diri, A. Hunter, D. Steel, "An active contour model for segmenting and measuring retinal vessels," IEEE Trans. Med. Imaging, vol. 28, pp. 1488–1497, 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 207–210

Comparison of methods for hand gesture recognition based on Dynamic Time Warping algorithm.

Katarzyna Barczewska
Department of Automatic Control and Biomedical Engineering, AGH University of Science and Technology,
Kraków, Poland
Email: kbarczew@agh.edu.pl

Aleksandra Drozd
Department of Measurement and Electronics, AGH University of Science and Technology,
Kraków, Poland
Email: drozd@agh.edu.pl

***Abstract*—Gesture recognition may find applications in rehabilitation systems, sign language translation or smart environments. The aim of nowadays science is to improve the recognition systems' efficiency but also to allow the user to perform the gesture in a natural way. The article presents different methods (DTW – Dynamic Time Warping, DDTW - Derivative Dynamic Time Warping, PDTW - Piecewise Dynamic Time Warping) based on Dynamic Time Warping algorithm, which is commonly used for hand gesture recognition using small wearable three-axial inertial sensor. Additionally, different approaches to signal definitions and preprocessing are discussed and tested.**

To verify which of the methods presented is more accurate in case of gesture recognition, database of 2160 simple gestures was collected, and recognition procedure was implemented. The main goal was to compare the efficiency of each method assuming that each person should perform the movement naturally. Obtained results suggest that the most efficient method for the presented problem was the DDTW. The worst recognition performance was achieved with the PDTW method.

I. Introduction

THE recent advance of sensor technologies allows engineers to use smaller and smaller devices capturing human motion. These devices are cameras[1], game controllers, such as Microsoft's Kinect [2] or sensors: inertial [3], [5], [6] or built in data gloves[4]. One of the areas of interest is gesture recognition, which may find its application in game interface design, controlling virtual reality, smart environments but also in biomedical science. For example gesture recognition system may be used as a rehabilitation instrument to improve the sensibility of hands for people recovering from physical accidents or cognitive disabilities [3]. Another application might be an assistive translating system for the deaf people, who use sign languages to communicate [1], [4]. Much research has been done on the topic of gesture recognition, and designers of recognition system should always bear in mind that hand gestures are complicated and the way of performing a gesture varies depending on the person. Solutions presented in the literature using inertial sensor systems reach effectiveness of simple gesture recognition from 69% to 96% for general recognition and from 98% to 99% in recognition of one person's set of gestures [3], [5], [6]. Above solutions, however, assume that all gestures examined are strictly defined or they should be performed in one plane. Our research goal is to compare different variants of widely used for gesture recognition Dynamic Time Warping (DTW) algorithm and conclude which method gives the best results in recognition effectiveness taking into consideration aspects such as: defining the distance between two signals and choosing proper signals for analysis and recognition (acceleration or orientation in space). All tests (in contrast to [3], [5], [6]) are done assuming that a person should do the gesture in a natural way, which implies that a gesture can be performed in 3D space, according to one's preferences and physical conditions. Therefore, small wearable device such as three-axial IMU sensor was considered to collect motion data. Authors based on previous research presented in [7], where information about acceleration and Euler angles in 3-dimensional space of a sensor placed on a forefinger was used to classify the gesture. The aim of this research was to compare different variations of Dynamic Time Warping algorithm (DTW) including Piecewise Dynamic Time Warping (PDTW) and Derivative Dynamic Time Warping (DDTW) in gesture recognition. Moreover, some solutions concerning signal preprocessing leading to the recognition effectiveness improvement are presented.

II. Materials

Authors used database collected for tests described in [7]. Data acquisition was performed using 9 DoF inertial sensor, NEC-TOKIN, Motion Sensor MDP-A3U9S, placed on a volunteer's forefinger (Fig. 1). The small size (20 × 20 × 15 mm) and weight (6g) of the measurement module allowed user to move his hand and to bend the forefinger in natural way. The sensor contains 3-axis accelerometer, magnetometer and gyroscope. The output raw data contain information from each of these 3-axis sensors as well as the angular orientation expressed in Euler angles. Data transfer to the PC was performed via USB wire (sampling rate - 25 Hz). The set of 10 simple gestures was the recognition subject, their scheme is shown on Fig. 2. Similar gestures were also recognized in [3], [5], [6]. The database consisted of 2160 gestures, each in the separate text file, collected during 3 mea-

This work is supported by Ministry of Science and Higher Education in Poland in years 2013-2014

Fig. 1 Measurement module and its attachment to the forefinger. Original picture from [7]

surement sessions. In a single measurement session each of 9 volunteers performed each gesture 8 times.

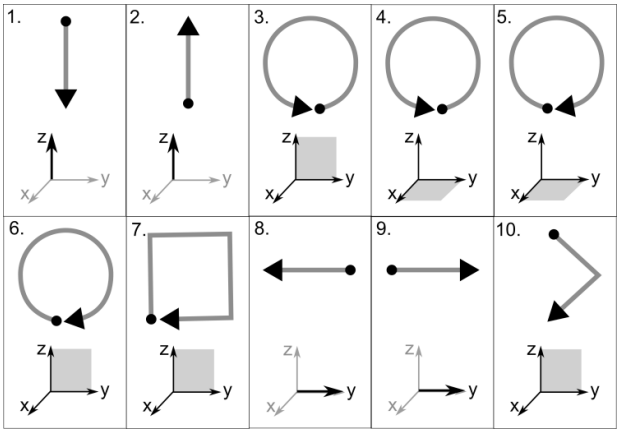


Fig. 2 Schemes of the gestures with coordinate system corresponding to that in measurement module. Main directions and planes of movement were marked.

In view of the fact that the individual gesture performance might vary significantly depending on the time, mood, tiredness or concentration, sessions were held at least one day apart. The database was divided into testing and training sets. 5 of 8 each gesture repetitions performed during measurement sessions were included into the training set, remaining 3 became the testing set. The training set contained 62.5% of all gestures, testing - 37.5%. Contrary to [3] and [6], authors did not assume that during gesture performance palm angular orientation does not change. In gestures made in natural way position changes can be observed, as well as changes of angular orientation. There were no restrictions about the way of gestures performing, what caused that some volunteers performed the same gesture using just one finger, others using the palm and some using whole arm.

III. METHODS

Preprocessing was applied to all measured signals, including mean filtering, signal values scaling to the interval [-1,1] as well as segmentation to obtain data corresponding only to the activity of performing a gesture uniformed for all examined people. Research on the topic of preprocessing was presented in [7] and showed that signal segmentation based on monitoring changes in Euler angles during movement improves the segmentation process.

Basic algorithm used to both determine exemplar set as well as to classify gestures was DTW algorithm (Dynamic Time Warping). This algorithm is commonly used to find the similarity between time series. Such a method was chosen because of the characteristic of collected gesture signals which were usually similar but transformed in time. The signals in our data base were different concerning their length, distribution of peak values and the velocity of performing particular gesture phases. Additionally two transformations of the base algorithm were implemented: Derivative Dynamic Time Warping (DDTW) and Piecewise Dynamic Time Warping (PDTW) [7-11].

A. Dynamic Time Warping (DTW)

DTW algorithm allows to compute the distance between two signals in the following procedure. Assuming there are two gesture signals (given by acceleration or angle signals changing in time): $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_m\}$ to align these two sequences using DTW for X and Y we need to define distance matrix *D* containing Euclidian distances between all pairs of points (x_i, y_j).

$$D(i,j) = d(x_i, y_j) \tag{1}$$

where

$$d(x_i, y_j) = |x_i - y_j|$$

Then we define cumulative matrix P recursively:

$$P(1,1) = 0$$
$$P(i,1) = D(i,1) + P(i-1,1)$$
$$P(1,j) = D(1,j) + P(1,j-1)$$

For $i,j>1$

$$P(i,j) = D(i,j) + \min\{P(i-1,j), P(i,j-1), P(i-1,j-1)\} \tag{2}$$

As a result of the DTW algorithm optimal total distance between X and Y after alignment was obtained, which is denoted by $q_{DTW} = P(n,m)$.

B. Derivative Dynamic Time Warping (DDTW)

DDTW algorithm is a variation of basic DTW algorithm. When the two series may have local differences in the Y-axis, it is useful to take into consideration derivative of the signals instead of the signals themselves. First, we calculate the estimate of derivatives of the signal. Then, as before according to equation (1) we construct an *n*-by-*m* matrix D where the (i_{th}, j_{th}) element of the matrix is the distance $d(x_i,y_j)$ between the two points x_i and y_j and finally calculate $q_{DDTW} = P(n,m)$.

C. Piecewise Dynamic Time Warping (PDTW)

PDTW algorithm uses time series transformed to reduced representation. A time series X of length n can be represented by a time series $\overline{X} = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_N}\}$, where $N<n$ and

the i-th element of $\overline{X}$ can be calculated from the following equation:

$$\bar{x}_i = \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \quad (3)$$

This means that data is reduced from n dimensions to N by averaging data in each of N frames. We denote the ratio of the length of the original time series to the length of the reduced representation by the compression rate c.

$$c = \frac{n}{N} \quad (4)$$

If the compression rate is high, it will reduce the time of performing calculations, but it will flatten the signal as well.

D. Different approaches to signal definitions

Since six different signal components (3 acceleration and 3 angle components) might be taken into consideration, same algorithms were used to compare the results of recognition based only on accelerations or only on angles or on both of the values together to show whether information about orientation in space may improve widely used technique based on acceleration analysis. As DTW algorithms compute distance between two signals, it can be used for two one-dimensional signals. However, the signals can be also treated as three-dimensional considering all acceleration or all orientation components. In this case distance matrix *D* (1) was modified and it consisted of distances between two points in three dimensional space defined by:

$$d(x_i, y_j) = \sqrt{(x_{i1} - y_{j1})^2 + (x_{i2} - y_{j2})^2 + (x_{i3} - y_{j3})^2}$$

for $x_i = (x_{i1}, x_{i2}, x_{i3}),\ y_j = (y_{j1}, y_{j2}, y_{j3})$

E. Exemplars

Authors proposed two approaches to the exemplars, that were the basis of gesture recognition: user-dependent and user-independent. In user-dependent approach, using samples from the training set, for each person for each of 10 gestures one gesture was indicated as an exemplar. The gesture from the training set became an exemplar, when it was the most similar to others in terms of one of the warping methods (DTW, DDTW and PDTW). For each warping method one individual (user-dependent) exemplar for every gesture was indicated. Then, taking into account all individual exemplars for each gesture, the one that was the most similar to others became general exemplar used in user-independent recognition. There were different sets of exemplars for acceleration, for angles, for all signals, also for 1D distance function and for 3D distance function.

IV. RESULTS

Gestures classification was performed using the testing set, all described below algorithms: DTW, DDTW and PDTW and corresponding sets of exemplars. Firstly, to recognize gestures authors used all acceleration and Euler angles signals at once. Secondly, to determine which parameter is more important for natural gestures recognition, classification was carried out for them separately. There were also two different approaches used in single parameter recognition:1) distance function in algorithms DTW, DDTW and PDTW was computed separately for each component of the parameter (for x, y and z axis in case of acceleration and for pitch, roll and yaw in case of Euler angles – 1D distance); 2) both acceleration and angles were treated as 3-dimensional signals and the distance in DTW algorithms was calculated like in 3-dimensional space giving one value for all acceleration components and one value for all angles (3D distance). Obtained results were also divided into individual and general cases. Individual are average efficiency values calculated for each person using his or her individual exemplar, general are average efficiency values calculated for everybody using the same general exemplar for each person. Efficiency was calculated as the sum of all correctly recognized gestures divided by the sum of all gestures. Efficiency values for each of the described recognition method are shown in Tab. I and Tab. II.

TABLE I.
CLASSIFICATION RESULTS FOR INDIVIDUAL CASE. GIVEN VALUES ARE RECOGNITION EFFICIENCIES.

Basis for classification	Distance function	DTW	DDTW	PDTW
Acceleration	1D	0,922	0,895	0,923
	3D	0,948	0,947	0,940
Angle	1D	0,757	0,861	0,747
	3D	0,789	0,921	0,784
All signals	1D	0,894	0,937	0,895
	Average	**0,862**	**0,912**	**0,858**

TABLE II.
CLASSIFICATION RESULTS FOR GENERAL CASE. GIVEN VALUES ARE RECOGNITION EFFICIENCIES.

Basis for classification	Distance function	DTW	DDTW	PDTW
Acceleration	1D	0,877	0,785	0,872
	3D	0,854	0,836	0,626
Angle	1D	0,584	0,674	0,579
	3D	0,638	0,822	0,626
All signals	1D	0,794	0,828	0,789
	Average	**0,749**	**0,789**	**0,698**

The PDTW method was conducted for three values of compression rate: 2, 3 and 5. All results for the PDTW method shown in tables above correspond to the compression rate equal to 2. The higher the value of this parameter, the lower recognition efficiency for collected database (the PDTW method might give better results for higher sampling rates or unfiltered signals). It can be observed that the highest value for recognition efficiency for individual case was obtained for classification based on 3-dimensional approach to acceleration signals and distance function – all warping methods lead to the value of 0.94 for this parameter. 3-dime-

sional approach to all other signals brought much better results in comparison to the corresponding methods with 1-dimensional distance function. Taking into account individual and general cases, statistic tests were conducted (at the 95% confidence level), indicating significant differences between results obtained by all described DTW methods. Tests revealed no difference only in one case: comparison of DTW and PDTW with compression rate c = 2 for general case. The most efficient warping method is the DDTW and the highest efficiency rate is also for acceleration signals. However, analyzing the various gestures separately, it appears that in some cases, the analysis of the angles lead to more accurate classification. Results of each gesture recognition efficiency for different basis of classification for the DDTW method are presented on Fig. 3.

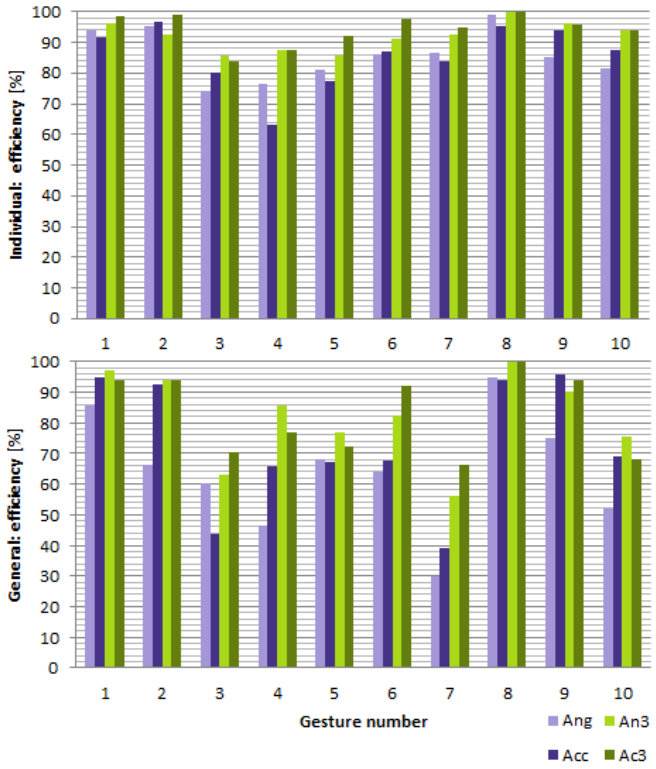


Fig. 3 Each gesture recognition efficiency for individual and general case for different classification basis: Ang – angles, An3 – angles and 3D distance function, Acc – acceleration, Ac3 – acceleration and 3D distance function. DDTW method.

V. Disscusion

Authors proved that recognition of gestures performed in a natural way without any constraints for the examined person is possible using methods based on Dynamic Time Warping Algorithms. Analyzing recognition efficiencies obtained for all DTW methods and taking into account the results of statistic tests which revealed significant differences between methods, it can be stated, that the most efficient method to described application is the DDTW. The worst recognition performance was achieved with the PDTW method. The higher compression rate, the higher reduction of information and the lower efficiency values. The reduction of information that was caused by the PDTW method was a disadvantage in this case, but it can occur that for higher sampling rates it will become an advantage.

Signals collected confirmed the theory that such aspects as time of the day, tiredness, concentration, experience may affect gesture performing which results in lower recognition efficiencies. As these problems are unavoidable, obtaining better results is a matter of algorithms and preprocessing.

Moreover, analyzing Fig. 3 it can be observed that there are gestures which are much better recognized while using angular orientation as a basis of classification. Solution to improve this method to recognize natural gestures would be to combine information about angular orientation and acceleration. Possibly treating the signal as a 6 dimensional (3 acceleration components and 3 angle components) would cause increase of recognition efficiency.

In future, research data base enlarging should be considered to verify the hypothesis which of the methods presented is more accurate in case of general (subject independent) gesture recognition. Another improvement may be creating more general exemplar set containing modeled signals (instead of signals registered by the sensor), which will reduce the influence of between subject variability.

References

[1] Paulraj M. P., Yaacob S., Azalan M. S. Z., Palaniappan R., *A Phoneme Based Sign Language Recognition System using 2D Moment Invariant Interleaving feature and Neural Network*; IEEE Student Conference on Research and Development, 2011.

[2] Wang Y., Yang C., Wu X., Xu S., Li H., *Kinect Based Dynamic Hand Gesture Recognition Algorithm System,* 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2012

[3] Hussain S.M.A., Harun-ur Rashid A.B.M., *User Independent Hand Gesture Recognition by Accelerated DTW,* IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision, Proceedings, Dhaka, Bangladesh 2012.

[4] Liang R.-H., Ouhyoung M., *A Real-time Continuous Gesture Recognition System for Sign Language*, Third IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings, 1998.

[5] Liu J., Wang Z., Zhong L., Wickramasuriya J., Vasudevan V., *uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications,* Pervasive and Mobile Computing Journal, Vol. 5, Issue 6, Elsevier, Amsterdam, The Netherlands, 2009.

[6] Akl A., Feng C., Valaee S., *A Novel Accelerometer-Based Gesture Recognition System,* Transactions on Signal Processing, IEEE, vol. 59, No. 12, December 2011.

[7] Barczewska K., Drozd A., Folwarczny Ł., *Rozpoznawanie gestów z wykorzystaniem czujników inercyjnych o 9 stopniach swobody,* Pomiary Automatyka Kontrola, Vol. 59, No. 3, March 2013.

[8] Keogh, E., Pazzani, M., *Derivative Dynamic Time Warping*. In First SIAM International Conference on Data Mining (SDM'2001), Chicago, USA

[9] Keogh, E., Pazzani, M., *Scaling up Dynamic Time Warping for Datamining Applications*. In 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, 2000

[10] Müller M.: Information Retrieval for Music and Motion. Chapter 4: Dynamic Time Warping. Springer Verlag 2007;

[11] Helwig N. E., Hong S., Hsiao-Wecksler T., *Time-Normalization Techniques for Gait Data*, 33rd Annual Meeting of American Society of Biomechanics Materials, State College, PA, USA, 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 211–218

Designing multiple user perspectives and functionality for clinical decision support systems

Christopher D. Buckingham
and Abu Ahmed
School of Engineering and Applied Science
Aston University
Birmingham, United Kingdom
Email: c.d.buckingham@aston.ac.uk

Ann Adams
Warwick Medical School
University of Warwick
Coventry, United Kingdom
Email: a.e.adams@warwick.ac.uk

***Abstract*—Clinical Decision Support Systems (CDSSs) need to disseminate expertise in formats that suit different end users and with functionality tuned to the context of assessment. This paper reports research into a method for designing and implementing knowledge structures that facilitate the required flexibility. A psychological model of expertise is represented using a series of formally specified and linked XML trees that capture increasing elements of the model, starting with hierarchical structuring, incorporating reasoning with uncertainty, and ending with delivering the final CDSS. The method was applied to the Galatean Risk and Safety Tool, GRiST, which is a web-based clinical decision support system (www.egrist.org) for assessing mental-health risks. Results of its clinical implementation demonstrate that the method can produce a system that is able to deliver expertise targetted and formatted for specific patient groups, different clinical disciplines, and alternative assessment settings. The approach may be useful for developing other real-world systems using human expertise and is currently being applied to a logistics domain.**

I. Introduction

MANY Clinical Decision Support Systems (CDSSs) with appropriate functionality have been successfully developed in academic institutions but never seen the light of day within healthcare practice. There are two fundamental reasons why these systems are not adopted. One is the failure to integrate with the way organisations and their individual employees work. The other is the inability to communicate information effectively beyond the immediate remit of the CDSS, which is often too narrow in the first place. This paper describes a research approach that attempts to circumvent both problems by developing a CDSS that has flexible requirements and data sharing protocols built into the design process from the very beginning. The CDSS is the Galatean Risk and Safety Tool, GRiST [1], [2], that helps assess and manage risks associated with mental-health problems.

The aim of the research was to design GRiST so that it could disseminate mental-health expertise using appropriate language for the particular type of recipient and in a format commensurate with the variable circumstances of assessment. This is no easy task because it would need to accommodate end users ranging from psychiatrists with years of specialist medical education to carers or charity workers who may have minimal training. In fact, GRiST was later adapted for self-assessments, by patients who do not have any predefined common ground apart from mental-health problems. Assessment contexts were also highly variable because GRiST was intended to be deployed for mental-health patients across the care pathway, from primary care, through secondary care and specialist services, and back to care in the community.

The complexity of health services in general and mental health in particular is one reason why the UK Government had so many problems with its National Programme for Information Technology [3] that was intended to revolutionise information systems and processes within the National Health Service (NHS). When GRiST was available for deployment in 2006, the oft-acknowledged "cinderella" mental-health services were still more paper-based than most in the NHS. GRiST set out to tackle barriers to information technology (IT) and its adoption by a design process dedicated to developing flexible interfaces and delivery formats for heterogeneous users and contexts. The research questions were: (i) how can the knowledge base be presented in the format and language most appropriate for each intended type of user? and (ii) how can the information technology generate flexible interfaces to the knowledge so that they fit with the different contexts of assessment?

The paper will first briefly review the clinical rationale for GRiST before describing the main functionality and underlying philosophy of the system. This will provide the context for the cognitive engineering approach that was used to develop knowledge structures providing risk assessments and advice. Their implementation as a sophisticated set of linked XML trees will be described, showing how they support the full GRiST CDSS and its deployment across health and community settings. Examples of the variety of interfaces and language used will be given along with an evaluation of the clinical implementation and adoption. The paper will end by considering the next steps for the research programme and how these have been facilitated by the knowledge structure design principles.

II. Background

To prevent serious untoward incidents (SUIs) amongst inpatients and in the community, clinicians need new and reliable

978-83-60810-53-8/$25.00

research evidence to help them detect high risk patients and to support risk management decisions. Despite patient safety being central to NHS policy [4], SUIs remain worryingly common [5]–[7]. Identified causes are lack of sufficient, accessible information about patients' risk profiles [5] and poor risk management or care planning [5], [6]. Risk assessment and management are core competencies for mental health clinicians [8], [9], but the two processes are often not properly connected [10].

There is a clear need to improve clinical practice, which was the motivation for GRiST. It is set apart from alternative risk-assessment and management tools by explicitly modelling human expertise within a generic model of psychological classification. This was a fundamental design principle; if GRiST is based on how humans in general organise their knowledge and reason with it, then its expertise will be in a universally accessible format. It enables GRiST to transcend disciplinary specialisms and opens its expertise up to people with no training at all.

GRiST is designed to assist the early detection of multiple risks amongst people with mental health problems, including suicide, self-harm, harm to others, self-neglect, and vulnerability. It records patient data (cues) to provide a precise information profile that supports the risk judgements given by clinicians.

Risk assessment can be formulated as a classification problem where each risk such as suicide or harm to others is a class and the support for each class determines the level of risk. The factors determining which risk gains most support will be the patient cues such as previous risk history, current intention, emotional and mental state, as deemed relevant by the assessor. The classification task is to formulate the support for each risk from the input data and activate appropriate interventions associated with the most supported class.

In GRiST, risk classes are represented by hierarchical knowledge structures or trees called galateas [11], which are used to represent mental-health expertise. The trunk or root node of the tree is the risk. It is deconstructed into subconcepts that are themselves trees until the leaf nodes are reached, representing the input data.

Figure 1 provides a hypothetical illustration of how the galateas represent classes and their support. The data used for input to the tree can be any type but it is then converted into a fuzzy-set membership grade, MG, from 0 to 1. Zero represents no support for the root decision class and 1 represents maximum support, but for this item of information alone; its MG at this point is independent of any other item. The main role of the MGs is in converting from real-world patient data to the model input. This is shown in Figure 1 by the MG row of the datum nodes, which defines a distribution of MGs matching the range of potential input data values. Values above or below the range take the MG associated with the maximum or minimum value respectively; values within the range are found by linear interpolation if they don't match a value specified in the distribution. For example, the "number of attempts" datum in Figure 1 has the value 3, which is 0.6

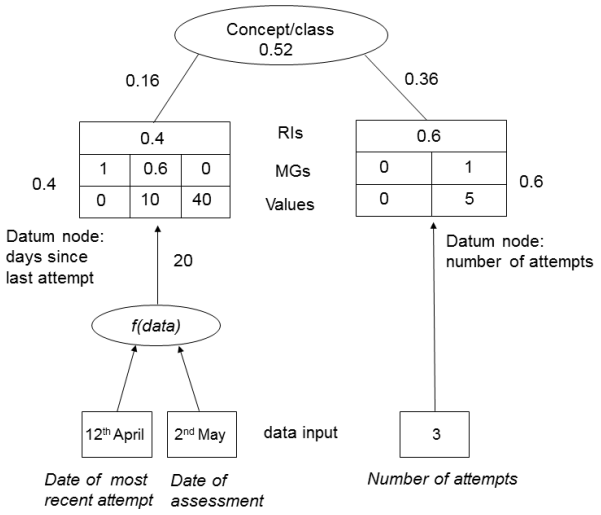


Fig. 1. Hypothetical example of how membership grades (MGs) are processed from patient input data to risk concept. RIs are relative influences, which represent the weights of data and concepts.

along the value range between 0 and 5 and so is assigned an MG that is 0.6 between 0 and 1; i.e. 0.6, as given by the MG outside the box for that datum. For others, such as "days since last attempt", the patient data is passed through a function, *f(data)*, before matching the value-mg distribution: two dates, in this case, which the function uses to generate the number of days between them, 20. Twenty is one-third between 10 and 40 and so is one third along the MG continuum between 0.6 and 0, producing an MG of 0.4, likewise shown outside the datum-node box. The MG is then multiplied by the RI associated with the datum, as shown in the RIs row, to give the MG contribution to the parent concept.

The parent concept MG is the sum of its children contributions, which is how 0.52 is assigned to the concept node in Figure 1. If this concept also had a parent, then the concept would have its own RI and its contribution to the parent would be the product of its RI and MG in the same way that it received its children MGs. The MGs percolate in this manner through to the root node to produce the overall class membership and thus the risk evaluation. Equation 1 formalises the process

$$MG_C = \sum_{i=1}^{n} MG_i RI_{pi} \tag{1}$$

where C is a concept, MG is the membership grade generated at each datum node, i, of the concept, and RI_{pi} is the product of all the RIs along the path, p, from the datum node to the concept.

The focus of this paper is how the hierarchical modelling of expertise translates into an ontology that can drive the GRiST CDSS. The added value of the hierarchy is that it represents the conceptual structure understood by human decision makers when relating influential factors to the decisions taken. There is plenty of evidence for the psychological validity of this

hierarchical knowledge structuring. For example, expert chess players "chunk" positions of chess pieces into hierarchical types of game states [12]. Similar strategies have been shown in other domains such as architecture [13], fault diagnosis [14], and medicine [15]. A review of the evidence [16] concluded that "on balance, it is difficult to dismiss hierarchical organisation as only a construct" (p31) and more recent research has begun to show its neural correlates [17], [18].

The psychological grounding of GRiST is not unique, of course, when it comes to intelligent knowledge-based systems (IKBSs) [19], [20]. However, it uses a generic classification model that represents expertise in a non-specialist format. It can be understood without requiring clinical training and makes it ideal for communicating to heterogeneous users (see [11] for more on the Galatean model rationale).

A. Cognitive engineering and the GRiST ontology

The GRiST approach to constructing decision support systems can be categorised as cognitive engineering because it is the application of cognitive science to IT systems that are intended to help solve real-world problems [21]. For cognitive engineering, models need to encapsulate expertise in a format that can be accessed by the experts and that is commensurate with the inputs and outputs those experts are familiar with in their problem-solving worlds [22]. IKBS Engineers were coming to this conclusion with the idea of situated cognition [23], which argues that thinking cannot be separated from the environment [24], [25]. These environments change and static IKBSs based on a single, giant elicitation exercise are doomed to fail because they will not be flexible enough to evolve or even be maintained easily [26, pg. 767].

There has been a change in tack from psychology to the data itself, with machine learning, data mining, and pattern recognition approaches coming to the fore. In a recent review of artificial intelligence in medicine [27] Peter Szolovits points out that in the early days, "we thought we knew a lot, but had little or no actual data. Today we are inundated with data, but have correspondingly devalued expertise" (pg. 12). The focus is on the machine, how knowledge can be structured for easy processing, and how useful outputs can be induced from the data. It is the same focus that stimulated the rise of ontologies for organising data into shared knowledge bases. Nevertheless, despite Musen's claim that cognitive models do not lead to scalable and maintainable IKBSs [28], "the symbiosis between cognitive science and cognitive engineering shows no sign of abating" [21, pg. 582] and continues to be the case in medicine [29].

The GRiST research tries to bring human and machine closer together using a form of ontology that has an intuitive connection with the knowledge used by mental-health experts. The interface between human and machine ontologies should be a primary focus for knowledge engineering [30], especially for CDSSs based on clinical expertise. The most basic form of ontology is a controlled and extensible vocabulary [31], [32], which means that dictionaries and thesauri would count. These are very familiar to people and emphasises the point that ontologies are not strictly the preserve of machines. Indeed, all sensate beings create some kind of ontology for interacting with their environment [33].

Maintaining the intuitive representation of the GRiST knowledge base meant that the terms should reflect the natural language of human users [34], [35]. This is particularly important in mental-health risk screening because of the diversity of information that relates to risk and the lack of any all-encompassing coding schemes. Where schemes do exist, e.g., ICD-10 [36] and DSM-IV [37], they focus on diagnostic categories for mental disorders such as depression and schizophrenia and do not encompass the diversity of peoples' histories and current behaviour that impact on risk. Attempts to create ontologies within mental health have also focused on diagnoses [38], not risk, and have been aimed at data interoperability rather than formalising expertise and clinical decision making.

The GRiST ontology development was designed to ensure the end product met the needs of its users and organisational settings [39] by extensive iteration between clinicians and the evolving CDSS. The galatean psychological model kept the human-machine interface open and intuitive. It has a precisely specified semantics for hierarchical knowledge, incorporating parameters required for processing uncertainty, and the mathematical functions for propagating them through the hierarchy. This coupling of the ontology with its problem-solving method (classification) helps construct a system that solves real-world tasks [40], but does so by emphasising the fluid relationships of human intuition rather than machine formalisms and logic-based reasoners [41]. The next section explains the method in detail.

III. METHOD

The goal of the methods reported in this paper was to create galatea knowledge structures that were able to evolve with expert consensus and support customised knowledge delivery for a variety of end users accessing it in different contexts. The first problem was how to develop and manage the hiearchical knowledge, which was solved using mind maps.

A. Mind maps

One of the most intuitive aids for note-taking, brainstorming, and generally organising ideas is the mind map [42]. Its layout reflects the goal of representing free-flowing, unconstrained associations of the mind at the same time as structuring knowledge hierarchically; it exactly accords with the knowledge-engineering requirements of GRiST.

There are many mind mapping software programs available. Freemind [43] was chosen because it is: (i) open-source; (ii) available across platforms; (iii) creates node structures that can be easily edited; (iv) enables icons to be incorporated into the nodes; (v) attaches notes to the node without obscuring the structure; and, most importantly, (vi) uses XML directly for representing the mind map rather than it being only an export choice. Its structure-editing role was integrated with the GRiST knowledge-engineering toolkit by creating an XSLT

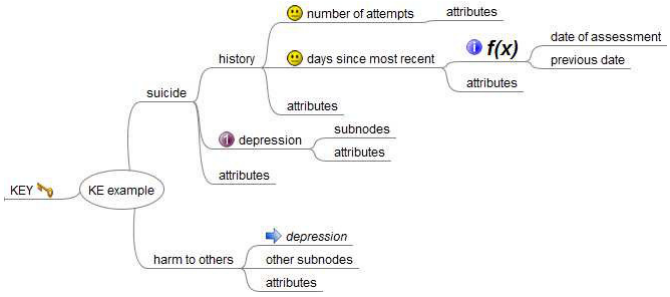


Fig. 2. Hypothetical and simplified mind map of risk assessment expertise

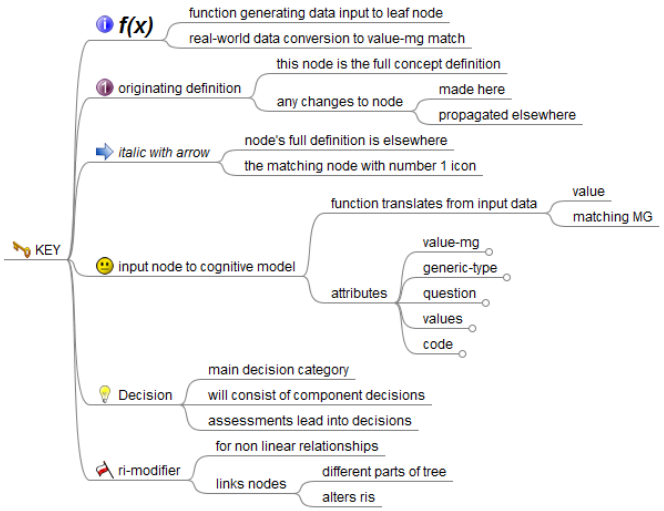


Fig. 3. Expanded key node showing icons that help drive knowledge engineering

document that transformed the Freemind mind map XML into the GRiST structure tree.

A useful resource for helping users control structure changes and also to direct the style sheet is Freemind's icons. Figure 2 shows a simplified example of how Freemind defines the knowledge structure; Figure 3 expands the "Key" node that explains the icons helping control the translation between mind map specification of knowledge structures and the subsequent GRiST XML trees. Many concepts, such as depression, underlie all risks and so are repeated in the knowledge hierarchy. The blue arrow icon enables the mind map to define the full structure in one place. When the style sheet detects the blue arrow, it looks for a node with the same name that has the round number 1 icon associated with it (see Figure 2). The other icons are similarly used to specify aspects of the galatean structure, such as the face, which identifies leaf nodes of the galatea where value-mg distributions are defined. The $f(x)$ node indicates which patient data are required to generate the matching value to the value-mg distribution. This is the case for the time period between the assessment and the most recent suicide attempt, for example, as shown in Figure 1.

Every risk node, both leaf and concept, has a subnode called "attributes", which contains attributes required by the GRiST XML trees. These enable the XSLT conversion document to translate between Freemind mind map format and GRiST XML nodes by making the attributes an explicit structure in Freemind. Otherwise, they would be unrecognised and ignored by Freemind when creating the mind map. The XSLT conversion document looks in the attributes subnode and creates them as well-formed attributes of the output XML tree that is at the root of the GRiST ontology. The next section introduces the GRiST XML trees and their attributes in more detail.

B. GRiST XML tree functionalities, attributes, and relationships

Once the initial knowledge structure has been specified in Freemind, it is translated via the XSLT specification into the GRiST initial XML. The idea is to have a base tree that incorporates the requirements for all patient types and assessment circumstances. It is a kind of "universal" or Everyman tree incorporating every issue for every user. The knowledge-engineering task is to encapsulate the different subtypes with their particular perspectives and priorities within the GRiST XML trees and extract them for delivery within the CDSS.

For GRiST, the subtypes reflect the variety of patient being assessed and the contexts of assessment. Four patient types or *populations* were quickly distinguished as GRiST developed: children and adolescents; working-age adults; older adults; and learning disabilities. Delivering GRiST across assessment contexts also required functional variations that may apply to more than one population (i.e. are not unified with populations) and so needed to be treated separately. Customisations of this type are called services. GRiST is thus tailored along the axes of populations for different assessment trees and services, if it turns out that the functionality needs customising as well as the underlying classification tree; service functionality can be applied to combined subsets of more than one population.

Three objectives were pursued when designing the GRiST XML structures: (i) define the structures of all trees; (ii) instantiate the trees with parameters required for classifying patients according to their particular population; and (iii) generate the specific data structures required for delivering the variety of CDSS functionality for end users. The trees can be summarised as follows (Table I defines the main attributes):

The Super Structure Tree (SST), which contains for all populations, all structural information about nodes and the questions attached to them, with associated values and membership grades. The SST also enumerates all the services that it may be used in with the accompanying more modest functional customisations required across different end users. The SST is the base "Everyman" tree that holds common information across them all as well as information about how to generate the distinctive sub-trees.

The Structure Tree (ST), which contains structural, question, value, and value-mg information for an individual population. It is generated from the SST, and can be conceived of as an SST tailored for one and only one population. Service

TABLE I
EXAMPLES OF ATTRIBUTES USED BY THE GRIST XML SPECIFICATIONS TO DRIVE KNOWLEDGE ENGINEERING TOOLS AND DELIVERY FUNCTIONALITY OF THE CDSS. THE TREES COLUMN IDENTIFIES THE TREES THAT CONTAIN THE ATTRIBUTE.

Attribute	Semantics	Trees
`label`	name of tree node that can vary for populations	all trees
`code`	code for tree node, which is invariant	all trees
`populations="(`*`population-name`*`)"`	different populations of users defined by the tree	SST
`services`	defines services with particular configurations	SST, ST, RIT,
`help="(`*`help text`*`)"`		SST, ST, RIT, QT
`generic="[path to generic node]"`	locates full definition of node	SST, ST, RIT
`generic-type="g"`	repeating nodes with invariant uncertainty parameters	SST, ST, RIT
`generic-type="gd"`	repeating nodes with varying uncertainty parameters	SST, ST, RIT, CAT
`generic-datum="[path to definition of datum]"`	locates full definition of node	SST, ST, RIT
`value-mg="((0 0)(7 1)(10 0.5))"`	association list of values and membership grades	SST, ST, RIT
`level`	prunes tree at different levels of assessor expertise	SST, ST, RIT
`question="`*`question`*	question for collecting item of information	SST, ST, RIT
`values="`*`values`*`"`	defines the type of the item of information	SST, ST, RIT, QT
`layer="n"`	specifies order of initial data collection	SST, ST, RIT, CAT
`filter-q="`*`question`*`"`	question indicating whether subtree is applicable or not	SST, ST, RIT, CAT
`persistent="hard/soft/value"`	carries data forward from previous assessment	SST, ST, RIT, QT
`service`	configures node with given services customisation	SST, ST, RIT, CAT
`prune-for`	removes branch/node for a population	SST
`other attributes ...`	emerging out of knowledge engineering	any

definitions contained in the SST will be carried over to the ST, meaning that customisations defined for a given service type will apply across all populations. The ST is used to generate the RIT corresponding to a population.

The Relative Influence Tree (RIT), which holds the RIs for all nodes. Nodes with generic-type attribute of "gd" are expanded in all locations that point to them because these nodes may have different internal RIs (for concepts) or value-mgs (for datum nodes) in the locations. The RIT structure is generated from the ST and used to elicit and store the RIs.

Together, the ST and RIT are sufficient to specify all the information required for the Galatean psychological model of classification to be instantiated. They provide the complete ontology and problem-solving package and are the end products of the knowledge elicitation stage. However, four more trees are required by the end-user decision support tools. Three are derived from the ST and/or RIT and one is generated by the decision tool during the assessment of decisions, as follows:

The Class Assessment Tree (CAT) is generated from the RIT because it needs the RI attribute. It produces the full galatean tree for classifying objects and so has all nodes fully expanded in all locations, with no paths to separate generic nodes.

The Question Tree (QT) is generated from the RIT and has all the information required to display questions, obtain associated answers, and generate membership grades for the answers.

The Answer Tree (AT) is generated during an assessment by the data-gathering tool, and stores all user-supplied data.

The Landmark Tree (LT) is a tree used for helping assessors to navigate the CAT during assessments. It is a reduced version of the CAT, and will highlight nodes that may be of particular interest or relevance to the current assessment. It is envisaged that this will be derived from the Freemind mind map defining the base structure.

Table I provides examples of attributes that represent data defining the psychological model and its specifications of variations for populations and services. For example, `prune-for="(older working-age)"` means remove this branch of Everyman for older-age and working-age adults but not for any of the other listed populations. Many of the node attributes have an "enhanced" form where they can have different values for the populations of classification trees. This was required to encompass the variety of end-user perspectives that went beyond simply providing different views of the Everyman tree structure but also different representations of the data. For example, the tree node labels for the service-user self-assessment population are different to those seen by the mental-health practitioners assessing services users, as shown for the suicide node label:

```
label="(((service-user) "ending your own
life") ((iapt learning-disabilities older
child-adolescent working-age) "suicide"))"
```

Lisp-like association lists are used to pair the population or service with its customised value. They provide great flexibility for dynamically creating and updating customisations within the SST. They are integral to giving the correct values to the tree transformation procedures that generate the correct trees for each population and service.

Customisation of the same population tree across different service provider contexts is effected by the `services` attribute in the top-level root node of the ST. Each service

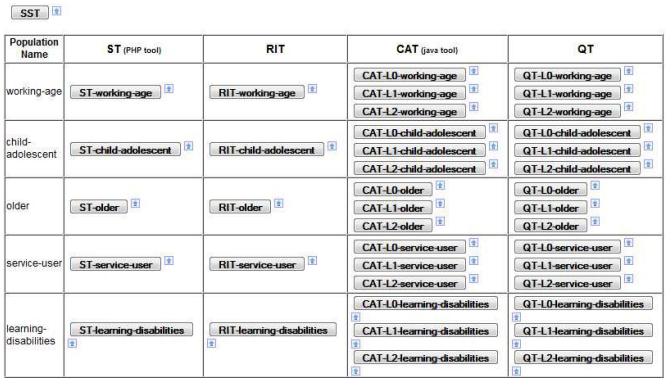


Fig. 4. Screenshot of the admin interface for managing population trees

provider represents a particular set of (minor) customisations/configurations that will be applied to GRiST's question set when conducting an assessment for that service. The top-level root node of the ST will have a:

```
services=
  "((structure
      ((service1 (association list of mods))
       (service2 (association list of mods))
       (...)))
    (rendition
      ((service1 (association list of mods))
       (service2 (association list of mods))
       (...)))"
```

attribute, defining all the services for which customisation/configuration data exists in the ST. Within the `services` attribute, these modifications (abbreviated to mods in the example) are organised as:

Structural modifications: those that involve dynamic (just-in-time) manipulations of the trees in some way prior to their being used to drive a GRiST assessment. Assessment tools will be agnostic of the structural changes, and will therefore not need to perform any additional processing.
Rendition modifications: those that involve dynamic (just-in-time) manipulations of the rendition of the GRiST assessment. Additional coding effort will be required in each assessment tool to realise the rendition manipulations.

The normal behaviour is that in order for a tree node to be amenable to the application of the defined modifications, it needs to "subscribe" to the modifications via a service attribute. For example, `service="rendition"` placed in a node will cause all the service-specified renditions to be applied to that node, such as providing it with a prefix or changing the font to bold, both of which were required for the IAPT service described in the results.

IV. Results

The principle behind managing the different XML trees is to have one single master tree (i.e., the SST) that is used to generate all the other ones. The conversions will be carried out using XSLT in the main and the resulting trees will be labelled so that they can be linked to the particular master tree from which they are derived.

An administrator's interface was provided for uploading, viewing and manipulating the trees (Figure 4). The website automates the derivation of all trees from each SST: namely each population's STs, RITs, and CATs and QTs at various levels (higher level trees pruned at concepts can be used for assessors with greater expertise who can use judgements in place of low-level data). Active trees (i.e. those delivered to end users) need to be marked so that experimental or legacy trees can be made or kept available alongside the current "live" ones for testing before deployment. When the GRiST CDSS is accessed from a patient record system, parameters are passed to the clinical server to indicate which patient assessent is being conducted and what population to invoke.

The success of the methodological approach has been clearly demonstrated over the years by the ability to create new trees suited to particular patient types when requested by mental-health practitioners. The first derivations enable GRiST to cover all age ranges, not just working age, and an additional specialist population was recently added for learning disabilities. Most conclusive was the need to produce a tool for self-assessments that patients could use in the community. This led to the development of myGRiST that exploited the enhanced attribute values to produce variations of nearly all the tree nodes. Not only were different data-collection questions and accompanying help text boxes required but also most tree branch names were changed to ones more suitable for non-clinicians. These substantial variations were all linked to the same common Everyman tree, which meant that clinical and patient answers were always directly comparable: the trees provide a common language and knowledge base despite their multiple manifestations.

The method facilitates a single genotype with a variety of phenotypes. But it also provides different drivers of the end-user CDSS tools. Clinicians tend to be under serious time pressure and want to collect data as efficiently and concisely as possible. This meant their tool was driven by the ST, which keeps class-specific questions separate from generic concepts that are applicable to all classes, and only expands generic concepts once. In other words, it reflects the original mind map with no structural redundancy. On the other hand, the service users conducting self assessments have more time to explore risks dynamically. They wanted less control over the order of access and they also wished to answer generic concepts such as their relationships, current behaviour, living conditions, etc. in the context of whichever risk they were considering. For them, their tool was driven by the CAT, which expands all concepts in all locations to provide full trees for every risk.

The first service customisations were also motivated by time pressures, especially for primary-care practices. Here, the visibility and rendition of risk tree data was needed across all populations (i.e. not limited to only working-age adults). At present, the main service customisation currently in use is for a primary-care service provided to general practitioners called Improving Access to Psychological Therapies (IAPT).

Since mental-health organisations started using the electronic (web service) version of GRiST in 2010, more than 2,000 clinicians have conducted the following number of completed assessments for the different populations and the IAPT service.

working age population: 50,193
children and adolescents population: 4,008
older adults population: 28,188
iapt service: 696

These results are testimony to the knowledge engineering method that facilitated accommodation of different end-user requirements. They also demonstrate the robustness of their implementation, with different populations and services being requested dynamically in real-time.

V. Conclusions

This paper has described a methodology for eliciting, evolving, and delivering complex mental-health expertise using a psychological model of classification. The new research reported here develops GRiST from its initial construction [44], [45] into a fully-fledged knowledge engineering environment that can manage both structure changes and subtle variations in knowledge parameters within an integrated system. The sophisticated application of attributes and XSLT to deliver customised services has been proven by continuous delivery to mental-health practitioners, every hour of every day of every week.

GRiST is generating an accumulating database of patient information that will help parameterise the galatean model underlying the CDSS. The RIs will be learned from the data to provide models of expert consensus that can detect risks more accurately and target appropriate advice. Efficiency of data collection will exploit the latest research on fast and frugal classification [46], [47] where the most important information is identified first and processed very rapidly. More in-depth analysis only needs to take place if the decision-maker remains ambivalent.

New GRiST versions have been requested for forensic services, accident and emergency, and ante-natal clinics. The knowledge engineering methodology means they can be delivered in short timespans because few or no changes are required in end-user applications. Structural information in the GRiST ontology is given by the node nesting but information about the role of the node, both for the psychological classification model and its delivery within a CDSS, is held by attributes. There is no limit to the number of attributes that can be added, which provides great flexibility for amending node behaviours as knowledge engineering progresses.

The disadvantage is that the meaning of attributes has to be recorded outside the XML. Ontology specification languages such as OWL [48] have more power to include semantics and logical relationships but the galatean approach traded flexibility of knowledge structure with the need for a detailed specification document to accompany it. Any tools operating on the XML would need to refer to the specification document and ensure their operations were in full accordance with the definitions. Once this has been achieved, then any changes in the ontology using existing attributes require no further software development.

The GRiST ontology was able to integrate natural language for people with codes more convenient to machines, helping to keep the human-machine interface closely aligned [34], [35]. Now that it has stabilised, it makes more sense to consider translating it into a formal specification language [49]. This would make it easier to validate the tree transformations and embed the rules more formally within the machine specification. An ontology language would also help with the problems of interoperability that are endemic within mental-health services, possibly to a greater extent than in other health areas. There is an inherent difficulty with categorising intangible mental-health constructs but data interchange on more straightforward health and social-care patient data would be beneficial. GRiST is explicit about how this generic information links to risks but mental-health organisations often collect it in other documentation not related to risk. Sharing it has proved problematic and its absence from risk documentation could present a danger to proper care [10].

Whether or not the current GRiST ontology is converted into a language such as OWL, instantiation of the galatean model as a formal specification of mental-health risk expertise has led to assessment tools that have generated considerable interest both in the UK and abroad. The method has applicability to any domain where human expertise can be disseminated within a DSS and where the DSS will be used by people with varying needs, characteristics, and work contexts. Evidence for this is emerging from the logistics domain where the galatean approach is being applied. The expertise and decision context is very different, with the goal being to optimise the use of vehicles for deliveries and collections based on predicting the number of orders. The knowledge trees have already successfully captured expertise that is able to represent alternative perspectives [50] and implementation of the CDSS is currently underway. Irrespective of the particular application domain, whether a CDSS is actually used in the real world depends on how flexible it is in meeting varying user requirements and *modus operandi*; this paper reports a method that should help improve its chances of adoption.

Acknowledgment

This research was partly supported by the Judi Meadows Memorial Fund and the European Commission through the 7th FP project ADVANCE (`http://www.advance-logistics.eu`) under grant No. 257398.

References

[1] GRiST, "Galatean risk and safety tool," www.egrist.org, 2013, accessed May 15th.

[2] C. D. Buckingham and A. Adams, "The grist web-based decision support system for mental-health risk assessment and management," in *Proceedings of the First BCS Health in Wales/ehi$_2$ joint Workshop*, 2011, pp. 37–40.

[3] R. M. Tackley, "The National Programme for Information Technology in the NHS," *Anaesthesia & intensive care medicine*, vol. 5, no. 12, pp. 400 – 401, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/B8CXB-4MBDBCS-3/2/343b3d085e232189d0974d9eba5fe0ef

[4] Department of Health, *Equity and excellence: liberating the NHS*. London: DH Publications, 2010.

[5] Health Care Commisssion, *Learning from investigations*. London: Commission for Healthcare Audit and Inspection, 2008.

[6] Centre for Suicide Prevention, "Independent homicide investigations: the national confidential inquiry into suicide and homicide by people with mental illness," University of Manchester, Tech. Rep., 2008.

[7] N. Rose, "Oxford serious incident review: 7 years on," *The Psychiatrist*, vol. 32, pp. 307–309, 2008.

[8] Department of Health, *The ten essential shared capabilities - a framework for the whole of the mental health workforce*. London: DH Publications, 2004.

[9] ——, *Best practice competencies and capabilities for pre-registration mental health nurses in England: the chief nursing officer's review of mental health nursing*. London: DH Publications, 2006.

[10] E. Gilbert, A. Adams, and C. D. Buckingham, "Examining the relationship between risk assessment and risk management in mental health," *Psychiatric and Mental Health Nursing*, no. 10, pp. 862–868, 2011.

[11] C. D. Buckingham, "Psychological cue use and implications for a clinical decision support system," *Medical Informatics and the Internet in Medicine*, vol. 27, no. 4, pp. 237–251, 2002.

[12] H. Freyhoff, H. Gruber, and A. Ziegler, "Expertise and hierarchical knowledge representation in chess," *Psychological Research*, vol. 54, pp. 32–37, 1992.

[13] O. Akin, *Models of Architectural Knowledge*. London: Pion, 1980.

[14] D. E. Egan and B. J. Schwartz, "Chunking in recall of symbolic drawings," *Memory & Cognition*, vol. 7, pp. 149–158, 1979.

[15] V. L. Patel and J. F. Arocha, "The nature of constraints on collaborative decision-making in health care settings," in *Linking expertise and naturalistic decision making*, E. Salas and G. A. Klein, Eds. Mahwah, NJ: Erlbaum, 2001, pp. 383–405.

[16] G. Cohen, "Hierarchical models in cognition: Do they have psychological reality?" *European Journal of Cognitive Psychology*, vol. 12, no. 1, pp. 1 – 36, 2000.

[17] J.Z.Tsien, "The memory code," *Scientific American*, pp. 52–59, June 2007.

[18] M. Declercq and J. De Houwer, "Evidence for a hierarchical structure underlying avoidance behavior," *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 35, pp. 123–128, 2009.

[19] A.-M. Chang, C. W. Holsapple, and A. B. Whinston, "A hyperknowledge framework of decision support systems," *Information Processing & Management*, vol. 30, no. 4, pp. 473 – 498, 1994.

[20] G. Lindgaard, C. Pyper, M. Frize, and R. Walker, "Does bayes have it? decision support systems in diagnostic medicine," *International Journal of Industrial Ergonomics*, vol. 39, no. 3, pp. 524 – 532, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6V31-4VJJW5J-2/2/1b62c3ff120a78426387049fe33c9de4

[21] W. D. Gray, "Cognitive modeling for cognitive engineering," in *The Cambridge handbook of computational psychology*, R. Sun, Ed. Cambridge: Cambridge University Press, 2008, pp. 565–588.

[22] G. Gigerenzer, *Adaptive thinking: rationality in the real world*, ser. Evolution and Cognition. Oxford University Press, 2000.

[23] W. J. Clancy, *Situated cognition*. Cambridge: Cambridge University Press, 1997.

[24] J. J. Gibson, *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum, 1979/1986.

[25] H. Heft, *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum, 2005.

[26] T. Menzies and W. Clancey, "Editorial: the challenge of situated cognition for symbolic knowledge-based systems," *International Journal of Human-Computer Studies*, vol. 49, no. 6, pp. 767–769, 1998.

[27] V. Patel, E. Shortliffe, M. Stefanelli, P. Szolovits, M. Berthold, R. Bellazzi, and A. Abu-Hanna, "The coming of age of artificial intelligence in medicine," *Artificial Intelligence in Medicine*, vol. 46, no. 1, pp. 5–17, 2009.

[28] M. A. Musen, "Technology for building intelligent systems: From psychology to engineering," in *Nebraska Symposium on Motivation*, B. Shuart, W. Spaulding, and J. Poland, Eds. Lincoln, Nebraska: U Nebraska P, 2009, vol. 52, pp. 145–184.

[29] A. Jalote-Parmar, P. Badke-Schaub, W. Ali, and E. Samset, "Cognitive processes as integrative component for developing expert decision-making systems: A workflow centered framework," *Journal of Biomedical Informatics*, vol. 43, no. 1, pp. 60 – 74, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6WHD-4WS2HXV-1/2/abbdd78b426b1f0c91d8684316a8eceb

[30] C. Brewster and K. O'Hara, "Knowledge representation with ontologies: Present challenges–Future possibilities," *International Journal of Human-Computer Studies*, vol. 65, no. 7, pp. 563–568, 2007.

[31] D. L. McGuinness and F. van Harmelen, "OWL web ontology language overview," W3C, W3C Recommendation, Feb. 2004, http://www.w3.org/TR/owl-features/.

[32] F. Villa, I. Athanasiadis, and A. Rizzoli, "Modelling with knowledge: A review of emerging semantic approaches to environmental modelling," *Environmental Modelling & Software*, vol. 24, no. 5, pp. 577–587, 2009.

[33] A. Sloman, "Putting the Pieces Together Again," in *The Cambridge handbook of computational psychology*, R. Sun, Ed. Cambridge: Cambridge University Press, 2008, pp. 684–709.

[34] B. Hu, S. Dasmahapatra, D. Dupplaw, P. Lewis, and N. Shadbolt, "Reflections on a medical ontology," *International journal of human-computer studies*, vol. 65, no. 7, pp. 569–582, 2007.

[35] Y. Wilks, "The Semantic Web: Apotheosis of annotation, but what are its semantics?" *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 41–49, 2008.

[36] World Health Organization, *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: World Health Organisation, 1992.

[37] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 4th ed. American Psychiatric Association, 2000.

[38] J. S. Kola, J. Harris, S. Lawrie, A. Rector, C. Goble, and M. Martone, "Towards an ontology for psychosis," *Cognitive Systems Research*, vol. 11, no. 1, pp. 42 – 52, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6W6C-4TBXGJ7-1/2/53faa35b5d32dc03e1cd54cb9c72d349

[39] B. Kaplan, "Evaluating informatics applications - clinical decision support systems literature review," *International Journal of Medical Informatics*, vol. 64, pp. 15–37, 2001.

[40] M. Musen, "Ontologies: Necessary–Indeed Essential–but Not Sufficient," *IEEE Intelligent Systems*, vol. 19, no. 1, pp. 77–79, 2004, january–February issue.

[41] M. K. Smith, C. Welty, and D. L. McGuinness, "OWL web ontology language guide," W3C, W3C Recommendation, Feb 2004, http://www.w3.org/TR/owl-guide/.

[42] T. Buzan, *The Mind Map Book*. BBC Consumer Publishing: London, 2003.

[43] Freemind, http://freemind.sourceforge.net/wiki/index.php/Main_Page, 2013, url accessed in May, 2013.

[44] C. D. Buckingham, A. Ahmed, and A. E. Adams, "Using XML and XSLT for flexible elicitation of mental-health risk knowledge," *Medical Informatics and the Internet in Medicine*, vol. 32, no. 1, pp. 65–81, 2007.

[45] C. D. Buckingham, A. E. Adams, and C. Mace, "Cues and knowledge structures used by mental-health professionals when making risk assessments," *Journal of Mental Health*, vol. 17, no. 3, pp. 299–314, 2008.

[46] G. Gigerenzer, U. Hoffrage, and D. Goldstein, "Fast and frugal heuristics are plausible models of cognition: Reply to Dougher ty, Franco-Watkins, and Thomas (2008)." *Psychological Review*, vol. 115, no. 1, pp. 230–237, 2008.

[47] R. Hertwig and U. Hoffrage, *Simple heuristics in a social world*. New York: Oxford University Press, 2013.

[48] W3C, "Owl," www.w3.org/TR/rif-rdf-owl, accessed 21-5-2013.

[49] S. Buckingham-Shum, "Contentious, dynamic, multimodal domains ... and ontologies?" *IEEE Intelligent Systems*, vol. 19, no. 1, pp. 80–81, 2004, january–February issue.

[50] C. D. Buckingham, P. Buijs, P. G. Welch, A. Kumar, and A. Ahmed, "Developing a cognitive model of decision-making to support members of hub-and-spoke logistics networks," in *Proceedings of the 14th International Conference on Modern Information Technology in the Innovation Processes of the Industrial Enterprises*, L. M. Elisabeth Ilie-Zudor, Zsolt Kemény, Ed. Hungarian Academy of Sciences, Computer and Automation Research Institute, 2012, pp. 14–30. [Online]. Available: igor.xen.emi.sztaki.hu/mitip/media/MITIP2012_proceedings.pdf

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 219–223

Towards Determining Syntactic Complexity of Visual Stimuli Used in Art Therapy

Bolesław Jaskuła, Jarosław Szkoła
University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszów, Poland
Email: {bjaskula, jszkola}@wsiz.rzeszow.pl

Krzysztof Pancerz
University of Management and Administration
Akademicka Str. 4, 22-400 Zamość, Poland
Email: kpancerz@wszia.edu.pl
University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszów, Poland

***Abstract*—In the paper, we deal with the problem of automatic determining syntactic complexity of visual stimuli. This problem is important in case of using paintings in eye-tracking based diagnosis and therapy of some kinds of neuropsychological and emotional disorders. Our approach to solving the considered problem is based on the clustering procedure using Self Organizing Feature Maps. The clustering results are compared with the heat maps obtained in the eye-tracking process.**

I. Introduction

ART therapy is based on the idea that the creative process of art making is healing and life enhancing and is a form of nonverbal communication of thoughts and feelings. Like other forms of psychotherapy and counseling, it is used to encourage personal growth, increase self-understanding, and assist in emotional reparation, and has been employed in a wide variety of settings with children, adults, families, and groups. Art therapy supports the belief that all individuals have the capacity to express themselves creatively and that the product is less important than the therapeutic process involved [1]. However, not only art making but also art viewing can have a therapeutic influence. Participating in the arts and viewing the arts have been found to have tremendous therapeutic impact [2]. Ulrich conducted a significant study of psychiatric patients' response to art from an extensive and diverse collection of wall-mounted paintings and prints [3].

Constant research is needed for increasing effectiveness of art diagnosis and therapy. Research in the area of neuroaesthetics is an example of this type of activities. As its name implies, neuroaesthetics is an attempt to combine neurological research with aesthetics by investigating the experience of beauty and appreciation of art on the level of brain functions and mental states. The first approach relies on observation of subjects viewing art samples and inspection of the mechanism of vision, with the aim of inducing general rules about aesthetics. This is the most popular approach to neuroaesthetics proposed by Zeki [12].The second approach aims at establishing the link between certain brain areas and artistic activity. In contrast to approaches focusing on the artistic abilities and creativity, the third approach investigates aesthetic enjoyment through brain-imaging experiments on subjects looking at pictures. A fundamental methodological crux for all these approaches is whether the aesthetic judgments are perceived as bottom-up processes driven by neural primitives or as top-down processes with high-level correlates [4].

Conclusions presented above enable us to hope that the visual art can be an effective tool in the diagnosis and therapy process, for example, in the treatment of some kinds of neuropsychological and emotional disorders. In order to do that, we need research systematizing the methodology of application of the art as a therapeutic tool. The first step of our research has been presented in [11]

II. Syntactic Complexity of Visual Stimuli

As it was mentioned earlier, painting perception is connected with the activity of a number of regions of the brain. A structure of visual stimuli, i.e., its complexity, influences which regions of the brain are activated by visual stimuli (painting), i.e., which cognitive functions (basic or higher) are initiated by the patient. Therefore, an important goal of conducted research is categorization of visual stimuli according to their complexity, i.e., their usefulness for diagnosis or therapy in the treatment of different kinds of neuropsychological and emotional disorders.

Paintings can represent many things and be analyzed on various syntactic and semantic levels [5] (cf. Figure 1). Jaimes and Chang [6] developed an overview structure of these content levels for the indexing of images. Jaimes and Chang's Index Pyramid consists of four syntactic and six semantic levels, where the width of each level gives an indication of the amount of knowledge required to describe the image content on that level. Even though the lower levels consist of upper levels, single levels can be seen as individual parts.

Syntactic levels define the visual elements such as colors and lines and are in close relation to visual perception. Semantic levels are concentrated on visual concepts and they define the meanings of the visual elements and of their arrangements. As can be seen from the Pyramid, semantics can be observed on the general, specific or abstract level. Jörgensen [7] divides painting attributes into three different groups; perceptual, reactive and interpretive attributes. Perceptual attributes are closely related to visual stimuli, e.g. the color "red". Reactive attributes are peoples' personal reactions to paintings, e.g. uncertainty, confusion and liking the image.

978-83-60810-53-8/$25.00

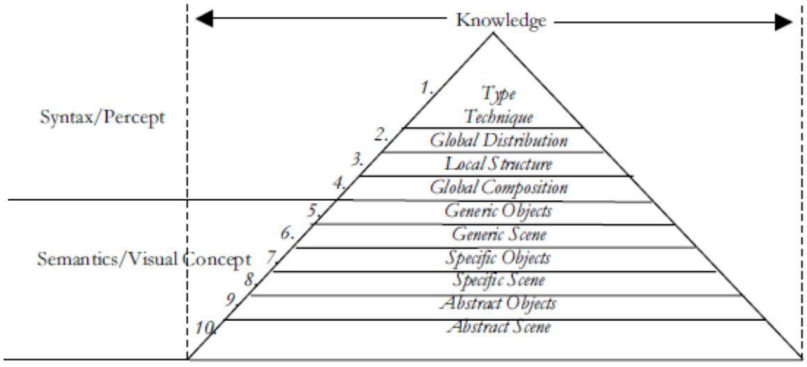


Fig. 1. The ten-level visual structure

Interpretive attributes include features like sentiment, abstract concepts and content elements like action and function.

The image complexity is a sum of syntactic and semantic complexity. At the most basic level, we are interested in the general visual characteristics of the image or the video sequence. Descriptions of the type of image or video sequence or the technique used to produce it are very general, but prove to be of great importance. Images, for example, may be placed in categories such as painting, black and white, color photograph, and drawing [6]. Global distribution aims to classify images or video sequences based on their global content and is measured in terms of low-level perceptual features such as spectral sensitivity (color), and frequency sensitivity (texture). Individual components of the content are not processed at this level. Global distribution features, therefore, may include global color (e.g., dominant color, average, histogram), global texture (e.g., coarseness, directionality, contrast), global shape (e.g. aspect ratio), global motion (e.g. speed, acceleration, and trajectory), camera motion, global deformation (e.g. growing speed), and temporal/spatial dimensions (e.g. spatial area and temporal dimension), among others. In contrast to Global Structure, which does not provide any information about the individual parts of the image or the video sequence, the Local Structure level is concerned with the extraction and characterization of the image's components. At the most basic level, those components result from low-level processing and include elements such as the Dot, Line, Tone, Color, and Texture. At this level, we are interested in the specific arrangement of the basic elements given by the local structure, but the focus is on the Global Composition. In other words, we analyze the image as a whole, but use the basic elements described above (line, circle, etc.) for the analysis.

One of the key goals realized by the observer at the syntactic level is to detect image contours. One common attribute of paintings in a broad array of artistic traditions, starting from the earliest surviving depictions on cave walls, is the use of boundary lines to depict the edges of objects. There are no actual contour lines dividing real objects from their backgrounds in most cases, which raises the question of why contour lines are so ubiquitous and effective in depiction. One theory is that line drawings are a convention that are imposed within a particular culture and passed down through learning [8].

Recapping, image complexity at the syntactic level is the degree of cognitive effort to which the observer is forced by the structure (arrangement) of the painting obtained by means of adequate painting techniques at the level of receiving physical stimuli. The degree of this type of complexity can be determined in terms of uncertainty or redundancy.

III. Procedure

Human eyes are unable to observe the whole image with the equal acuity. The area of acute vision covers a field of less than 1.5° of arc. A standard eye-tracking examination is performed within 50 cm of the screen. An image has a horizontal resolution of 1200 - 1600 pixels. Therefore, the area of acute vision covers a field with a diameter of 25 pixels. In the performed examinations, the size of a segment has been selected proportionally to the image with a horizontal resolution of 1200 pixels. Art perception is dependent on physical features of the human eye. Therefore, splitting the painting into smaller parts is important in syntactic processing. Research carried out using eye-tracking shows that data processing is realized in the form of fixations followed by saccades (i.e., transitions between fixations). We have created a specialized computer tool for the objective analysis of painting complexity - entropy. The painting with a low information content has a small value of entropy. According to information theory, the smaller probability of object occurrence, the greater information. Results obtained by means of the created tool have been compared with results obtained by means of eye-tracking in the form of the so-called heat maps. Heat maps represent the fixation locations and the duration of the fixations. The regions indicated by our tool, in most cases, are covered by experimental results obtained in the eye-tracking process. On

this basis, one can determine that in the syntactic analysis of the painting, most vision fixations concern regions whose structures differ from other regions of the same painting. It is the foundation of the syntactic analysis. The main advantage of the computer tool is independence from individual features. Moreover, results are not sensitive to subjective factors.

The approach presented in this paper is based on a clustering procedure using Self Organizing Feature Maps. The concept of a Self-Organizing Feature Map (SOM) was originally developed by Kohonen [9]. SOMs are neural networks composed of a two-dimensional grid (matrix) of artificial neurons that attempt to show high-dimensional data in a low-dimensional structure. Each neuron is equipped with modifiable connections. Self-Organizing Feature Maps possess interesting characteristics such as self-organization, i.e., networks organize themselves to form useful information, as well as competitive learning, i.e., network neurons compete with each other. The winners of the competition strengthen their weights while the losers' weights are unchanged or weakened. SOMs are used in a clustering process. Input data are vectors composed of m features (elements).

In our approach, each SOM is a feedforward three-layer neural network, one layer for one color component (R, G, or B). This architecture can also be used for other color models, e.g. YUV.

The training process of SOMs can be presented as a list of the following steps:

1) **Initialization**. At the beginning, weights are initialized with random values from a given interval (w_{min}, w_{max}), i.e.:
$$map(i,j,k) = random(w_{min}, w_{max}),$$
where k determines a layer of the multilayer map, i and j are coordinates in the k-th map layer. An initial value for a learning rate α is equal to α_{top}. During the learning process, a value of α is changed from α_{top} to α_{bottom} with step α_{step}, where p is a number of epochs. The initial size of the map is equal to $s_{min} \times s_{min}$ of neurons. This size is progressively increased to $s_{max} \times s_{max}$, where $s_{max} = max(s_{min}, \sqrt{2n} + 1)$.
2) **Reading input data**. Input vectors are normalized to the interval $[0, 1]$. All input vectors have the same dimension, i.e., m.
3) **Iterations**.
   a) At each iteration, a random input vector is entered to the input layer $x_{current} = \mathbf{x}(random(1, n))$.
   b) A winning Kohonen neuron is determined, i.e., the neurons compete on the basis of which of them have their associated weight vectors "closest" to $x_{current}$. The winner is selected on the basis of minimization of the mean squared error, i.e.:
   $$\min_{i,j} \sum_{k=0}^{m} \left(x_{current}(k) - map(i,j,k)\right)^2$$
   c) For the winner and direct neighbours only, weights are modified according to:
      - winner:
      $$map(i,j,k) = \alpha \left(x_{current}(k) - map(i,j,k)\right),$$
      - direct neighbours:
      $$map(i,j) = 0.6\alpha \left(x_{current} - map(i,j)\right).$$
   d) The learning rate is updated according to $\alpha_{e+1} = \alpha_{top} - e\alpha_{step}$, where e is the current epoch number indicator.
   e) The size of the map is updated:
   $$s_{e+1} = s_{max} \frac{e}{p},$$
   if s_{e+1} is greater than the current size, where p is a number of epochs.
   f) If the size of the map has been changed, weights are updated according to:
   $$map_{new}(i,j,k) = \\ = \sum_{neighbours\ of\ (i,j)} 0.6 map_{neighbour}(i,j,k) + \\ + map(i,j,k).$$

In a standard algorithm, we can distinguish the following steps. An image is recorded in the RGB format. Next, SOM is trained on the basis of the RGB components, separately for each component. However, in each step, weights of maps are also averaged (using the weighted average) for color components. Application of SOMs causes generalization of relationships between pixels of an input image by grouping similar regions close to each other. SOMs are widely used in problems demanding reduction of input data dimension. The obtained map represents classes into which input data space can be divided. In most cases, this is the last step of grouping data. The disadvantage of such an approach is a region of tolerance too broad for neighborhoods of individual classes. Generally, for each class, only its centroid is indicated.

In this paper we propose to extend the standard algorithm:

1) An input image is split into segments constituting a grid of the size $N \times N$. The size of the individual segment depends on the size of the image.
2) Each segment is independently processed. Data for a given segment are passed to the input of SOM. As a result, we obtain maps (i.e., layers of SOM) for each component: R, G, and B. SOM is trained using the procedure described earlier.
3) On the basis of maps for components, a minimal spanning tree (MST) is created using the weighted average of color components. The spanning tree includes information about correlations of pixels of the input image:
$$tree[i][j] = \begin{cases} w_{ij} & \text{if } (i,j) \text{ belongs to MST,} \\ 0 & \text{otherwise,} \end{cases}$$
where i and j are pixel indexes, w_{ij} is a correlation factor of pixels.
4) The coefficient $C_{seg}(m,n)$ is calculated for each segment (see Algorithm 1) with coordinates m and n. In

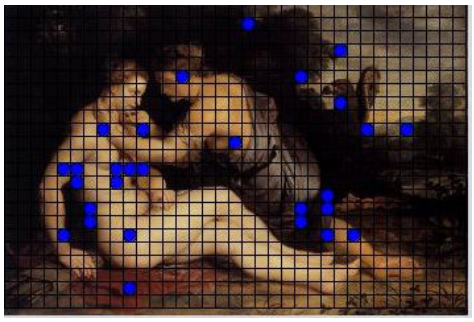
a)  b) 

Fig. 2. Example 1: a) the clustering result; b) the heat map

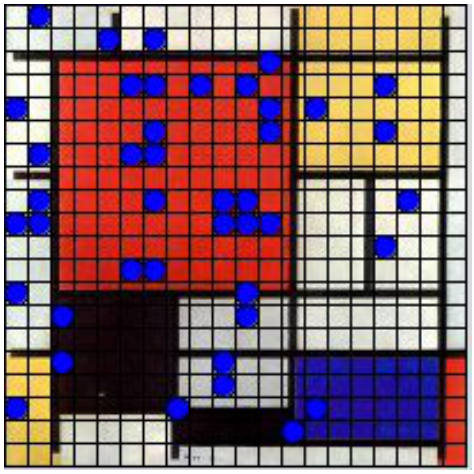
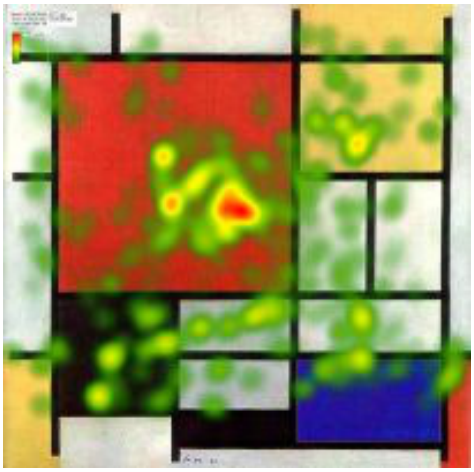
a)  b) 

Fig. 3. Example 2: a) the clustering result; b) the heat map

Algorithm 1: Algorithm for calculating the coefficient $C_{seg}(m,n)$

```
C_seg(m,n) ← 0;
for each pixel i do
    for each pixel j do
        if tree[i][j] > CORR_min and
        dist(i,j) < DIST then
            C_seg(m,n) ← C_seg(m,n) + 1;
        end
    end
end
```

this algorithm, pairs of pixels are compared. Moreover, $CORR_{min}$ is a threshold correlation between pixels and $DIST$ is a threshold distance between pixels (for example, the Manhattan distance).
This step enables us to omit pixels which potentially are not important for visual perception. Moreover, correlation between pixels which are far apart is not taken into consideration. Thresholds $CORR_{min}$ and $DIST$ have been determined experimentally, and they are equal to:

$$CORR_{min} \approx 2,$$

$$DIST = \frac{D_{segm}}{2},$$

where D_{segm} is the width of the segment.
The greater value of the coefficient $C_{seg}(m,n)$ indicates a more complex structure of the segment, e.g. gradients, colors, shapes, etc.

5) We search for numbers of segments with the same coefficient $C_{seg}(m,n)$. Let $VAL_{C_{seg}}$ be a set of all values of coefficients $C_{seg}(m,n)$ calculated in the previous step. For each value v in $VAL_{C_{seg}}$, we calculate its number $OCC_{C_{seg}}(v)$ of occurrences.
6) We calculate the entropy for the analyzed image:

$$H = -\sum_{v=1}^{k} p(v) \ln p(v),$$

where:

- k is a cardinality of $VAL_{C_{seg}}$,
- $p(v) = \frac{OCC_{C_{seg}}(v)}{N}$,
- $N = \sum_{v=1}^{k} OCC_{C_{seg}}(v)$.

The greater entropy indicates that the complexity of the image is greater, i.e., a number of segments with different coefficients $C_{seg}(m,n)$ is greater.

a)  b) 

Fig. 4. Example 3: a) the clustering result; b) the heat map

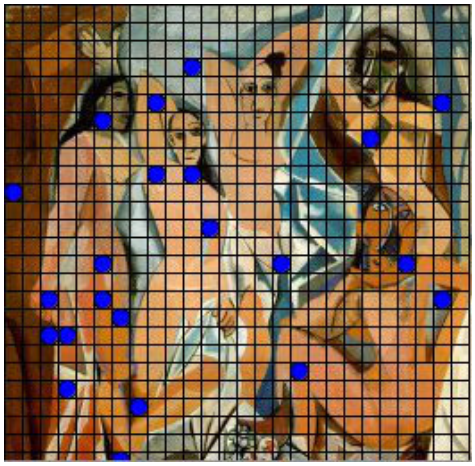
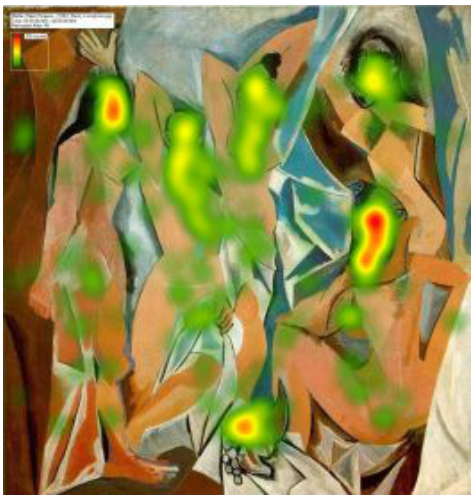
a)  b) 

Fig. 5. Example 4: a) the clustering result; b) the heat map

IV. Experiments

Experiments have been carried out using some well known paintings. Figures 2 - 5 present exemplary comparisons of the clustering results and the heat maps obtained in the eye-tracking process.

V. Conclusions and Further Works

In the paper, we have shown the computer tool based on Self-Organizing Feature Maps enabling us to automatically determine complexity of visual stimuli used in syntactic analysis. Such a method is necessary in diagnosis and therapy of some neuropsychological and emotional disorders using eye-tracking. This problem is the main task for our future work.

References

[1] C. A. Malchiodi, Ed., *Handbook of Art Therapy*. New York: The Guilford Press, 2003.

[2] J. Rollins, J. Sonke, C. R., A. Boles, and J. Li, Eds., *State of the Field Report: Arts in Healthcare*. Washington, DC: Society for Arts in Healthcare, 2009.

[3] R.S. Ulrich, "Effects of interior design on wellness: theory and recent scientific research," *Journal of Health Care Interior Design*, vol. 3, pp. 97–109, 1991.

[4] A. A. A. Salah and A. A. Salah, "Technoscience art: A bridge between neuroesthetics and art history?" *Review of General Psychology*, vol. 12, no. 2, pp. 147–158, 2008.

[5] V.-M. Saarinen, M. Laine-Hernandez, and H. Saarelma, "The influence of image content levels and looking type on eye movements," *Graphic Arts in Finland*, vol. 35, no. 2, pp. 1–12, 2006.

[6] A. Jaimes and S. fu Chang, "A conceptual framework for indexing visual information at multiple levels," in *Proceedings of SPIE INTERNET IMAGING 2000*, 2000, pp. 2–15.

[7] C. Jörgensen, "Indexing Images: Testing an Image Description Template," in *ASIS Annual Conference Proceedings*, 1996.

[8] D. Melcher and P. Cavanagh, "Pictorial cues in art and in visual perception," in *Art and the senses*, F. Bacci and D. Melcher, Eds. Oxford University Press, 2011, pp. 359–394.

[9] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[10] B. Jaskuła and K. Pancerz, "Toward interactive computer systems based on eye-tracking technology modernizing didactics of visual art perception," *CyberEmpathy: Visual Communication and New Media in Art, Science, Humanities, Design and Technology*, vol. 1, 2012.

[11] B. Jaskuła, J. Szkoła, and K. Pancerz, "SOM Based Segmentation of Visual Stimuli in Diagnosis and Therapy of Neuropsychological Disorders," in *Proceedings of the International Conference on Man-Machine Interactions*, 2013.

[12] S. Zeki, Ed., *Inner vision: an exploration of art and the brain*. Oxford University Press, 1999.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 225–230

Simulating of Schistosomatidae (Trematoda: Digenea) Behavior by Physarum Spatial Logic

Andrew Schumann
University of Information Technology and Management in Rzeszow, ul. Sucharskiego 2, 35-225 Rzeszów, Poland
Email: Andrew.Schumann@gmail.com

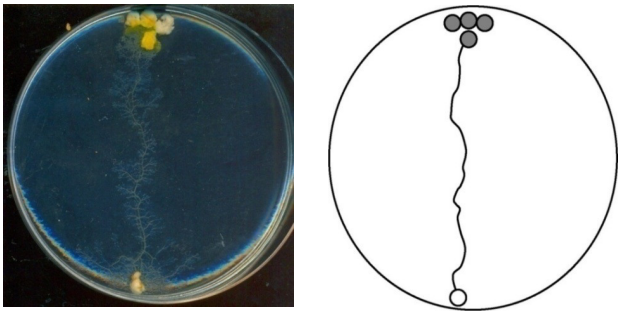
Ludmila Akimova
Scientific and Practical Center of the National Academy of Scientific on Bioresources, Akademicheskaya str. 27, 220072 Minsk, Belarus Email: akimova_minsk@mail.ru

***Abstract*—In this paper we consider possibilities of simulating behavior of the group of trematode larvae (miracidiae and cercariae) by the abstract slime mould based computer that is programmed by attractants and repellents. For describing this simulation, we appeal to the language which is a kind of π-calculus called Physarum spatial logic. This language contains labels for attractants and repellents. Taking into account the fact that the behavior of miracidiae and cercariae can be programmed only by attractants (repellents for miracidiae and cercariae are not known still), we can claim that the behavior of miracidiae and cercariae is a restricted (poorer) form of Physarum spatial logic.**

I. Introduction

IN *Physarum Chip Project: Growing Computers From Slime Mould* [3] supported by FP7 we are going to implement programmable amorphous biological computers in plasmodium of Physarum. This abstract computer we are going to obtain is called *slime mould based computer*. The plasmodium behaves and moves as a giant amoeba and its behavior can be considered as a biological implementation of Kolmogorov-Uspensky machines [2]. This allows us to use the plasmodium of Physarum for solving different tasks that can be solved in Kolmogorov-Uspensky machines as well. The slime mould based computer is programmed using attractants and repellents (fig. 1). On the one hand, it was experimentally proved that the slime mould prefers substances with potentially high nutritional value, e.g. it is attracted by peptones, aminoacids phenylalanine, leucine, serine, asparagine, glycine, alanine, aspartate, glutamate, and threonine. On the other hand, repellents for Physarum polycephalum can be presented by some illumination-, thermo- and salt-based conditions. Usually the plasmodium forms a congregation of protoplasm in food sources to surround them, secret enzymes and digest the food. Slime mould based computer can be regarded as a parallel computing substrate complementary to existing massive-parallel reaction-diffusion chemical processors [1].

In papers [15; 16] we showed that the behavior of plasmodium of *Physarum polycephalum* has an own spatial logic which is one of the natural implementations of π-calculus. This logic called *Physarum spatial logic* can be used as a programming language for the slime mould based computer. Taking into account the fact that within π-calculus we can formalize and describe different concurrent processes, within Physarum spatial logic we can do the same as well.



(a) (b)

Fig. 1 An 'ideal' example of plasmodium attracted by source of nutrients. Initially an oat flake colonized by plasmodium is placed in the southern part of Petri dish, and a group of intact oat flakes in the northern part. Plasmodium propagates towards the intact flakes and occupies them. (a) Snapshot the experimental Petri dish with the plasmodium. (b) Scheme of the plasmodium attraction: initial position of the plasmodium is shown by circle, newly occupied oat flakes (attractants) by solid discs, trajectory of the plasmodium by curve. This figure is from the book [20]. *Courtesy of Andy Adamatzky*

In this paper we will show that the behavior of local group of the genus *Trichobilharzia* Skrjabin & Zakharov, 1920 (Schistosomatidae Stiles & Hassall, 1898) can be simulated by Physarum logic. This means that, first, a local group of Schistosomatidae can behave as a programmable biological computer, second, a biologized kind of π-calculus such as Physarum spatial logic can describe concurrent biological processes at all.

II. Physarum Spatial Logic

In this section we will consider some basics of Physarum spatial logic.

The behavior of Physarum plasmodium can be divided into the following elementary processes: inaction, fusion, cooperation, and choice, which could be interpreted as unconventional (spatial) falsity, conjunction, weak and strong disjunction respectively, denoted by *Nil*, &, |, and +. These

This research is being fulfilled by the support of FP7-ICT-2011-8.

operations differ from conventional ones, because they cannot have a denotational semantics in the standard way. However, they may be described as special spatial transitions over states of Physarum machine: inaction (*Nil*) means that pseudopodia has just stopped to behave; fusion (&) means that two pseudopodia come in contact one with another and then merge; cooperation (|) means that two pseudopodia behave concurrently; choice (+) means a competition between two pseudopodia in their behaviors.

A π-calculus for describing the dynamics of Physarum machine is presented as a labeled transition system with some logical relations.

Assume that there are N active species or growing pseudopodia and the state of species i is denoted by $p_i \in$ L. These states are time dependent and they are changed by plasmodium's active zones interacting with each other and affected by attractants or repellents. Plasmodium's active zones interact concurrently and in a parallel manner. Foraging plasmodium can be represented as a set of the following abstract entities.

- The set of actions (growing pseudopodia), T' = {α, β, ...}, localized in *active zones*. The actions from T' are called the *simplest transitions*, the latter are defined as $\{p \xrightarrow{\alpha} q : p, q \in L, \alpha \in T'\}$. Notice that we also have transitions that do not belong to T'. Assume that the set of all transitions is denoted by T.

On a nutrient-rich substrate plasmodium propagates as a typical circular, target wave, while on the nutrient-poor substrates localized wave-fragments are formed. Each action $\alpha \in$ T', starts on a state p_i, which is its current position, and says about a transition (propagation) of a state p_i to another state of the same or another computation cell. Part of plasmodium feeding on a source of nutrients may not propagate, so its transition may be *Nil*, but this part can always start moving.

- The set of *attractants* {A_1, A_2, ...} $\subset$ T \ T' are sources of nutrients, on which the plasmodium feeds. It is still subject of discussion how exactly plasmodium feels presence of attractants. Each attractant A is a function from T' to T'.
- The set of *repellents* {R_1, R_2, ...} $\subset$ T \ T'. Plasmodium of Physarum avoids light and some thermo- and salt-based conditions. Thus, domains of high illumination are repellents such that each repellent R is characterized by its position and intensity of illumination, or force of repelling. In other words, each repellent R is a function from T' to T'.
- The set of *protoplasmic tubes* {C_1, C_2, ...} $\subset$ T \ T'. Typically plasmodium spans sources of nutrients with protoplasmic tubes/veins. The plasmodium builds a planar graph, where nodes are sources of nutrients, e.g. oat flakes, and edges are protoplasmic tubes. $C(\alpha)$ means a diffusion of growing pseudopodia $\alpha \in$ T'.

Hence, T = T' $\cup$ {A_1, A_2, ...} $\cup$ {R_1, R_2, ...} $\cup$ {C_1, C_2, ...}.

Our process calculus contains the following basic operators: *Nil* (inaction), $\cdot$ (prefix), | (cooperation), \ (hiding), & (reaction/fusion), + (choice), a (constant or restriction to a stable state), $A(\cdot)$ (attraction), $R(\cdot)$ (repelling), $C(\cdot)$ (spreading/diffusion). Let T = {a, b, ...}, the set of all actions (evidently, this set is finite), be considered as a set of names. A name refers to a communication link or channel. With every $a \in$ T we associate a complementary action $\overline{a}$. Let us suppose that a designates an input port and $\overline{a}$ designates an output port. Any behavior of Physarum will be considered as outputs and any form of outside control and stimuli as appropriate inputs. For instance, T' is to be regarded just as the set of output ports and thereby T \ T' contains ports that can be interpreted under different conditions as input ports or output ports. So, for each $X \in$ {A_1, A_2, ...} $\cup$ {R_1, R_2, ...}, we take that $X(\gamma) = \delta$ is a function from T' to T' such that X is a stimulus for γ and X makes an output $\delta \in$ T' along γ. Evidently that $X(Nil) = Nil$. Let $X(\gamma)$ denote an input and $\overline{X(\gamma)}$ an output.

Define **L** be the set of labels built on T (under this interpretation, $a = \overline{\overline{a}}$). Suppose that an action a communicates with its complement $\overline{a}$ to produce the internal action τ and τ belongs to **L**, too.

We use now the symbols γ, δ, … , etc., to range over labels (actions), with $a = \overline{\overline{a}}$, and the symbols P, Q, etc., to range over processes on states p_i. The processes are given by the syntax:

$$P, Q ::= Nil \mid \gamma.P \mid A(\gamma).P \mid \overline{A(\gamma)}.P \mid R(\gamma).P \mid \overline{R(\gamma)}.P \mid C(\gamma).P \mid (P \mid Q) \mid P \setminus Q \mid P \,\&\, Q \mid P + Q \mid a$$

Each label is a process, but not vice versa, because a process may consists of many labels combined by the basic operators.

An operational semantics for this syntax is defined as follows: $\gamma ::= p_i$, where $p_i \in$ L.

$$\text{Prefix}: \frac{}{\gamma.P \xrightarrow{\gamma} P},$$

$$\frac{}{A(\gamma).P \xrightarrow{\delta} P}\ (A(\gamma) = \delta), \quad \frac{}{\overline{A(\gamma)}.P \xrightarrow{\delta} P}\ (\overline{A(\gamma)} = \delta),$$

$$\frac{}{R(\gamma).P \xrightarrow{\delta} P}\ (R(\gamma) = \delta), \quad \frac{}{\overline{R(\gamma)}.P \xrightarrow{\delta} P}\ (\overline{R(\gamma)} = \delta),$$

(the conclusion states that the process of the form $\gamma.P$ (resp. $A(\gamma).P$, $\overline{A(\gamma)}.P$, $R(\gamma).P$, $\overline{R(\gamma)}.P$) may engage in γ (resp. $A(\gamma)$, $\overline{A(\gamma)}$, $R(\gamma)$, $\overline{R(\gamma)}$) and thereafter they behave like P; in the presentations of behaviors as trees, $\gamma.P$ (resp. $A(\gamma).P$, $\overline{A(\gamma)}.P$, $R(\gamma).P$, $\overline{R(\gamma)}.P$) is understood as an

edge with two nodes: γ (resp. $A(\gamma)$, $\overline{A(\gamma)}$, $R(\gamma)$, $\overline{R(\gamma)}$) and the first action of P),

$$\text{Diffusion:}\quad \frac{P \xrightarrow{\gamma} P'}{P \xrightarrow{\gamma} C(\gamma)} \quad (C(\gamma) = P'),$$

$$\text{Constant:}\quad \frac{P \xrightarrow{\gamma} P'}{a \xrightarrow{\gamma} P'} \quad (a = P, a \in L),$$

$$\text{Choice:}\quad \frac{P \xrightarrow{\gamma} P'}{P + Q \xrightarrow{\gamma} P'}, \quad \frac{Q \xrightarrow{\gamma} Q'}{P + Q \xrightarrow{\gamma} Q'},$$

(these both rules state that a system of the form $P + Q$ saves the transitions of its subsystems P and Q),

$$\text{Cooperation:}\quad \frac{P \xrightarrow{\gamma} P'}{P \mid Q \xrightarrow{\gamma} P' \mid Q}, \quad \frac{Q \xrightarrow{\gamma} Q'}{P \mid Q \xrightarrow{\gamma} P \mid Q'},$$

(according to these rules, the cooperation | interleaves the transitions of its subsystems),

$$\frac{P \xrightarrow{\gamma} P' \quad Q \xrightarrow{\bar{\gamma}} Q'}{P \mid Q \xrightarrow{\tau} P' \mid Q'},$$

(i.e. subsystems may synchronize in the internal action τ on complementary actions γ and $\bar{\gamma}$),

$$\text{Hiding}: \quad \frac{P \xrightarrow{\gamma} P'}{P \setminus Q \xrightarrow{\gamma} P' \setminus Q} \quad (\gamma \notin Q, Q \subseteq \mathbf{L}),$$

(this rule allows actions not mentioned in Q to be performed by $P \setminus Q$),

$$\text{Fusion}: \quad \frac{}{\gamma.P \,\&\, \breve{P} \xrightarrow{\gamma} Nil}$$

(the fusion of dual processes are to be performed into the inaction, e.g. a fusion of an attractant/repellent P and appropriate repellent/attractant $\breve{P}$),

$$\frac{P \xrightarrow{\gamma} P' \quad Q \xrightarrow{\gamma} P'}{P \wedge Q \xrightarrow{\gamma} P'}, \quad \frac{P \xrightarrow{\gamma} P' \quad Q \xrightarrow{\gamma} P'}{Q \wedge P \xrightarrow{\gamma} P'}$$

(this means that if we obtain the same result P' that is produced by the same action γ and evaluates from two different processes P and Q, then P' may be obtained by that action γ started from the fusion P & Q or Q & P),

$$\frac{P \xrightarrow{\gamma} P'}{P \wedge Q \xrightarrow{\gamma} Nil + C(\gamma) + P'}, \quad \frac{P \xrightarrow{\gamma} P'}{Q \wedge P \xrightarrow{\gamma} Nil + C(\gamma) + P'}$$

(these rules state that if the result P' is produced by the action γ from the processes P, then a fusion P & Q (or Q & P) is transformed by that same γ either into the inaction or diffusion or process P').

These are inference rules for basic operations. The ternary relation $P \xrightarrow{\gamma} P'$ means that the initial action P is capable of engaging in action γ and then behaving like P'.

Now we can show that in the behavior of any local group of Schistosomatidae we can observe the same elementary processes: inaction, fusion, cooperation, and choice, which are defined in the same way.

III. Life Cycle of Schistosomatidae (Trematoda : Digenea)

All representatives of subclass Digenea Carus, 1863 (Platyhelminthes: Trematoda) are exclusively endoparasites of animals. The digenean life cycle has the form of heterogony, i.e. there is a natural alternation of amphimictic (usually synarmophytous) and parthenogenetic stages. At these stages digeneae have different outward, different means of reproduction and different adaptation to different hosts. The interchange of hosts is necessary for a successful realization of digenean flukes life cycle. The majority of representatives of this subclass have a complete life cycle with participation of three hosts: intermediate, additional (metacercarial) and definitive. Molluscs are always the first intermediate hosts, while different classes of vertebrate animals are definitive hosts.

Among digeneae there is a bunch of parasites, belonging to the family of Schistosomatidae, which represents an isolated bunch which has adapted to parasitizing in the circulatory system of vertebrate animals. Puberal representatives of this family are diecious individuals (in other digenean families maritas are hermaphrodites). The family includes the following three subfamilies: Schistosomatinae Stiles and Hassall, 1898, they parasitize a variety of birds and mammals, including human being; Bilharziellinae Price, 1929 and Gigantobilharziinae Mehra, 1940, they parasitize birds. Representatives of the first subfamily (in particular the genus *Schistosoma* Weinland, 1858) parasitize mammals, including human being. In the tropical and subtropical countries, about 200 million persons are infected by them from which 11 thousand persons annually die because of the given infestation [13]. Representatives of the latter two subfamilies, the so-called avian schistosomatidae, have been observed on all continents, including Europe. In a puberal state they parasitize birds, however they are capable to incorporate into a human organism as nonspecific host. After they penetrate human skin, where they perish, they invoke thereby allergic dermatitis. The fact of incorporation of these larvae into nonspecific hosts invokes interest to avian schistosomatidae. Therefore the simulation of behavior of their local groups can be interesting from a medicine view, be-

cause it allows to perceive better features of digenean behavior. Simulating their behavior is possible by means of Physarum spatial logic as we will show.

The life cycle of all representatives of the family Schistosomatidae is identical, it passes with participation of two hosts.

From an egg which got to water from an organism of definitive host, a miracidia hatchs. It is a settle free-swimming larval stage of parthenogenetic generation of Schistosomatidae. Miracidia for a short span should find a molluc of a certain kind to insinuate into it, otherwise it perishes. Molluscs, thus, are attractants for miracidiae. More precisely, in respect to miracidiae the chemotaxis as attractant holds (miracidia moves towards a chemical signal proceeding from a molluc). Other kinds of attractants for miracidiae are presented by light (there is a positive phototaxis) and gravitation (negative geotaxis). We will designate all miracidian attractants by $A^m_1, A^m_2, \ldots, A^m_n$.

Miracidian repellents have not been detected still, i.e. $\{R^m_1, R^m_2, \ldots, R^m_n\} = \varnothing$.

The continuation of digenean life cycle will take place just in case a miracidia detects a mollusc for which a certain kind of digeneas has a hostal specificity [10]. Otherwise the miracidia dies. Miracidiae *Trichobilharzia szidati* Neuhaus, 1952 can look for a intermediate host only for the period of 20 h at temperature 20 °C [14].

In a body of mollusc, a miracidia undergoes metamorphosis and it is transformed into a mother sporocyst in which daughter sporocysts educe. In the latter then cercariae start to be formed. This state can be called the miracidian diffusion. We will designate these diffusions by $C^m_1, C^m_2, \ldots, C^m_n$.

Now we can construct a version of Physarum spatial logic for simulating the behavior of local groups of miracidiae. The processes have the following syntax which is defined in the way of Physarum logic:

$$P, Q ::= Nil \mid \gamma.P \mid A^m(\gamma).P \mid \overline{A^m(\gamma)}\ .P \mid$$
$$C^m(\gamma).P \mid (P \mid Q) \mid P \setminus Q \mid P \,\&\, Q \mid P + Q \mid a$$

For the simulation we need also to have two sets of actions T and T^m, where T contains actions of Physarum plasmodium, T^m includes actions of local group of miracidiae. These sets should have the same number of members (the same cardinality), namely we should have the same number of active zones (growing pseudopodia and active miracidiae), the same number of attractants, and the same number of diffusions (motions of protoplasmic tubes towards food and miracidian motions towards chemical signals of eventual hosts to transform into a maternal sporocyst). For instance, if we have five molluscs in one experimental dish with water and a suspension of miracidiae, then we can try to simulate the miracidian processes by Physarum spatial logic for stimuli of five nutrient sources with similar localizations as that for molluscs.

IV. The Behavior of Cercariae of Bird Schistosomes (Genus *Trichobilharzia*)

The cercarial behavior of bird schistosomes (family *Schistosomatidae*) is well studied due to representatives of the genus *Trichobilharzia* [11]. Their behavior is characterized by specific taxises which are referred to an effective search of necessary definitive hosts. These taxises developed by evolution of larvae of bird schistosomes allow their looking for specific hosts to be successful, forward their affixion to a surface of host body as well as their incorporation into a host cutaneous covering and their penetration into a circulatory system, where a parasite reaches sexual maturity. Thus, taxises form an enough large family of attractants for cercariae.

In a resting state, cercariae are attached to a vascular wall or on a water film by means of acetabulum. Active motions are characteristic only by the strong shaking of pot or by the water interfusion. At a weak rotation of pot it is visible that the cercarial body and its tail follow the water stream, while their acetabulums keep cercariae on the pot wall. Any continuous active motion is not observed.

Cercariae of the genus *Trichobilharzia* after leaving a mollusc actively swiming in the water for an hour. Such an active behavior of larvae after leaving a mollusc provides a cercarial allocation in water space. Then cercariae pass to a passive behavior. They are attached by a ventral sucker to a surface film of water or to various subjects near a water surface, getting thus a characteristic resting state. The resting state allows cercariae in absence of specific to them stimulants to stop the search of host and to conserve their energy.

Free-swiming cercariae need to insinuate into a definitive host during the limited time interval (1–1.5 days at temperature 24°C) since otherwise larvae perish [14].

For successful search of hosts, larvae of digeneae of *Trichobilharzia* have developed by evolution a behavior facilitating this problem. They possess a positive phototaxis, negative geotaxis, chemotaxis, and also actively react to turbulence of water [7]. It means that for cercariae there are already many other attractants.

The light sensitivity of cercariae of *Trichobilharzia* is very high. As experiments show, cercariae always move towards a light source, and then take a resting state on the lighted side of capacity in which they are. The given taxis, and also negative geotaxis allow cercariae to be kept in the nature at a surface of water in expectation of suitable hosts.

Cercariae actively react to changes in intensity of illumination (shadings) and to turbulence of water [5]. These external factors, corresponding to possible appearances of definitive hosts in water, stimulate the cercarial transition from a resting state into actions that enlarge their chances to meet hosts.

Cercariae possess a chemotaxis in relation to specific hosts. On body surfaces larvae of the genus *Trichobilharzia* have chemoreceptors which receive appropriate chemical signals proceeding from a skin of potential host. The similarity of compound of fatty acids of bird and human skin leads to that cercariae equally react to the bird and human appearance in water: they move in their direction, and then

they are attached to skin and begin penetration into it [9]. So, the chemotaxis from a skin of potential hosts (surface lipids of skin of human being and swimming bird), the positive phototaxis, the negative geotaxis and the water turbulence present cercarial attractants of different degree of appeal. We will designate these attractants by $A^c_1, A^c_2, \ldots, A^c_n$.

In experimental researches it has been shown that any attachment of cercariae of *Trichobilharzia* to skin is stimulated by cholesterol and ceramides, and incorporation into skin by linoleic and linolenic acids, all these materials are present on skin of both bird and human being [12; 8]. Thereby surface lipids of human skin invoke higher frequency of cercarial incorporations into skin, than surface lipids of birds [9]. One more reason that cercariae of *Trichobilharzia* successfully insinuate into human skin is the fact of that the skin of duck foots has thicker keratinized surface which, possibly, is more difficult for overcoming, than that of human being [8].

On the basis of experiments the rate of penetration of larvae of schistosomes *Trichobilharzia szidati* into human skin [8] has been defined. The larva begins incorporation into human skin approximately in 8 seconds (range from 0 to 80 seconds) after first contacts. The process of full penetration into skin takes about 4 minutes (range from 83 till 13 minutes 37 seconds). The given numerals testify that it is enough if the person has even a short-term contact to water where there are cercariae of bird schistosomes to give them possibility to incorporate into skin.

In some cases, for example children, cercarial larvae can "chip" skin and be brought by venous blood to lungs, invoking there hemorrhages and inflammation. If cercariae are lucky to insinuate into blood and then to lungs, the disease can get harder by the pulmonary syndrome from small cough to symptoms of bronchial obstruction [18].

At the same time, repellents for cercariae have not been found yet. For example, Ludmila Akimova's experience shows that cercarial motions towards a smaller concentration of material which invokes a destruction of larvae are not observed at all. The experience principle consists in that in a small cavity with the length of 10 cm, the width and depth of 0.5 cm there is water with a suspension of cercariae. Then a thin essential oil is added in one of the side of this small cavity. Cercariae, which are nearby, quickly perish, although other cercariae do not move aside where the reacting material is absent. Cercariae simply freely float and as soon as they appear in that part where there is the reacting material they perish. Thus, $\{R^c_1, R^c_2, \ldots, R^c_n\} = \varnothing$.

In definitive hosts cercariae reach diffusion states. We will designate these diffusions by $C^c_1, C^c_2, \ldots, C^c_n$.

The behavior of local groups of cercariae can be simulated within a version of Physarum spatial logic, where the processes have the following syntax defined in section II:

$$P, Q ::= Nil \mid \gamma.P \mid A^c(\gamma).P \mid \overline{A^c(\gamma)}\ .P \mid$$
$$C^c(\gamma).P \mid (P \mid Q) \mid P \setminus Q \mid P \,\&\, Q \mid P + Q \mid a$$

The sets of actions T and T^c, where T consists of actions of Physarum plasmodium, T^c contains actions of local group of cercariae, should have the same number of members. For example, if we have three human beings in one lake with cercariae, then we can simulate the cercarial processes by Physarum spatial logic where three nutrient sources with similar localizations as that for human beings act as stimuli. Hence, the behavior of local groups of cercariae is another biological implementation of Kolmogorov-Uspensky machines. It can build planar graphs as well.

V. Arithmetic Operations in Physarum Spatial Logic and In Schistosomatidae Behavioral Logic

We know that within π-calculus we can convert expressions from λ-calculus. In particular, it means that we can consider arithmetic operations as processes. Physarum spatial logic as well as its modification in the form of behavioral logic for local groups of miracidiae (cercariae) is a biologized version of π-calculus. Therefore we can convert arithmetic operations into processes of either Physarum spatial logic or schistosomatidae behavioral logic.

Indeed, growing pseudopodia may represent a natural number *n* by the following parametric process:

$$\underline{n}(x, z) ::= \underbrace{\overline{x}.\overline{x}.\cdots.\overline{x}}_{n}.\overline{z}.Nil$$

The process $\underline{n}(x, z)$ proceeds *n* times on an output port called the successor channel $\overline{x} \in \{A_1, A_2, \ldots\} \cup \{R_1, R_2, \ldots\}$ (e.g. it is the same output of attractant) and once on the zero output port $\overline{z} \in \{A_1, A_2, \ldots\} \cup \{R_1, R_2, \ldots\}$ before becoming inactive *Nil*. Recall that it is a "Church-like" encoding of numerals used first in λ-calculus. Notice that in case of miracidiae or cercariae $\overline{x} \in \{A_1, A_2, \ldots\}$ and $\overline{z} \in \{A_1, A_2, \ldots\}$.

An addition process takes two natural numbers *i* and *j* represented using the channels *x*[*i*], *z*[*i*] and *x*[*j*], *z*[*j*] and returns their sum as a natural number represented using channels *x*[*i*+*j*], *z*[*i*+*j*]:

$$Add(x[i], z[i], x[j], z[j], x[i+j], z[i+j]) ::=$$
$$(x[i].\ \overline{x}[i+j].\ Add(x[i], z[i], x[j], z[j], x[i+j], z[i+j])) +$$
$$z[i].Copy(x[j], z[j], x[i+j], z[i+j])).$$

A multiplication process takes two natural numbers *i* and *j* represented using the channels *x*[*i*], *z*[*i*] and *x*[*j*], *z*[*j*] and returns their multiplication as a natural number represented using channels $x\left[\underbrace{i+\ldots+i}_{j}\right]$, $z\left[\underbrace{i+\ldots+i}_{j}\right]$:

$$Mult(x[i], z[i], x[j], z[j], x[i*j], z[i*j]) ::= Add(x[i], z[i],$$
$$x[j], z[j], x[i+\ldots+i], z[i+\ldots+i]).$$

The *Copy* process replicates the signal pattern on channels *x* and *y* on to channels *u* and *v*. It is defined as follows:

$$Copy(x, y, u, v) ::= (x.\ \overline{u}.\ Copy(x, y, u, v) + y.\ \overline{v}.\ Nil)$$

As we see, within Physarum spatial logic and its poorer version in the form of schistosomatidae behavioral logic we can consider some processes as arithmetic operations. Also, we can combine several arithmetic operations within one process. Let us regard the following expression:

$$(10 + 20) * (30 + 40)$$

An appropriate process is as follows:

Mult(*Add*(*x*[10], *z*[10], *x*[20], *z*[20], *x*[10+20], *z*[10+20]), *z*[10+20], *Add*(*x*[30], *z*[30], *x*[40], *z*[40], *x*[30+40], *z*[30+40]), *z*[30+40], *Add*(*x*[30], *z*[30], *x*[70], *z*[70], *x*[2100], *z*[2100])).

VI. Conclusion

We show that many biologized versions of π-calculus are possible: Physarum spatial logic, schistosomatidae behavioral logic, etc. One of its basic versions, Physarum spatial logic, can be used for constructing slime mould based computer. This logic is richer than schistosomatidae behavioral logic and may be involved for simulations of the latter. The fact that we can formalize biological behaviors as kind of logic confirms that biological processes can be considered as forms of concurrent and parallel computations.

References

[1] A. Adamatzky, B. De Lacy Costello, T. Asai, *Reaction–Diffusion Computers*. Amsterdam: Elsevier, 2005.

[2] A. Adamatzky, "Physarum machine: implementation of a Kolmogorov-Uspensky machine on a biological substrate," *Parallel Processing Letters*, vol. 17, No. 04, pp. 455–467, December 2007.

[3] A. Adamatzky, V. Erokhin, M. Grube, Th. Schubert, A. Schumann, "Physarum Chip Project: Growing Computers From Slime Mould," *International Journal of Unconventional Computing*, 8(4), pp. 319–323, 2012.

[4] W. W. Cort, "Schistosome dermatitis in the United States (Michigan)," *Journal of the American Medical Association*, vol. 90, pp. 1027–1029, 1928.

[5] W. Feiler, W. Haas, "*Trichobilharzia ocellata*: chemical stimuli of duck skin for cercarial attachment," *Parasitology*, vol. 96, pp. 507–517, 1988.

[6] W. Haas, "Host finding mechanisms – a physiological effect," in *Biology, Structure, Function: Encyclopedic Reference of Parasitology* (Y. Mehlhorn, ed.), edn 2, 1988, pp. 382–383.

[7] W. Haas, "Physiological analysis of cercarial behavior," *Journal of Parasitology*, vol. 78, pp. 243–255, 1992.

[8] W. Haas, S. Haeberlein, "Penetration of cercariae into the living human skin: *Schistosoma mansoni* vs. *Trichobilharzia szidati,*" *Parasitology Research*, vol. 105, No. 4, pp. 1061–1066, 2009.

[9] W. Haas, A. Roemer, "Invasion of the vertebrate skin by cercariae of *Trichobilharzia ocellata*: penetration processes and stimulating host signals," *Parasitology Research*, vol. 84, No. 10, pp. 787–795, 1998.

[10] P. Horák, L. Kolárová, "Bird schistosomes: do they die in mammalian skin?" *Trends in Parasitology*, vol. 17, No. 2, pp. 66–69, 2001.

[11] P. Horák, L. Kolárová, C.M. Adema, "Biology of the schistosome genus *Trichobilharzia,*" *Advances in Parasitology*, 52, pp. 155–233, 2002.

[12] L. Mikes, L. Zidková, M. Kasný, J. Dvorák, P. Horák, "In vitro stimulation of penetration gland emptying by *Trichobilharzia szidati* and *T. regenti* (Schistosomatidae) cercariae. Quantitative collection and partial characterization of the products," *Parasitology Research*, vol. 96, No. 4, pp. 230–241, 2005.

[13] D. H. Molyneux, "Control of human parasitic diseases: context and overview," *Advances in Parasitology*, vol. 61, pp. 1–45, 2006.

[14] W. Neuhaus, "Biologie und Entwicklund von *Trichobilharzia szidati* n. sp. (Trematoda, Schistosomatidae), einem Erreger von Dermatitis beim Menschen," *Zeitschrift für Parasitenkunde*, vol. 15, pp. 203–266, 1952.

[15] A. Schumann, A. Adamatzky, "Logical Modelling of Physarum Polycephalum," *Analele Universitatii de Vest, Timisoara, Seria Matematica – Informatica* XLVIII, 3, pp. 175–190, 2010.

[16] A. Schumann, A. Adamatzky, "Physarum Spatial Logic," *New Math. and Nat. Computation*, vol. 7, No. 3, pp. 483–498, 2011.

[17] M. Podhorsky, Z. Huzova, L. Mikes, P. Horak, "Cercarial dimensions and surface structures as a tool for species determination of *Trichobilharzi*a spp." *Acta Parasitologica*, vol. 50, pp. 343–365, 2009.

[18] С. А. Беэр, М. В. Воронин, *Церкариозы в урбанизированных экосистемах*. Москва: Наука, 2007.

[19] Ginetsinskaya T. A. ***Trematodes**, their Life Cycles, Biology and Evolution*. Leningrad: Nauka, 1968 (in Russian).

[20] A. Adamatzky, Physarum Machines: Computers from Slime Mould. World Scientific Publishing Company, 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 231–234

A Fuzzy Logic Approach to The Evaluation of Health Risks Associated with Obesity

Tadeusz Nawarycz
Department of Biophysics
Medical University of Lodz, Poland
Email: tadeusz.nawarycz@umed.lodz.pl

Krzysztof Pytel
Faculty of Physics and Applied Informatics
University of Lodz, Poland
Email: kpytel@uni.lodz.pl

Wojciech Drygas
Department of Social and Preventive Medicine
Medical University of Lodz, Poland

Maciej Gazicki-Lipman, Lidia Ostrowska-Nawarycz
Department of Biophysics
Medical University of Lodz, Poland

***Abstract*—Excessive body weight, especially in the form of the so-called abdominal obesity (AO) is an important factor of the cardio-metabolic risks (CMR). The paper presents a fuzzy model of AO and CMR assessments based on such key indicators of anthropometric measurements as body mass index (BMI as a measure of the global adiposity) as well as waist circumference (WC) and waist-to-height ratio (WHtR) as AO indicators. For the construction of a membership function (MF) the Zadeh's Extension Principle (EP) and mapping of the BMI fuzzy sets into adequate AO fuzzy sets using different transformation functions have been applied. Taking advantage of the results of a screening study, the AO membership functions for adult population of Lodz (WHO-CINDI project) are presented. MF design based on the EP theory is a useful methodology for assessing the AO and, consequently for a better assessment of CMR.**

I. Introduction

OBESITY is an important risk factor for cardiovascular diseases, and a primary public health problem in many countries of the world [1]. The trend in obesity is especially alarming among children and adolescents [2].

Pol-MONICA (Multinational Monitoring of Trends and Determinants in Cardiovascular Diseases) study has shown that overweight is exhibited by about 67% of the Polish population, while approximately 30% of women and 20% of men suffer from obesity [3].

As far as obesity is concerned, its most disadvantageous form is known as abdominal obesity (AO) also called visceral or central obesity [4]. An excess of visceral fat surrounding abdominal organs substantially increases a risk of life-threatening diseases such as type 2 diabetes and heart diseases [5]. These diseases are the most common cause of death and have the greatest impact on the financial burden of both individual and public health [6].

AO evaluation based on exact measurements of visceral fat is very difficult and it is only possible to be carried out with the help of complex and expensive imaging techniques [7]. In practice, an unfavorable distribution of adipose tissue and a presence of AO are usually estimated by indirect indices such as waist circumference (WC), waist to hip ratio (WHR) or recent the waist-to-height ratio (WHtR) [8], [9], [10].

Current recommendations establish sharp boundaries between terms like overweight, obesity, etc. They are often of a contractual nature of a consensus obtained in the form of recommendation. Issues related to the assessment of the AO belong to such unspecific problems of body composition which could be well approached by a theory of fuzzy sets [11], [12]. The paper presents a concept of fuzzy AO models developed on the basis of BMI, WC and WHtR measurements. Results of the respective studies in Lodz population were used to construct AO membership function (MF) with Zadeh Extension Principle [13], [14].

II. Categorization of health risk using body mass index (BMI) and waist circumference (WC)

Both in epidemiological studies and in clinical practice, overweight is most frequently determined on the basis of the BMI value. BMI is a simple index of "weight-for-height" calculated as weight in $[kg]$ divided by squared height in meters $[m^2]$. Clinically individuals with normal weight (NO) are those with BMI is lower than 25 $[kg/m^2]$. Overweight (OW) is defined as BMI of between 25 and 30 $[kg/m^2]$ and obesity (OB) is defined as BMI of 30 $[kg/m^2]$ [1].

Recently, the WC and WHtR indices have been successfully used as effective measures of AO both in adult and in children or adolescent populations. According to WHO and to National Cholesterol Education Program-Adult Treatment Panel (NCEP-ATP III), AO in the adult population refers to: WC higher than 102 cm for men (M) and higher that 88 cm for women (F) [8], [9]. The newest criteria developed by International Diabetes Federation (IDF) are much more demanding and they recommend an identification of AO obesity at lower levels of WC (94 cm for men and 80 cm for women) [10]. For WHtR, its value > 0.5 indicating AO (central fat distribution), regardless of gender and age [8], [11].

Classification based on a combination of BMI and WC is useful in identifying people at increased CMR. Table 1 presents health risk (CMR) classification according to BMI (three stages) and WC (two stages against NCEP-ATP-III) [8].

978-83-60810-53-8/$25.00

TABLE I
HEALTH RISK CLASSIFICATION FOR ADULT MEN (M) AND FEMALE (F) ACCORDING TO BODY MASS INDEX AND WAIST CIRCUMFERENCE [8].

Waist	**Body Mass Index (BMI in $[kg/m^2]$)**		
Circumference (WC in [cm])	Normal (NO) BMI $<$ 25	Overweight (OW) 25 $\leqslant$ BMI $<$ 30	Obese (OB) BMI $\geqslant$ 30
$<$ 102 cm (M) $<$ 88 cm (F)	Least Risk **(LRisk)**	Increased Risk **(IRisk)**	High Risk **(HRisk)**
$\geqslant$ 102 cm (M) $\geqslant$ 88 cm (F)	Increased Risk **(IRisk)**	High Risk **(HRisk)**	V. High Risk **(VHRisk)**

TABLE II
CARDIO-METABOLIC RISK (CMR) CLASSIFICATION ACCORDING TO BMI AND THREE AREA OF WC (CNO - WITHOUT CENTRAL OBESITY; COW - CENTRAL OVERWEIGHT AND COB - CENTRAL OBESITY).

Waist Circumference (WC)	**Body Mass Index (BMI)**		
	Normal BMI $<$ 25	**Overweight** 25 $\leqslant$ BMI $<$ 30	**Obese** BMI $\geqslant$ 30
cNO $<$ 94 cm (M) $<$ 80 cm (F)	**(LRisk)**	**(IRisk)**	**(HRisk)**
cOW (94 - 102) cm (M) (80 - 88) cm (F)	**(IRisk)**	**(HRisk)**	**(VHRisk)**
cOB $\geqslant$ 102 cm (M) $\geqslant$ 88 cm (F)	**(HRisk)**	**(VHRisk)**	**(VHRisk)**

In the light of the proposals by Lean et al. [15] as well as the previously mentioned recommendations by NCEP-ATP III and IDF, a more appropriate classification of CMR is based on BMI and two different WC cut-offs, which define three areas of central adiposity: cNO - without central obesity; cOW - central overweight; cOB - central obesity (Table 2). Similar classification of CMR can be made on the basis of BMI and are recommended by some authors of the two levels of WHtR cut-offs (WHtR $>$ 0.5 as a cOW and WHtR $>$ 0.6 as a cOB) [11]. It should be noted that, due to the dichotomous charkter of the analyzed risk factors as well as the problem of their cut-off points, CMR classifications shown in Tables 1 and 2 constitute rather a qualitative description of the cardio-metabolic complications.

III. EVALUATION OF THE HEALTH RISK ASSOCIATED WITH OBESITY USING A FUZZY LOGIC APPROACH

A. Extension Principle theory

The extension method proposed by Zadeh, usually called Extension Principle (EP) only, is one of the basic ideas to process the extansion of the classical mathematical concepts into fuzzy ones [13], [14]. Let us consider two crisp sets X and Y and f a mapping from X to Y, f: X $\rightarrow$ Y.

Let A be a fuzzy subset of X, $A \in X$ (Figure 1). So, the EP allows to built the image of A under the crisp mapping f as a fuzzy set B:

$$B = f(A) = (y, \mu_B(y) \mid y = f(x), x \in X \qquad (1)$$

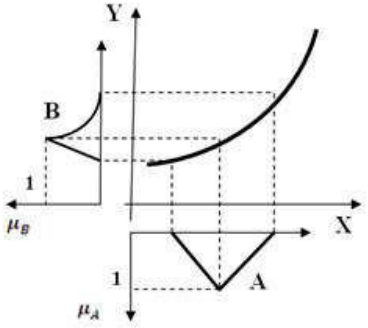


Fig. 1. Illustration of the Extension Principle for a mapping a single variable (explanations in text).

whose, membership function is given by:

$$\mu_B(y) = \begin{cases} sup\ \mu_A(x),\ if\ f^{-1}(y) \neq \emptyset \\ \\ 0,\ if\ f^{-1}(y) = \emptyset \end{cases} \quad for\ all\ y \in Y \qquad (2)$$

where: $f^{-1}(y)$ denotes the set of all points $x \in X$ such that $f(x) = y$.

In the case of X with an infinite number of elements a fuzzy set B induced by the mapping f can be expressed as:

$$B = f(A) = \int_y \frac{\mu_A(x)}{f(x)} \qquad (3)$$

Figure 1 illustrates the EP action considering the mapping $f : X \rightarrow Y$ and A a triangular fuzzy set.

B. Fuzzy models of global and abdominal obesity

The fuzzy model of global fatness was based on BMI that has a defined three fuzzy subsets (NO-normal, OW - overweight and OB - obesity) with the following MF:

$$NO = \begin{cases} 1 & for\ BMI \leq 25 \\ 6 - 0.2BMI & for\ 25 < BMI \leq 30 \\ 0 & for\ BMI > 30 \end{cases} \qquad (4)$$

$$OW = \begin{cases} 0 & for\ BMI \leq 25 \\ 0.2BMI - 5 & for\ 25 < BMI \leq 30 \\ 7 - 0.2BMI & for\ 30 < BMI \leq 35 \\ 0 & for\ BMI > 35 \end{cases} \qquad (5)$$

$$OB = \begin{cases} 0 & for\ BMI \leq 30 \\ 0.2BMI - 6 & for\ 30 < BMI \leq 35 \\ 1 & for\ BMI > 35 \end{cases} \qquad (6)$$

A graphic interpretation of fuzzy subsets of MF for NO, OW and OB is illustrated in Figure 2. Based on the defined BMI fuzzy model and EP theory described in section 3.1, the following transformation functions for men (M) and female (F) in construction of fuzzy AO subsets (cNO, cOW, cOB) have been used:

- f1: WC = f1(BMI); linear equations based on the NCAP-ATP III / IDF criteria [9].

$$f1 : \begin{cases} WC_M = 1.6 * BMI + 54 \\ WC_F = 1.6 * BMI + 40 \end{cases} \qquad (7)$$

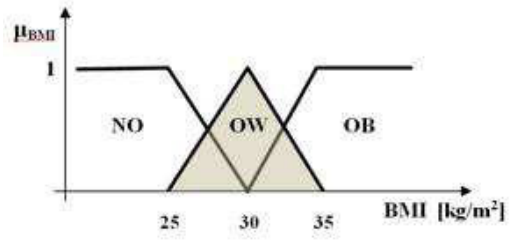


Fig. 2. Illustration of the Extension Principle for a mapping a single variable (explanations in text).

- f2: WC = f2(BMI); linear regression equations defined on the basis of epidemiological studies of the population in Lodz, designed by the WHO-CINDI (Countrywide Integrated Noncommunicable Diseases Intervention Programme: 1264 men and 1330 female aged 18 - 80 y.) [16]:

$$f2: \begin{cases} WC_M = 2.66 * BMI + 22.9 \\ WC_F = 2.27 * BMI + 23.1 \end{cases} \quad (8)$$

- f3: WHtR = f3(BMI); linear equation prepared on the basis on Ashwell recommendations [11]:

$$f3: WHtR_{M/F} = 0.02 * BMI \quad (9)$$

- f4: WHtR = f4 (BMI); regression equations defined on the basis of the WHO-CINDI population of Lodz [15]:

$$f4: \begin{cases} WHtR_M = -16 * 10^{-5} * BMI^2 + 25 * 10^{-3} * \\ \qquad * BMI - 0.012 \\ WHtR_F = -12 * 10^{-5} * BMI^2 + 22.7 * 10^{-3} * \\ \qquad * BMI + 0.0125 \end{cases} \quad (10)$$

Example: Mapping of the cOW fuzzy subset based on the proposed model and the fuzzy BMI predefined transformation functions f1-f4 ($OW_{BMI} \rightarrow cOW$).

Let:

$$OW_{BMI} = \{\frac{0}{25} + \frac{0.2}{26} + \frac{0.6}{28} + \frac{1}{30} + \frac{0.6}{32} + \frac{0.2}{34} + \frac{0}{35}\} \quad (11)$$

will be selected fuzzy numbers belonging to OW fuzzy set defined by the formula (5). Using the transformation functions f1-f4 (7 - 10) defined for men (M) let us calculate appropriate numbers belonging to a fuzzy subset cOW

- for f1:

$$\begin{aligned} cOW_{f1} &= \{\frac{0}{f(25)} + \frac{0.2}{f(26)} + \frac{0.6}{f(28)} + \frac{1}{f(30)} + \\ &+ \frac{0.6}{f(32)} + \frac{0.2}{f(34)} + \frac{0}{f(35)}\} \end{aligned} \quad (12)$$

and after substituting the values we get:

$$\begin{aligned} cOW_{f1} &= \{\frac{0}{94} + \frac{0.2}{95.6} + \frac{0.6}{98.8} + \frac{1}{102} + \\ &+ \frac{0.6}{105.2} + \frac{0.2}{108.4} + \frac{0}{110}\} \end{aligned} \quad (13)$$

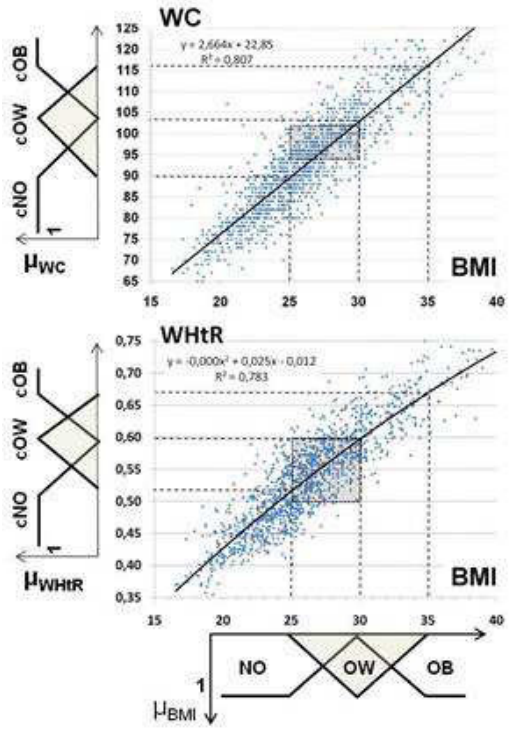


Fig. 3. Example of mapping the WC (top) and WHtR (bottom) fuzzy sets for a men population from Lodz [15]. The regression equations were used as a transformation functions. The gray rectangles - the cOW areas based on the NCEP-ATPIII/IDF recommended classification [8],[9].

- for f2:

$$\begin{aligned} cOW_{f2} &= \{\frac{0}{89.5} + \frac{0.2}{92.1} + \frac{0.6}{97.4} + \frac{1}{102} + \\ &+ \frac{0.6}{105.2} + \frac{0.2}{108.4} + \frac{0}{110}\} \end{aligned} \quad (14)$$

- for f3:

$$\begin{aligned} cOW_{f3} &= \{\frac{0}{0.5} + \frac{0.2}{0.52} + \frac{0.6}{0.56} + \frac{1}{0.6} + \\ &+ \frac{0.6}{0.64} + \frac{0.2}{0.68} + \frac{0}{0.7}\} \end{aligned} \quad (15)$$

- for f4:

$$\begin{aligned} cOW_{f4} &= \{\frac{0}{0.51} + \frac{0.2}{0.53} + \frac{0.6}{0.56} + \frac{1}{0.59} + \\ &+ \frac{0.2}{0.62} + \frac{0.2}{0.65} + \frac{0}{0.67}\} \end{aligned} \quad (16)$$

Example of the construction fuzzy sets AO (cNO, cOW, cOB) based on regression relationships WCM = f2 (BMI) (formula 8) and $WHtR_M$ = f4 (BMI) (formula 10) for the population of Lodz are presented in Figure 3 and Figure 4. WC and WHtR cut - off points designated for men (M) and females (F) based on mapping functions f1 - f4 are summarized in Table 3. As we can notice, the cut-off points for the fuzzy

TABLE III
WAIST CIRCUMFERENCE (WC) AND WAIST CIRCUMFERENCE-TO HEIGHT RATIO (WHtR) CUTOFFS DESIGNATED FOR MEN (M) AND FEMALES (F) BASED ON DIFFERENT MAPPING FUNCTIONS ($f1 - f4$)

BMI	WC cut-off [cm] (mapping functions)				WHtR cut-off [-] (mapping functions)			
	(f1)		(f2)		(f3)		(f4)	
	M	F	M	F	M	F	M	F
18.5	83,6	69,6	72,1	65,1			0,396	0,391
25	94	80	89,4	79,8	0,5	0,5	0,513	0,504
30	102	88	102,7	91,2	0,6	0,6	0,594	0,585
35	110	96	116	102,6			0,667	0,659

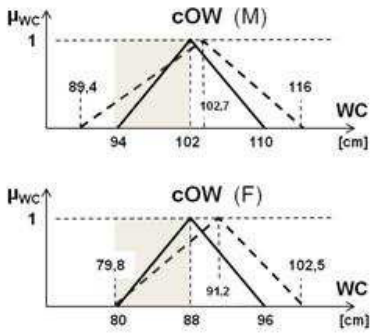


Fig. 4. Graphical form of triangular membership function for the fuzzy set "central overweight" (cOW) for men (M) and females (F). Solid line - developed based on NCEP-ATP III and IDF criteria (function f1); dotted line - developed for the population of Lodz (function f2).

subsets AO are different between themselves and depend on the transforming function. For example, with respect to the triangular MF describing the fuzzy subset cOW of the Lodz inhabitants, both for men and women (Figure 4) we observed the relation:

$$Supp(cOW)_{f1} < Supp(cOW)_{f2} \tag{17}$$

In case of the fuzzy subset cOW determined by WHtR, this relation is not so evident (Table 3). Both for men and women from Lodz, the WHtR cut-off points determined with the use of correlation functions f4 were more close to the current experts recommendations described by function f3.

IV. FINAL REMARKS

Adipose tissue, especially centrally distributed in the body, meets the complex and difficult to unambiguously describe regulatory functions. The excess of visceral fat is the cause of many cardio-metabolic disturbances mainly in the form of insulin resistance and secondary hyperinsulinemia [11], [16]. The current AO classifications are based mainly on the arbitrarily defined WC or WHtR cut-off points, which often are not fully adequate for the characteristic distribution of these indicators for a given population (region).

The paper presents a fuzzy models of AO developed on the basis on simple anthropometric indices such WC and WHtR. For the construction of a MF, the Zadeh's Extension Principle and mapping of the BMI fuzzy sets into adequate AO fuzzy sets using different transformation functions have been applied. The results obtained for the population of Lodz indicate the potential usefulness of the fuzzy set theory in the evaluation of both, the AO and the CMR. For the construction of the MF fuzzy subsets describing the AO it is reasonable to apply the Zadeh's Extension Principle and as a mapping function, the population-specific correlation function.

The results of our study may be a starting point for the construction of more complex fuzzy inference system for the assessment of global cardiovascular or cardiometabolic risk where AO is one of the more important risk factor. Taking into account the growing scale of the adverse health consequences of the obesity epidemic in the world, the artificial intelligence methods including particular fuzzy inference systems should be more widely used.

REFERENCES

[1] WHO, 2000, *Obesity: preventing and managing the global epidemic*. WHO Technical Report Series No.894. World Health Organization, Geneva.
[2] Branca, F., Nikogosian, H., Lobstein, T., 2007, *The challenge of obesity in the WHO European region and the strategies for response: Summary*. WHO & WHO European Ministerial Conference on Counteracting Obesity. Copenhagen: World Health Organization, Regional Office for Europe.
[3] Rywik S., Broda G., Piotrowski W. et al., 1996, *Epidemiology of Cardiovascular Disease*. Pol-MONICA Program., Kard. Pol., 1996 (supl.2), pp. 7–35 (in Polish).
[4] Larsson B et al., 1984, *Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow-up of participants in the study of men born in 1913*. Br Med J., 288, pp. 1401–4.
[5] Wang Y et al., 2005, *Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men*. Am J Clin Nutr., 81: 555–63.
[6] Hammond R.A., Levine R., 2010, *The economic impact of obesity in the United States*, Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 3, pp. 285–295.
[7] Despres J.P, Ross R, Lemieux S., 1996, *Imaging techniques applied to the measurement of human body composition*. In: Roche AF, Heymsfield SB, Lohman TG, eds. Human body composition. Chicago, IL: Human Kinetics, pp. 149-166.
[8] World Health Organisation. *Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation 2008*. World Health Organisation: Geneva, 2011.
[9] Grundy S.M., Brewer H.B., Cleeman J.I., et al. 2004, *Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition*., Arteriosclerosis, Thromb. and Vasc. Biol., 2 (24), pp. e13–8.
[10] IDF consensus worldwide definition of the metabolic syndrome. (www.idf.org - accessed January 2013).
[11] Ashwell M, Gunn P, Gibson S. *Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis*. ObesRev 2012; 13: 275–286.
[12] Rutkowska, D., Starczewski, A. *Fuzzy Inference Neural Network and Their Applications to Medical Diagnosis*. In: Szczepaniak, P., Lisboa, P., Kacprzyk, J. (eds.) Fuzzy System in Medicine. Physica - Verlag, Heidelberg (2000).
[13] Massad E, Ortega N.R, Barros L.C., Struchiner C.J., *Fuzzy Logic in Action: Applications in Epidemiology and Beyond* (Studies in Fuzziness and Soft Computing, v. 232), Springer, 2009.
[14] Zadeh L.A., 1975, *The concept of a linguistic variable and its application to approximate reasoning I*. Information Science, 8, pp. 199–251.
[15] Lean M.E, Han T.S, Morrison C.E., 1995, *Waist circumference as a measure for indicating need for weight management*. BMJ, 311, pp. 158–61.
[16] Leparski E., Nussel E. *Protocol and guidelines for monitoring and evaluation procedure CINDI - Countrywide Integrated Noncommunicable Diseases Intervention Programme*, 1987, 60, Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokio.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 235–240

Failure Analysis and Estimation of the Healthcare System

Elena Zaitseva, Jozef Kostolny, Miroslav Kvassay,
Vitaly Levashenko
University of Zilina
Department of Informatics
Zilina, Slovakia
Email: {jozef.kostolny, morislav.kvassay,
vitally.levashenko, elena.zaitseva}@fri.uniza.sk

Krzysztof Pancerz
University of Management and Administration
Zamość, Poland
Institute of Biomedical Informatics
University of Information Technology and
Management
Rzeszów, Poland
Email: kpancerz@wszia.edu.pl

***Abstract*—The principal goal of information technologies application in medicine is improvement and conditioning of medical care. Modern healthcare systems have to perfect the care of a patient. Therefore, the healthcare system has to be characterized, first of all, by high reliability and reliability analysis of such a system is an important problem. The new method for estimation of system reliability is considered in this paper. This method permits to investigate the influence of any system component failure to the system functioning.**

I. Introduction

INITIATIVES for implementing healthcare systems based on the information technologies are now a principal part of the development in medicine. The development of these systems depends on organization of the healthcare provision in each country and the presence of the information and telecommunication technologies in the healthcare sector [1–3]. There is one principal characteristic for all healthcare systems. It is reliability that is defined as the probability that a system will perform its intended function during a period of running time without any failure. A fault is an erroneous state of the system. The system reliability is a complex characteristic that depends on the functioning of separate parts (components) of the system.

Based on bibliography in reliability analysis of the healthcare domain, we can show two principal approaches. The first of them is reliability estimation of medical equipment and devices that includes reliability quantification of hardware and software of the healthcare system [4–6]. The second approach agrees with examination of human errors [7, 8]. However, independent evaluation of these principal parts of the healthcare system does not allow providing detail and actual reliability analysis. In [9], new tendencies in reliability engineering are considered. According to [9], the reliability analysis has to be based on joint evaluation of all principal parts (components).

The typical healthcare system structure consists of some principal components from the point of view of reliability analysis [4, 7, 9]. In [4], two of them have been defined: equipment/device and human factors. We need to note that the human factor has been considered as errors of operators of medical equipment or devices in [4]. A detailed structure of the human factor and human errors for the healthcare system is presented in [7]. The healthcare system structure includes three components: technical, human and organization [9]. The technical component includes two types of medical devices/equipment that are based on special and standards-based technologies according to [10]. For example, the first type is the medical decision support system, the system for integrating electronic medical records or picture archiving communication systems. The second type is the special medical device and equipment that can be used for a special operation only (as magnetic resonance imaging scanners, for example). The human component of the healthcare system causes medical errors. The organization component of the system joins management aspects and maintenance of the healthcare system.

In this paper, we develop results that have been presented in [9] as well as methods proposed for estimation of system components based on a single approach. Particularly, we consider the Importance Analysis of the healthcare system. This analysis allows investigation of every system component functioning/failure into the system reliability. In Section II, the typical approach for the reliability examination from the step of mathematical modelling to the calculation of the reliability indices is considered basing on an example of the mathematical model of performance shaping factors for human errors in the healthcare system. The Direct Partial Logic Derivatives are also proposed in this section for description of the system behaviour. Section III presents most frequently used importance measures that allow defining the system component with minimal or maximal influence to the system reliability. The algorithms for calculation of these measures based on Direct Partial Logic Derivatives are developed in this section. The new algorithm for calculation of one of the possible importance measures is proposed in Section IV. The possible development of the proposed methodology is analysed in Conclusions.

II. Mathematical Background

The reliability analysis of a system includes three principal steps [11]:

- the quantification of the system model;
- the representation and modelling of the system;

This work is supported by the grant of Slovak Research and Development Agency SK-PL-0023-12

- the representation, propagation and quantification of the uncertainty in the system behaviour.

I. Quantification of the System

Quantification of the system is a principal step and it causes development of a mathematical model. There are two approaches to the quantification in reliability engineering.

The first of them defines only two states of the system reliability: the functioning and failure. The mathematical model for the representation of this quantification is called a Binary-State System (BSS). The system and its components are allowed to have only two possible states (completely failure and functioning) in BSS (Fig. 1). This approach is well known and widely used in reliability engineering. The system failure can be investigated in detail based on this quantification. However, the analysis of other performance levels, before the system failure, has some difficulties for BSS. In this case, the quantification of the system reliability to some performance levels is used. The mathematical model with some performance levels is called a Multi-State System (MSS).

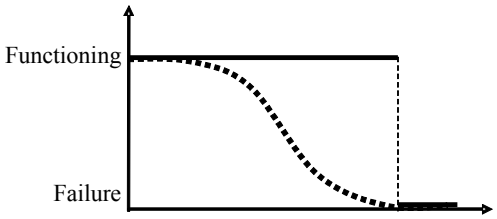


Fig. 1 System reliability interpretations for BSS

MSS reliability analysis is a more flexible approach to evaluating system reliability, as it can be used when both the system and its components may experience more than two states, including, for example, completely failed, partially failed, partially functioning and perfect functioning (Fig. 2). The MSS scientific achievement has been documented in [11 – 13]. However, a mathematical approach to analysing such a system is complex. In many applications, the definition of the system failure is the principal problem. Therefore, the system quantification can be simplified and considered as the BSS (Fig. 1).

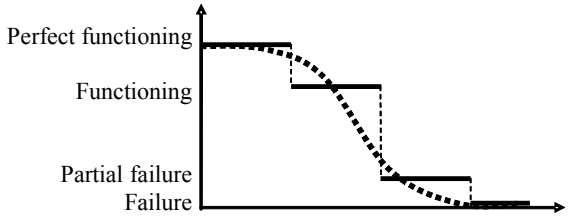


Fig. 2 System reliability interpretations for MSS

The BSS permits to investigate the system failure that has the high priority in reliability engineering. Therefore, mathematical models and methods for the estimation of the system failure are developed dominantly based on the system representation by BSS, because MSS complicates the system failure investigation.

In this paper, the analysis of the healthcare system failure is considered on the basis of the use of BSS.

II. Modelling of the System

The next step, after the definition of quantification, is the mathematical model development. There are some types of the BSS representation as the mathematical model. These representations (mathematical models) correlate with the mathematical methods for the calculation of the system reliability indices and measures. One of these representations is the structure function. This function allows the mathematical description of a system with any complexity [12, 13].

The system reliability in the stationary state depending on component states is defined by *structure function* [14]:

$$\phi(x_1,...,x_n) = \phi(x) : \{0,1\}^n \rightarrow \{0,1\}. \qquad (1)$$

A coherent system is considered in the paper below. The important assumptions for this system [12, 14] are as follows: (a) the structure function (1) is monotone, and (b) the system component failure does not improve the system reliability.

In the considered mathematical model, every system component x_i is characterized by probability of the reliability:

$$p_i = \Pr\{x_i = 1\}. \qquad (2)$$

The system component unreliability is defined as:

$$q_i = \Pr\{x_i = 0\} = 1 - p_i. \qquad (3)$$

For example, the structure function of the human sub-system (component) can be defined on the basis of the mathematical model of performance shaping factors for human errors in the healthcare system that is proposed in [10]. According to the model in [10], the analysis of the human error has to include social, personal, organization and technological aspects (Fig. 3). We can interpret this model as the structure function:

$$\varphi(x) = x_1 \wedge \left((x_2 \wedge x_3) \vee (x_2 \wedge x_4) \vee (x_3 \wedge x_4) \right), \qquad (4)$$

where $\wedge$ and $\vee$ are symbols of the operations AND and OR accordingly.

The structure function (4) defines correlation of the social, technical and organizational aspects as the system 2-out-of-3, that is to say, the combination of these aspects is dependable if two or more of these aspects are reliable. These aspects and the personal aspect are correlated as the series system. We use the term "component" for any of these aspects and indicate as x_i in this paper in the examples below.

The BSS behavior specified by the structure function is described by the Direct Partial Logic Derivative. In this case, the structure function variables are interpreted as the component state, and the function value is agreed with the system state (reliability).

The Direct Partial Logic Derivative with respect to variable x_i for the BSS structure function (1) permits to

The mathematical model of performance shaping factors for human errors

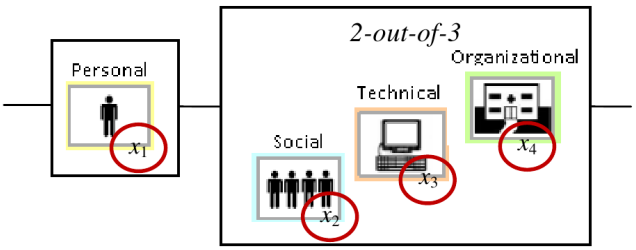


Fig. 3 The structure function for the healthcare system human component

analyse the system reliability change from j to $\bar{j}$ when the i-th component state changes from a to $\bar{a}$ [14]:

$$\partial\varphi(j\to\bar{j})/\partial x_i(a\to\bar{a})= =\begin{cases}1, & if\ \varphi(a_i,x)=j\wedge\varphi(\bar{a}_i,x)=\bar{j}\\ 0, & otherwise\end{cases} \quad (5)$$

where $\phi(a_i, \boldsymbol{x}) = \phi(x_1,\ldots, x_{i-1}, a, x_{i+1},\ldots, x_n)$; $\phi(\bar{a}, \boldsymbol{x}) = \phi(x_1, \ldots, x_{i-1}, \bar{a}, x_{i+1},\ldots, x_n)$; $a, j \in \{0, 1\}$ and $\bar{a} = 1-a$, $\bar{j} = 1-j$.

Let us consider the system failure in the Direct Partial Logic Derivative terminology. The system failure is represented as a change of the structure function value $\phi(\boldsymbol{x})$ from state 1 into 0. This change can be caused by the i-th variable change from 1 to 0 if we consider a coherent system. Therefore the Direct Partial Logic Derivative for the BSS failure analysis is defined by the equation

$$\partial\varphi(1\to0)/\partial x_i(1\to0)= =\begin{cases}1, & if\ \varphi(1_i,x)=1\ and\ \varphi(0_i,x)=0,\\ 0, & otherwise.\end{cases} \quad (6)$$

The Direct Partial Logic Derivative (6) $\partial\varphi(1\to0)/\partial x_i(1\to0)$ allows investigating boundary states of this system for which failure of one component x_i causes the system breakdown. For example, these states for the system in Fig. 3 are shown in Table 1. Therefore, there are 4 boundary states for the first component and 2 states for every other component of the mathematical model of performance shaping factors for human errors in the healthcare system.

The Direct Partial Logic Derivative can be used for the investigation of the influence of the system components failure to the failure of BSS. This investigation is a subject of the Importance Analysis in the reliability engineering [12–15].

III. Mathematical Method

Importance analysis allows examining different aspects of reliability changes and the uncertainty in the system failure. In particular, the importance analysis is used for BSS reliability estimation depending on the system structure and its component states. The various evaluations of the BSS component importance are called Importance Measures (IMs). IM quantifies the criticality of a particular component within BSS. They have been widely used as tools for identifying system weaknesses, and to prioritise reliability improvement activities.

The most frequently used IMs as Structural Importance (SI), Birnbaum importance (BI), Fussell-Vesely Importance (FVI) are shown in Table II [12, 13].

TABLE I.
BOUNDARY STATES FOR THE HEALTHCARE SYSTEM HUMAN COMPONENT (FIG.3)

$\partial\phi(1\to0)/\partial x_1(1\to0)$	$\partial\phi(1\to0)/\partial x_2(1\to0)$	$\partial\phi(1\to0)/\partial x_3(1\to0)$	$\partial\phi(1\to0)/\partial x_4(1\to0)$
(1→0, 0, 1, 1)	(1, 1→0, 0, 1)	(1, 0, 1→0, 1)	(1, 0, 1, 1→0)
(1→0, 1, 0, 1)	(1, 1→0, 1, 0)	(1, 1, 1→0, 0)	(1, 1, 0, 1→0)
(1→0, 1, 1, 0)			
(1→0, 1, 1, 1			

TABLE II.
IMPORTANCE MEASURES

Short name	Description
SI	SI concentrates on the topological structure of the system and determines the proportion of working states of the system in which the working of the i-th component makes the difference between system failure and working state.
BI	BI of a given component is defined as the probability that such a component is critical to MSS functioning and represents loss in MSS when the i-th component fails.
FVI	FVI quantifies the maximum decrement in MSS reliability caused by the i-th system component state deterioration and if $a = 0$, the measure allows estimating system performance level decrease for full unreliability of the i-th system component.

Calculation of IMs is based on different mathematical approaches and the Direct Partial Logic Derivative is one of them. This approach has been proposed in [14]. According to [14], the Direct Partial Logic Derivative has been used for calculation of SI and BI. The FVI definition is based on the minimal cuts of BSS. In this paper, a new algorithm for calculation of the minimal cuts of the system based on the Direct Partial Logic Derivative is proposed.

III. IMPORTANCE MEASURES

IV. Structural Importance

SI is one of the simplest measures of the component importance and this measure focuses on the topological aspects of the system. According to the definition in [16], this measure determines the proportion of working states of the system in which the working of the i-th component makes the difference between system failure and its working:

$$IS_i=\frac{\rho_i}{2^{n-1}} \quad (7)$$

where ρ_i is a number of system states when the change component state results in the system failure.

For example, calculated IMs (7) based on Direct Partial Boolean Derivatives for the system are shown in Fig. 3. Values of SI (8) and intermediate values of ρ_i are shown in Table III. According to this table, the first component has maximal influence to the system reliability from the point of view of the system topology.

TABLE III.
STRUCTURAL IMPORTANCE FOR THE SYSTEM IN FIG.3

i	ρ_i	IS_i
1	4	4/8 = 0.500
2	2	2/8 = 0.250
3	2	2/8 = 0.250
4	2	2/8 = 0.250

V. Birnbaum Importance

BI of a given component is defined as the probability that the system is sensitive to inoperative of the *i*-th system component [17]. Let us consider the Direct Partial Logical Derivatives for calculation of BI. In [14], BI has been defined as

$$IB_i = \Pr\left\{\partial\varphi(1\to 0)/\partial x_i(1\to 0)=1\right\} \qquad (8)$$

For example, let us consider the system shown in Fig. 3. Probabilities of the system element reliability and unreliability are shown in Table IV. According to the data in Table I, elements of this system have the following values of BIs:

$$IB_1=\Pr\{\partial\phi(1\to 0)/\partial x_1(1\to 0)=1\}=$$
$$=q_2q_3p_4+ q_2p_3q_4+ q_2p_3p_4+ p_2q_3q_4+ p_2q_3p_4+ p_2p_3q_4+ p_2p_3p_4=0.964;$$

$$IB_2=\Pr\{\partial\phi(1\to 0)/\partial x_2(1\to 0)=1\}=p_1q_3q_4=0.096;$$

$$IB_3=\Pr\{\partial\phi(1\to 0)/\partial x_3(1\to 0)=1\}=p_1q_2q_4=0.072.$$

$$IB_4=\Pr\{\partial\phi(1\to 0)/\partial x_4(1\to 0)=1\}=p_1q_2q_3=0.096.$$

BI for the first component has the maximal value. The BIs, as the SIs, show that the first system component is more important for reliability.

TABLE IV.
PROBABILITIES OF ELEMENTS FOR THE SYSTEM IN FIG.3

	x_1	x_2	x_3	x_4
p_i	0.80	0.70	0.60	0.70
q_i	0.20	0.30	0.40	0.30

VI. Fussell-Vesely Importance

FVI represents the contribution of each component to the system failure probability and for BSS it is calculated by the following equation [17]:

$$I_{FV}(x_i)=\frac{F_{\min\ cut}(x_i)}{Q} \qquad (9)$$

where $F_{\min\ cut}(x_i)$ is the system minimal cut that includes the *i*-th system component, Q is the function of the system unreliability [14, 17]:

$$Q = \Pr\{\phi(\boldsymbol{x})=0\}. \qquad (10)$$

Therefore, for calculation of this measure, the minimal cut set is needed. In the next section, we propose a new algorithm for calculation of the minimal cut set for BSS by Direct Partial Logic Derivatives.

IV. MINIMAL CUT SET IN IMPORTANCE ANALYSIS

VII. Minimal Cut Set and Minimal Cut Vector

Let us consider the conception of the cut set. The cut set is the set of the system components whose simultaneous failure results in the system failure (if the system has been functioning). As a rule, the number of the cut set components k is changed from 1 to n. The system failure is caused by one component reduction only if $k = 1$ and all components have to fail that to cause the system failure if $k = n$. The minimal cut set is a cut set in which any subset remaining after the removal of any of its components is no longer the cut set.

Let $\boldsymbol{a} = (a_1 \ldots a_n)$ and $\boldsymbol{b} = (b_1 \ldots b_n)$ be two state vectors for system component states or values of structure function (1). The vector $\boldsymbol{a} < \boldsymbol{b}$ if $a_i < b_i$ for $i = 1, \ldots, n$.

The state vector $\boldsymbol{a} = (a_1 \ldots a_n)$ is a cut set vector if $\phi(\boldsymbol{a}) = 0$.

The cut set vector $\boldsymbol{a}$ is minimal, if $\phi(\boldsymbol{b}) = 1$ for any $\boldsymbol{b} > \boldsymbol{a}$.

For example, the system shown in Fig. 3 has 12 cut set vectors (if $\phi(\boldsymbol{x}) = 0$, $\phi(\boldsymbol{x})$ is defined by (4)) and 4 minimal cut set vectors:

$$\{(x_1), (x_2\ x_3), (x_2\ x_4), (x_3\ x_4)\}. \qquad (11)$$

Therefore, FVI for this system, according to (9), is calculated as:

$$IFV_1=\Pr\{(x_1)\}\ /\ Q = q_1\ /\ Q = 0.492$$

$$IFV_2=\Pr\{(x_2\ x_3), (x_2\ x_4)\}\ /\ Q = q_2q_3 + q_2q_4\ /\ Q = 0.517$$

$$IFV_3=\Pr\{(x_2\ x_3), (x_3\ x_4)\}\ /\ Q = q_2q_3 + q_3q_4\ /\ Q = 0.591$$

$$IFV_4=\Pr\{(x_2\ x_4), (x_3\ x_4)\}\ /\ Q = q_2q_4 + q_3q_4\ /\ Q = 0.517$$

where the system unreliability Q (10) is calculated as:

$$Q = \Pr\{\phi(\boldsymbol{x})=0\} = q_1 + p_1(q_2q_3 + q_2q_4 + q_3q_4) = 0.4064.$$

FVI of the first component has the minimal value. Therefore, the first component does not have most significant influence to the system reliability if the component combination failure is considered. According to value IFV_3, the third component refuse causes the system failure in combination with other component refuses predominantly.

FVI is an alternative measure of Importance Analysis that allows estimating influence of the particular component to the system reliability and functioning. However, the minimal cut sets for the calculation of this measure are needed.

VIII. Minimal Cut Set Vectors and Direct Partial Logic Derivatives

Let us compare two definitions of the Direct Partial Logic Derivatives and minimal cut set vectors. Let the Direct Partial Logic Derivative be $\partial\phi(1\to0)/\partial x_i(1\to0)$. This derivative permits to determine the structure function state vectors that are boundary for the structure function value with respect to the variable x_i. The minimal cut set vector is the boundary state vector too, but for some variables (components). Therefore, the set of Direct Partial Logic Derivatives with respect to some variables can be defined by the minimal cut set vector. This supposition has been verified and tested. The result of testing confirms the supposition.

The Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_i(1\to0)$ indicates state vectors $(0_i, \boldsymbol{x})$, in which improvement of component i results in the system improvement. To identify minimal state vectors, i.e., state vectors for which improvement of any broken component results in the improvement of the whole system, we have to compute $\partial\phi(1\to0)/\partial x_i(1\to0)$ (5) for every component and then compute the intersection of these derivatives. To compute the intersection, the modified type of derivative has to be used that is defined as:

$$\partial\varphi(j\to\bar{j})/\partial x_i(s\to\bar{s})=$$
$$=\begin{cases}1 & \text{if } x_i=s \text{ and } \varphi(s_i,x)=j \text{ and } \varphi(\bar{s}_i,x)=\bar{j}\\ 0 & \text{if } x_i=s \text{ and } \varphi(s_i,x)=\varphi(\bar{s}_i,x)\\ * & \text{if } x_i\neq s\end{cases} \tag{12}$$

The rule for intersection of two modified derivatives (11) is defined in Table V. This intersection identifies state vectors, in which improvement of both components (if the component can be repaired) results in improvement of the system.

Let us continue the hand calculation example for the mathematical model of performance shaping factors for human errors in the healthcare system (Fig. 3) that is defined by the structure function (4). The Direct Partial Logic Derivatives $\partial\phi(1\to0)/\partial x_i(1\to0)$ for system components have been calculated and shown in Table I. The intersection of these derivatives, according to the rule in Table V, allows getting 4 cut set vectors:

$$\{0***,\ *00*,\ *0*0,\ **00\},$$

which are consistent with the minimal cut sets (12).

TABLE V.
DEFINING THE INTERSECTION OF TWO MODIFIED DPLD

		$\partial\varphi(j\to\bar{j})/\partial x_i(s\to\bar{s})$		
		*	0	1
$\partial\varphi(j\to\bar{j})/\partial x_i(s\to\bar{s})$	*	*	0	1
	0	0	0	0
	1	1	0	1

The test of the proposed algorithm has been implemented on the basis of the sets of the benchmarks LGSynth91 [18]. Testing characteristic is a number of cut set vectors and time for computation (Fig. 4). The numbers in the left part of the graphs indicate the time and the numbers in the right part are numbers of the cut set vectors for the system. There is the proportional correlation between the number of cut set vectors in the system and time for the computation.

V. CONCLUSION

In this paper, the problem of calculation of IMs is considered. The IM definitions based on the Direct Partial Logic Derivatives [4] are provided. A new algorithm for calculation of FVI by the Direct Partial Logic Derivatives is presented in this paper. The experimental investigation corroborates the possible application of this algorithm for investigation of the large dimension system and computation of the IMs. The development of the presented result in future investigation will be adaptation of this algorithm for analysis of Multi-State System. This system allows the analysis of some (more than two) states in the reliability [12].

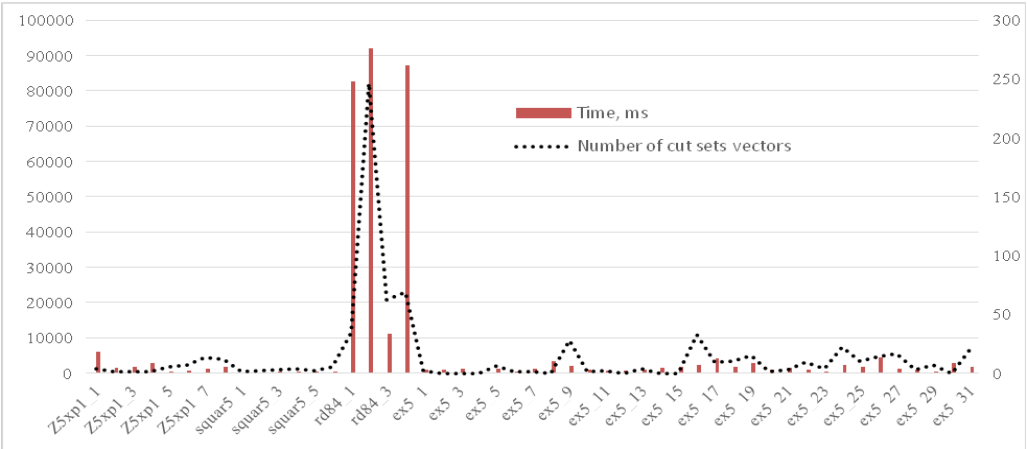


Fig. 4 Computational complexities

Reliability analysis of the health care system is an important issue. The principal problem in this analysis is the development of methodology that permits to investigate every system component and the system based on united approaches. This conception has been presented in details in [9, 15]. We propose and develop one of possible approaches that allow investigating the healthcare system component importance. The mathematical background of this approach is Direct Partial Logic Derivatives. The advantage of this approach is the possibility to use it for estimation of every system that is defined by the structure function.

REFERENCES

[1] M. S. Bogner, *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994, 428 p.

[2] D. Castro, "Meeting national and international goals for improving health care: The role of information technology in medical research," in *Proc. Atlanta Conf. on Science and Innovation Policy, 2009*, Atlanta, 2009, pp. 1–9.

[3] T. Cohen, "Medical and information technologies converge," *IEEE Engineering in Medicine and Biology Magazine*, vol. 23, pp. 59–65, May-June 2004.

[4] B. S. Dhillon, *Medical Device Reliability and Associated Areas*. Boca Raton, FL: CRC Press, 2000, 264 p.

[5] A. Taleb-Bendiab, D. England, M. Randles, P. Miseldine and K. Murphy, "A principled approach to the design of healthcare systems: Autonomy vs. governance," *Reliability Engineering and System Safety*, vol. 91, pp. 1576–1585, December 2006.

[6] S. Spyrou, P. D. Bamidis, N. Maglaveras, G. Pangalos and C. Pappas, "A methodology for reliability analysis in health networks," *IEEE Trans. on Information Technology in Biomedicine*, vol. 12, pp. 377–386, May 2008.

[7] M. Vaughn-Cooke, H. B. Nembhard and J. Ulbrecht, "Reformulating human reliability in healthcare system," in *Dig. Proc. IIE Annual Conference and Expo 2010*, Cancun, 2010.

[8] M. Lyons, S. Adams, M. Woloshynowych and C. Vincent, "Human reliability analysis in healthcare: A review of techniques," *Int. Journal of Risk and Safety in Medicine*, vol. 16, pp. 223–237, 2004.

[9] E. Zaitseva, V. Levashenko and M. Rusin, "Reliability analysis of healthcare system," in *Proc. Federated Conf. on Computer Science and Information Systems (FedCSIS)*, Szczecin, 2011, pp. 169–175.

[10] K. Siau, "Health care informatics," *IEEE Trans. on Information Technology in Biomedicine*, vol. 7, pp. 1–7, March 2003.

[11] E. Zio, "Reliability engineering: Old problems and new challenges," *Reliability Engineering and System Safety*, vol. 94, pp. 125–141, February 2009.

[12] A. Lisnianski and G. Levitin, *Multi-state System Reliability: Assessment, Optimization and Applications*. Singapore: World Scientific, 2003, 358 p.

[13] E. Zaitseva, "Importance analysis of a multi-state system based on multiple-valued logic methods", in: *Recent advances in system reliability: Signatures, Multi-state systems and Statistical inference*, A. Lisnianski, I. Frenkel (Eds), Springer, London, 2012, pp. 113-134.

[14] E. N. Zaitseva and V. G. Levashenko, "Importance analysis by logical differential calculus," *Automation and Remote Control*, vol. 74, no. 2, pp. 171–182, Feb. 2013.

[15] E. Zaitseva, "Importance Measures in Reliability Analysis of Healthcare System", in: *Human–Computer Systems Interaction: Backgrounds and Applications 2, part 1*, Z.S. Hippe, J.L. Kulikowski, T. Mroczek (Eds.), Springer, London, 2012, pp. 119-133.

[16] M. Armstrong, "Reliability-Importance and dual failure-mode components", *IEEE Trans. on Reliability*, vol. 46, no. 2, pp. 212-221, 1997.

[17] F.C. Meng, "On some structural importance of system components", *Journal of Data Science*, no,7, pp.277-283, 2009.

[18] http://www.cbl.ncsu.edu:16080/benchmarks/LGSynth91/twolexamples/

3rd International Workshop on Advances in Semantic Information Retrieval

Recent advances in semantic technologies form a solid basis for a variety of methods and instruments that support information retrieval, knowledge representation, and text analysis. They influence the way and form of representing documents in the memory of computers, approaches to analyze documents, techniques to mine and retrieve knowledge. The abundance of video, voice and speech data also raises new challenging problems to information retrieval systems.

We believe that our workshop will facilitate discussion of new research results in this area, and will serve as a meeting place for researchers from all over the world. Our aim is to create an atmosphere of friendship and cooperation for everyone, interested in computational linguistics and information retrieval. The second ASIR workshop will continue to maintain high standards of quality and organization, set in the previous year. We welcome all the researchers, interested in semantics and information retrieval, to join our event.

Topics

The workshop addresses semantic information retrieval theory and important matters, related to practical Web tools. The topics and areas include, but not limited to:

- Domain-specific semantic applications.
- Evaluation methodologies for semantic search and retrieval.
- Models for document representation.
- Natural language semantic processing.
- Ontology for semantic information retrieval.
- Ontology alignment, mapping and merging.
- Query interfaces.
- Searching and ranking.
- Semantic multimedia retrieval.
- Visualization of retrieved results..

Event Chairs

Klyuev, Vitaly, University of Aizu, Japan
Mozgovoy, Maxim, University of Aizu, Japan

Program Committee

Borgo, Stefano, Laboratory for Applied Ontology, Italy
Budzynska, Katarzyna, Institute of Philosophy and Sociology of the Polish Academy of Sciences, Poland
Carrara, Massimiliano, Universita di Padova, Italy
Cybulka, Jolanta, Poznan University of Technology, Poland
Dobrynin, Vladimir, Saint Petersburg State University, Russia
Goczyła, Krzysztof, Gdansk University of Technology, Poland
Haralambous, Yannis, Institut Telecom - Telecom Bretagne, France
Homenda, Wladyslaw, Warsaw University of Technology, Poland
Jin, Qun, Waseda University, Japan
Kaczmarek, Janusz, Lódz University, Poland
Kakkonen, Tuomo, University of Eastern Finland, Finland
Kulicki, Piotr, John Paul II Catholic University of Lublin, Poland
Lai, Cristian, CRS4, Italy
Leonelli, Sabina, University of Exeter, United Kingdom
Ludwig, Simone, North Dakota State University, United States
Martinek, Jacek, Poznan University of Technology, Poland
Mirenkov, Nikolay, University of Aizu, Japan
Morshed, Ahsan, CSIRO ICT Centre, Commonwealth Scientific and Industrial Research Organisation, Australia
Mozgovoy, Maxim, University of Aizu, Japan
Nalepa, Grzegorz J., AGH University of Science and Technology, Poland
Palma, Raúl, Poznan Supercomputing and Networking Center, Poland
Piasecki, Maciej, Wroclaw University of Technology, Poland
Pyshkin, Evgeny, Saint Petersburg State Polytechnical University, Russia
Reformat, Marek, University of Alberta, Canada
Shtykh, Roman, CyberAgent Inc., Japan
Soldatova, Larisa, Brunel University, United Kingdom
Suárez de Figueroa Baonza, Mari Carmen, Ontology Engineering Group, Scool of Computer Science at Universidad Politécnica de Madrid, Spain
Tadeusiewicz, Ryszard, AGH University of Science and Technology, Poland
Trypuz, Robert, John Paul II Catholic University of Lublin, Poland
Vacura, Miroslav, University of Economics, Czech Republic
Vargiu, Eloisa, Barcelona Digital Technology Centre, Spain
Vazhenin, Alexander, University of Aizu, Japan
Wang, Haofen, Shanghai Jiao Tong University, China
Wu, Shih-Hung, Chaoyang University of Technology, Taiwan
Zadrozny, Slawomir, Systems Research Institute, Poland
Ławrynowicz, Agnieszka, Poznan University of Technology, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 243–249

Information Retrieval Using an Ontological Web-Trading Model

José Andrés Asensio
University of Almería
Applied Computing Group
04120, Almería, Spain
Email: jacortes@ual.es

Nicolás Padilla
University of Almería
Applied Computing Group
04120, Almería, Spain
Email: npadilla@ual.es

Luis Iribarne
University of Almería
Applied Computing Group
04120, Almería, Spain
Email: luis.iribarne@ual.es

***Abstract*—One of the biggest problems facing Web-based Information Systems (WIS) is the complexity of the information searching/retrieval processes, especially the information overload, to distinguish between relevant and irrelevant content. In an attempt to solve this problem, a wide range of techniques based on different areas has been developed and applied to WIS. One of these techniques is the information retrieval. In this paper we described an information retrieval mechanism (only for structured data) with a client/server implementation based on the Query-Searching/Recovering-Response (QS/RR) model by means of a trading model, guided and managed by ontologies. This mechanism is part of SOLERES system, an Environmental Management Information System (EMIS).**

I. Introduction

Nowadays, *Web-based Information Systems* (WIS) have become popular as they favour universal access to the information, helping their users to analyze the information from different viewpoints and support group work, decision-making, etc. However, one of the biggest problems of this kind of systems is the complexity of the information searching/retrieval processes, largely due to the huge amount of information they manage.

Their users depend on web sites, digital libraries, engines and other information searching/retrieval systems [1], [2] to help them in this tedious process and, even so, they deal with an overload of information in which they must distinguish between the relevant and irrelevant content. In an attempt to solve this problem, a wide range of techniques based on different areas has been developed and applied: information retrieval, information filtering, studies on information search behavior, etc. Of all these techniques, we focused on the information retrieval in a client/server model for Web systems. In this context, the term "information retrieval" refers to a set of techniques that satisfy the users' information requirements [3].

The main WIS information retrieval mechanism, based on the client/server model, is *the Query-Searching/Recovering-Response* (QS/RR), showed in Figure 1. On one hand, the term "Query" refers to the whole process of creating and formulating the client's request. The term "Searching" refers to the process of locating the data sources (repositories, data storage or databases, regardless of the model) where the information is found, and the term "Recovering" refers to the process of locating, identifying and selecting the data from these sources. Finally the term "Response" refers to the whole process of formulation, preparation and creation of the response by the server to the client. The "Query-Searching" pair is a process that goes from the client to the server. The "Recovering-Response" pair goes from the server to the client.

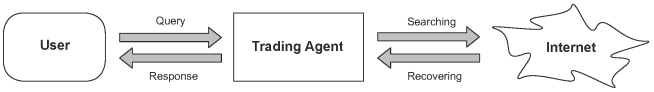


Fig. 1. Overview of the QS/RR mechanism.

A solution to QS/RR mechanism is the UDDI (*Universal Description Discovery and Integration*) specification and WSDL (*Web-Services Definition Language*) for SOA (*Service Oriented Architecture*). They are based on client/server implementations for Web systems. Nevertheless, these techniques allow us to respect a *subscribe/publish/response* model (a QS/RR information retrieval approach) for locating WSDL documents (i.e., XML specifications of web-services) and connecting web services in WIS, but not for different types of information (non-WDSL information). Traders [4] are another solution for open and distributed systems that extend the OMA (*Object Management Architecture*) ORB (*Object-Request Broker*) mechanism. From the viewpoint of the *Open Distributed Processing* (ODP), a Trader (also called trading service, trading function or mediator) is the software object that mediates between objects that offer certain capacities or services and other objects that demand their use dynamically. As is shown in Figure 2, objects that offer their services are called "exporters" and provide the Trader with a description (extra-functional aspects) and an interface (functional aspects) of their service, whereas objects that demand these services are called "importers" and ask the Trader for services with certain characteristics. The function of the Trader, therefore, consists of checking the characteristics required in the descriptions of the services

This work has been supported by the EU (FEDER) and the Spanish Ministry MINECO under grant of the TIN2010-15588, and also by the JUNTA DE ANDALUCÍA excellent project TIC-6114. http://acg.ual.es.

offered (stored in a local repository) and indicating the importer the interfaces of the selected services for his interaction with the exporter.

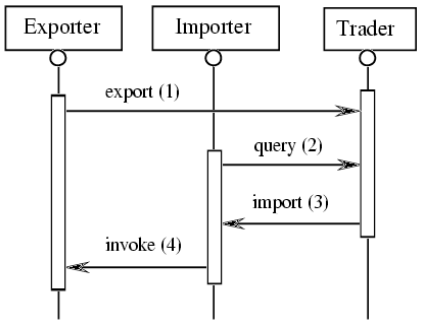


Fig. 2. Roles of the ODP Trader.

There is a large number of studies in which the trading service follows the ODP specification. For instance, in [5] a trading service called DOKTrader is presented, which acts on a federated database system called *Distributed Object Kernel* (DOK). Another example is found in [6]. This study concentrates on the creation of a framework to develop distributed applications for a *Common Open Service Market* (COSM), making use of a *Service Interface Description Language* (SIDL) to describe the services manipulated by the trader. These approaches of the ODP trading implementations have several shortcomings like component interactions, object communications or language description, which have been improved using **ontologies**.

The use of ontologies in trading services has spread, especially in web information services. Ontologies are being used to describe the services offered as well as communication primitives employed by system components. In [7] authors present the design of a market managed by ontologies. Within this system, an ontological communication language is used to represent queries, offers and agreements. Furthermore, in [8], ontologies are used to describe information shared by different system components. To achieve greater operability and autonomous, many systems have chosen to encapsulate the trader object within a software agent. In [9] the MinneTAC agent is described, like a trading agent developed to participate in the Trading Agent Competition (TAC). Through the description of this agent, implementation of a trader as a software agent is shown to maximize benefits from scenarios that require cooperation and negotiation between the trader and the rest of the system components, as well as systems that require communication among various trading objects, making use of ontologies to represent information shared by the agents, whether to describe data and the relationships among variables, as is the case in [10], or defining communication primitives and interaction among agents [11].

In this paper, we propose the Ontological Web Trading (OWT) model that implements a mechanism for solving the complexity of information retrieval in the SOLERES system by means of a trading model for WIS, guided and managed by ontologies. OWT has been implemented in this system as a software agent. SOLERES [12] is an Environmental Management Information System (EMIS) based on satellite images, neural networks, cooperative systems, multi-agent architectures and commercial components. This multi-agent system implements a user Information Retrieval mechanism that implements the QS/RR model and uses the SPARQL query language and the OWL ontology description language to operate. In this system, the ontologies are used in two different contexts: (a) to represent the application domain information itself (**data ontology**), and (b) to request services between agents during their interaction (**service ontologies**). Although a trader agent has five interfaces (i.e., Lookup, Register, Link, Proxy and Admin), this paper discusses only the service and data ontology design features of the Lookup interface, which is used for searching and recovering information. This information should be only structured data. All research work presented here is part of a complete design strategy for *Ontology-Driven Software Engineering* (ODSE) that we are developing in SOLERES.

The remainder of the paper is organized as follows. Section 2 shows the SOLERES system architecture. Section 3 identifies the requirements that an ontological trading service should meet for open and distributed environments as well as the operation models it may carry out. Section 4 describes the Web Trading Agent. Section 5 shows the Lookup ontology used by such agent. We end with some conclusions and prospects for future work in section 6.

II. A Case Study: The SOLERES System

This section presents the main SOLERES system architecture (Figure 3), a spatio-temporal information system for environmental management (an example of EMIS). The general idea of the system is a framework for integrating the disciplines above for "Environmental information" as the application domain, specifically ecology and landscape connectivity. The system has two main subsystems, SOLERES-HCI and SOLERES-KRS. The first is the framework specialized in human-computer interaction. This subsystem is beyond the scope of this article and will not be described. On the other hand, SOLERES-KRS is used to manage environmental information. Examining Figure 3, the `IMI Agent` is like a gateway between the user interface and the rest of the modules, and is responsible for the management of user demands.

Given the magnitude of the information available in the information system, and that this information may be provided by different sources, at different times or even by different people, the environmental information (i.e., the knowledge) can be distributed, consulted, and geographically located in different ambients (i.e., locations, containers, nodes or domains) called Environmental Process Units (`EPU`). Thus the system is formed by a cooperative group of knowledge-based EPUs. These groups operate separately by using an agent to find better solutions (queries on ecological maps).

We accomplished the distributed cooperation of these EPUs by developing a Web Trading Agent (WTA) based on the ODP trader specification and extended to agent behavior.

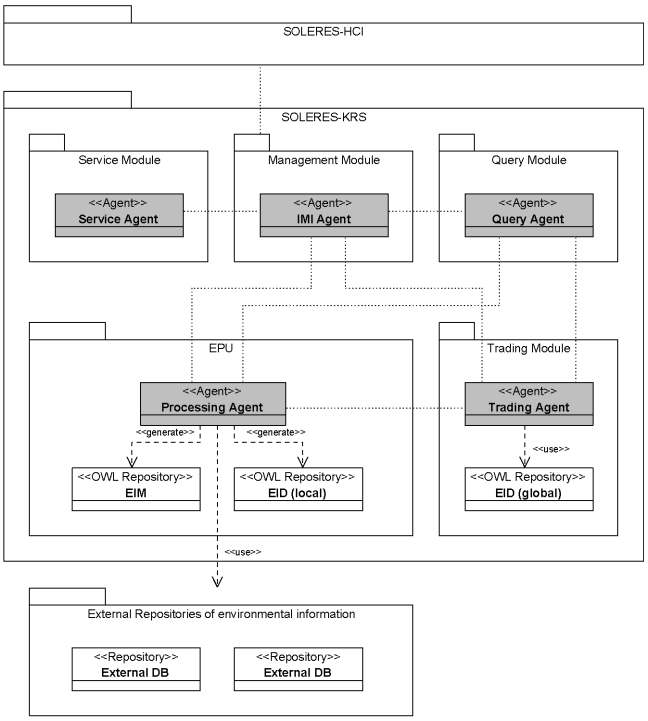


Fig. 3. SOLERES architecture.

Our trading agent mediates between HCI requests and EPU services. EPUs manage two local repositories of environmental information. One of these repositories contains metadata of the information in the domain itself (i.e., basically information related to ecological classifications and satellite images), called *Environmental Information Map* data: EIM documents (`EIM`). This information is extracted from external databases (`External DB` repository in the Figure). The EIM documents are specified by an ontology in OWL [13] (`<<OWL Repository>>`). These EIM documents are the first level of information in SOLERES-KRS.

The second repository contains metadata called Environmental Information metaData, or EID documents (`EID`). These documents contain the most important EIM metadata that could be used by the information retrieval service, and further, incorporate other new metadata necessary for agent management itself. To a certain extent, an EID document represents a "template" with the basic metadata from the EIM document. The EID documents have also been specified by an ontology to accomplish open distributed system requirements. EID documents represent the second level of information in the KRS subsystem. Each EPU keeps its own EID document (or sets of documents) locally and also registers them with the Web Trading Agent (WTA). This way, the WTA has an overall repository of all the EID documents from all EPUs in an ambient and can thereby offer an information search service, as described in the following sections.

III. Requirements and Trading models for OWT

The OWT model developed here is based on the traditional functionality of a trading service, adapted to the management of any type of information (not only on services) through ontologies. Next we will identify a set of requirements necessary for the design of an ontological trading service for open and distributed environments. Later on, we will describe the trading models implemented in our model.

I. Requirements for OWT

For the design of an OWT model, we established a set of properties or requirements that must be met. Table I shows a list of such properties from the ODP standard constraints.

TABLE I. OWT PROPERTIES.

Property	Name
#1	Heterogeneous data model
#2	Federation
#3	Composition and adaptation of services
#4	Weak pairing
#5	Usage of heuristics and metrics
#6	Extensible and scalable
#7	"Storage and forwarding" policy
#8	Delegation
#9	Push and pull storage model

Property **#1** (Heterogeneous data model) means that a trading service should be able to work with different data models and platforms and should not be restricted to just one data model. Thus it should be able to mediate with different protocols of access to information and adapt to the evolution of current and future models.

Property **#2** is related to the federation. For the cooperation among traders there should exist a federation among trading services by using different strategies. For instance, a "repository-based" federation strategy allows more than one service to read and write on the same repository, each being unaware of the presence of others inside the federation, and thus allowing a scalable approach.

Current trading services use "one-to-one" pairing according to the clients' demands and availability of services stored in the repositories they can access. Nevertheless, the ontological trading service should also provide "one-to-many" pairing linking (property **#3**), where a client's query should be satisfied through the composition of two or more instances of metadata available in the repositories.

In the trading service processes, especially those working for open systems (like Internet) where methods and operations refer to the services offered, it is essential to consider the kind of pairing imposed (weak or accurate) (property **#4**), as services are chosen randomly, in an unstandarized way and without agreement. That is why a trading service, when getting the list of chosen metadata during the information searching/retrieval processes, should allow using partial pairing to select (from repositories) those metadata that completely adapt to the request for information or just to a part of it.

Property **#5** points out that a trading service should allow users to specify heuristics and metrics functions when searching for metadata, especially for weak pairing. Thus, among other aspects, the trading service would return results organized according to a search conditions.

Property **#6** defines the extensibility and scalability characteristics of a trading service. Here the trading service should consider any piece of information on services (or metadata) such as data of creators, marketing information and so on, and allow users to independently include new pieces of information for metadata they export (register). In turn, it should be able to use the new piece of information as part of the exported metadata.

In view of a client metadata query, a trading service should retrieve a result. Such result can refer either to a list of chosen metadata that satisfy the query or to a "fail" message if there is no search result. In the latter case, we should also be able to require a trading service to compulsorily satisfy the query or, if that is not the case, store it with the information available by that time and postpone the response until one (or several) metadata providers register (export) a metadata that satisfies the client query. This "response-query" behavior is called behavior "on hold" or "storage-and-forwarding" behavior (property **#7**).

Regarding the previous property, a trading service should also allow delegating (property **#8**) (complete or partial) queries to other trading services if the trading service itself were not able to satisfy such queries.

Property #9 defines the push and pull storage models of a trading service. A push model is the model in which exporters directly get in touch with the trading service to register their metadata. An alternative for metadata registration, suitable for trading services, which work in open and distributed environments on a broad scale, consists of making use of a pull storage model. Here, exporters do not get in touch with traders but rather publish metadata on their websites so that the trading services themselves later on "track" the network in search of new metadata.

Now that the requirements demanded of the trading service have been identified, the OWT model operations in the query process can be described.

II. OWT Model Operations

Let us now see how the OWT model operates in the query process, since an object or component makes a query until the results are retrieved.

This model is a trading-based version of the three-level client/server model. It is comprised basically of a series of elements <I,T,D>, each of which intervenes on a different level, depending on the treatment of the query. Level 1 (L1) is like the client side. Queries are generated and dealt with by an interface object (I). Level 3 (L3) is the server side. System data (D) reside on this level. In our case, these are the EIM repositories with the environmental information. Level 2 (L2) is the middleware that enables the source information to be located. This is the level where the trader objects (T) operate. Associated with the trader (T), the EID repositories with the source environmental information metadata (EIM) also reside there. All three objects use the Lookup ontology (described later) to communicate between them. As the premise for their functioning, an interface object must be associated with a trader object. However, a trader object can also be associated with one or more external data sources or resources, in our case, with the environmental source data (which reside in the EPU, as discussed above). This "trader-information source" association arises from the production of environmental information, where each EPU has an associated trader in which a subset (metadata) of environmental information generated by it is registered. On the other hand, each trader can be associated with one or more traders in federations.

In this three-level architecture, three operating scenarios are possible: *Trading Reflection*, *Trading Delegation*, and *Trading Federation*. Figure 4 shows the three levels (L1, L2, L3), where the three basic objects (I,T,D) reside, and the three scenarios permissible in OWT, as described below.

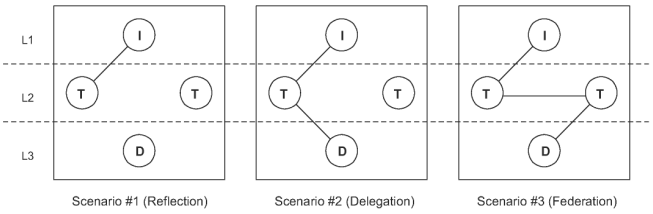


Fig. 4. Operational models of trading.

The **Trading Reflection** scenario in which the query may be solved directly by the trader. The query is generated on the interface and the information can be reached by the metadata that reside in the repository associated with the trader. In this case, the model <I,T> pair intervenes.

The **Trading Delegation** scenario indirectly mediates with the trader. The query is partly resolved by the trader. A query is generated on the interface level that goes on to the trading level (T). The trader locates the data source (or sources) (D), inferring this information from its metadata repository. Therefore, the trader delegates the query to the outside data source (D). In this case, the object series is <I,T,D>.

Finally, the **Trading Federation** scenario is a case in which two or more trader objects are able to federate. As in the cases above, the query remains preset on the interface. This query is passed on to the associated trader object. It can propagate the query to another federated trader object, who locates the external data source (D). In this case the object series intervening is <I,T,T,D>.

For design reasons, the three basic OWT model levels <I,T,D> have been implemented by agents using the JADE platform in the following way. The interface (I) was implemented by means of two agents: the Interface Agent and the IMI Agent. The trading level (T) was implemented by using two other agents: Query Agent and Trading Agent (WTA). The data level (D) was implemented by means of a Resource Agent. From the work perspective presented here, we are interested in the information searching/retrieval

processes, so that the explanation concentrates only on the WTA and the Lookup ontology used for it.

IV. Web Trading Agent

This section describes the internal structure of our Trading Agent and some details about its design and implementation. It should be emphasized that this agent, like all SOLERES system agents, was modeled, designed and implemented based on run-time management of the ontologies used. The trader therefore manages two kinds of ontologies, data and service (or process):

(a) The first is related to the ecological information repositories the trader can access. The information is distributed in different OWL repositories on two levels, as described in *Section II-A*. Some of them contain environmental metadata (EIM repositories) and others contain metadata from the first (EID repositories). A trader manages an EID repository.

(b) The second kind of ontology refers to trader functionality, that is, actions it can do and demand from others. In this case, behavior and interaction protocols must also be defined. These definitions set the operating and interaction rules for agents, governing how the functions the trader provides and demands to work (behavior) are used and the order they are called up in (protocols/choreography).

Figure 5 shows a data ontology from an EID repository (described formally in UML). Let us recall that the application domain to be modeled is ecological information (a type of environmental information) on cartographic maps and satellite images. Advanced algorithms based on neuronal networks find correlations between satellite and cartographic information. For the calculation of this correlation, prior treatment of the satellite images and maps is necessary (an image classification, `Classification`).

A cartographic map stores its information in layers (`Layer`), each of which is identified by a set of variables (`Variable`). For instance, we are using cartographic maps classified in 4 layers (climatology, lithology, geomorphology and soils) with over a hundred variables (e.g., scrubland surface, pasture land surface, average rainfall, etc.).

Satellite images work almost the same way. The information is also stored in layers, but here they are called bands. An example of satellite images is the LANDSAT image, which has 7 bands (but no variables stored in this case). Finally, both the cartographic and satellite classifications have geographic information associated (`Geography`), which is made at a given time (`Time`) by a technician or group of technicians (`Technician`).

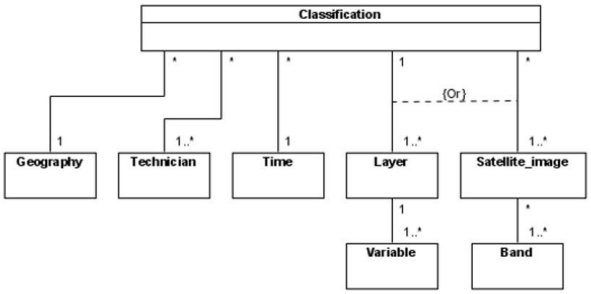


Fig. 5. Ontology of the EID metadata that traders use.

As a complement and formalization of this conceptual model, Table II shows the complete assertions of the eight ontology entities expressed in OCL (*Object Constraint Language*). As an example, we can describe two assertions. The assertion #2 for the `Classification` entity shows it has two required properties, `Classification_id` and `Classification_name`. This entity (classification) is related: (i) either with at least one `Layer` or `Satellite_image` entity (never with both entities simultaneously) through the `classification_shows_layer` or `classification_uses_satellite_image` relationships, respectively; (ii) always with one `Geography` entitythrough the `classification_shows_geography` relationship; (iii) with at least one `Technician` entity through `classification_is_made_by_technician`; and (iv) also with two `Time` entities, `classification_starts_time` and `classification_ends_time`. Analogously, the assertion #4 for the `Layer` entity indicates that it has two

TABLE II. EID ONTOLOGY ASSERTIONS IN OCL.

#	Entity	Assertions
#1	Band	(band_id **exactly** 1) **and** (band_is_shown_by_satellite_image **min** 0) and (band_name **exactly** 1)
#2	Classification	(classification_id **exactly** 1) **and** ((classification_shows_layer **min** 1) **or** (classification_uses_satellite_image **min** 1)) **and** (classification_ends_time **exactly** 1) **and** (classification_is_made_by_technician **min** 1) **and** (classification_name **exactly** 1) **and** (classification_shows_geography **exactly** 1) **and** (classification_starts_time **exactly** 1)
#3	Geography	(geography_id **exactly** 1) **and** (geography_is_shown_by_classification **min** 0) **and** (geography_locality **exactly** 1) **and** (geography_name **exactly** 1) **and** (geography_town **exactly** 1)
#4	Layer	(layer_id **exactly** 1) **and** (layer_has_variable **min** 1) **and** (layer_is_shown_by_classification **exactly** 1) **and** (layer_name **exactly** 1) **and** (layer_observations **max** 1)
#5	Satellite_image	(satellite_image_id **exactly** 1) **and** (satellite_image_is_used_by_classification **min** 0) **and** (satellite_image_shows_band **min** 1)
#6	Technician	(technician_id **exactly** 1) **and** (technician_first_name **exactly** 1) **and** (technician_last_name **exactly** 1) **and** (technician_makes_classification **min** 0) **and** (technician_organization **max** 1)
#7	Time	(time_id **exactly** 1) **and** (time_day **exactly** 1) **and** (time_month **exactly** 1) **and** (time_year **exactly** 1) **and** (time_is_started_by_classification **min** 0)
#8	Variable	(variable_id **exactly** 1) **and** (variable_name **exactly** 1) **and** (variable_is_had_by_layer **exactly** 1)

required properties, `layer_id` and `layer_name`, as well as another optional, `layer_observations`, and it is always related with `layer_is_shown_by_classification` and, at least with one `Variable` through `layer_has_variable`.

The functionality of our trader [14], [15] is divided into three clearly differentiated components (see Figure 6): (a) a component that manages the agent-communication mechanism (**Communication**); (b) a parser that codes and decodes the trading ontology-based messages exchanged (**Parser**); and (c) trading itself (**Trader**).

The third component is inspired by the ODP specification, which indicates how offers and demands must be implemented among objects in a distributed environment and proposes grouping all the different functionalities that a trader may include. Although the standard specifies five trader interfaces (i.e., Lookup, Register, Admin, Link and Proxy), its specification does not demand a trader to implement these five interfaces to work. In fact, we have only developed ontologies for the Lookup, Register, Admin and Link interfaces, but none has been implemented for the last one yet. The Lookup interface offers the search-information in a repository under certain query criteria. The Register interface enables objects in this repository to be inserted, modified and deleted. The Admin interface can modify the main parameters of the trader configuration, and finally, the Link interface makes trading agent federation possible.

As previously explained, this paper focuses on identifying and explaining how ontologies appear and intervene in the Web Trading Agent. Of the interfaces implemented, we only explain here the Lookup interface works, because it takes part in the search, which is the primary subject of this article.

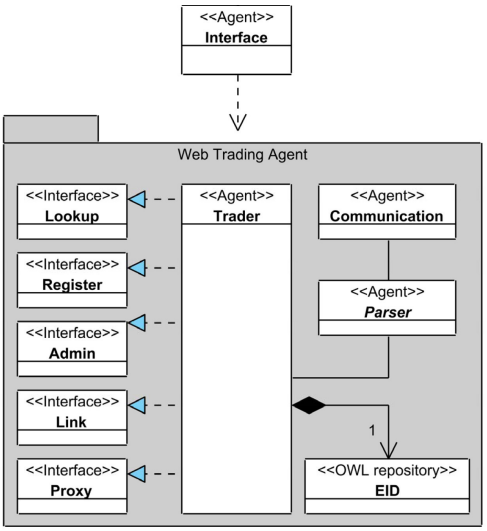


Fig. 6. Web Trading Agent view.

V. The Lookup ontology in OWT

The Lookup ontology (Figure 7) is used between system objects. The trader uses the `Query` action and the `QueryForm` concept. The `QueryForm` concept expresses the query in a specific language, whose properties, among others are: an `id` (a query identifier) and an `uri` (reference to the file where the query is stored). In addition, there could be a set of query policies (`Policy`) through the `PolicySeq` concept, and each "policy" is represented by means of a tuple (name, value). For instance, some of the tuples implemented are:

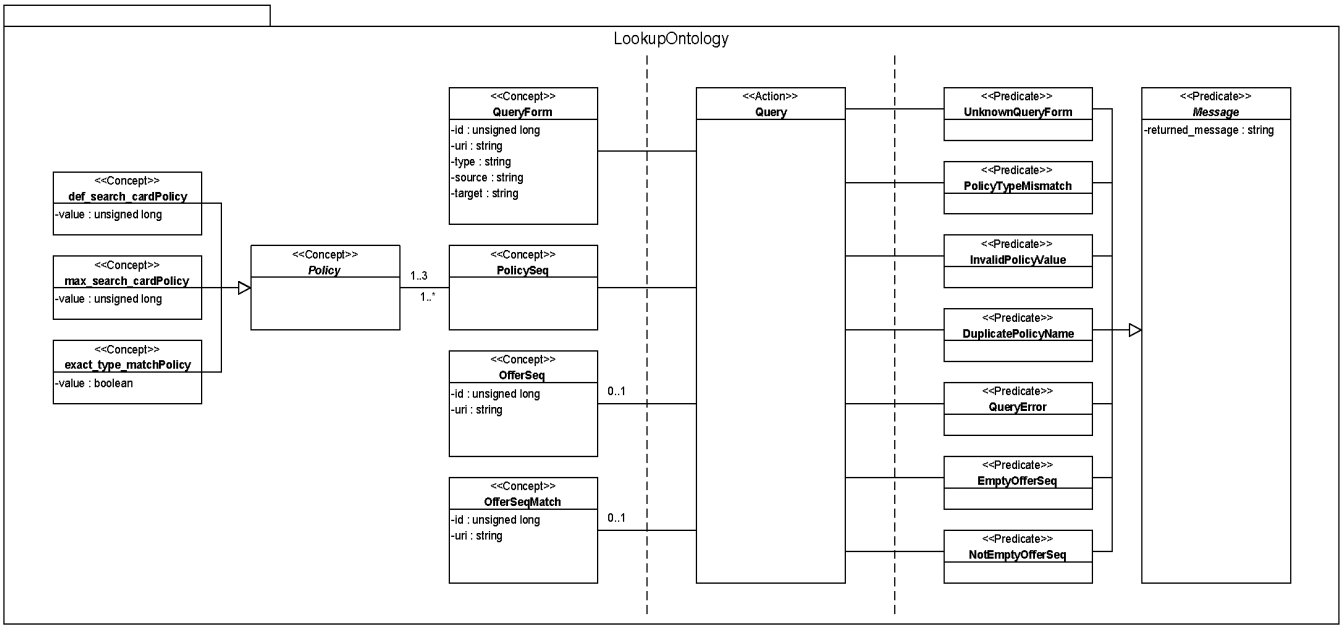


Fig. 7. Lookup Ontology metamodel expressed in UML.

`def_search_cardPolicy` or `max_search_cardPolicy`, indicating the number of records to be located by default, and the maximum number of records to be located in the query, respectively. It is possible some exceptions.

Thus, `UnknownQueryForm` indicates that the query cannot be answered because the file specified in the `uri` is not accessible; `PolicyTypeMismatch` indicates that the type of value specified is not appropriate for the `Policy`; `InvalidPolicyValue` indicates that the Policy value specified is not within the permissible value range for that Policy; `DuplicatePolicyName` indicates that more than one value for the same `Policy` has been specified in the `PolicySeq`; and `QueryError` indicates that an error has occurred during the query. If there is no exception and the query is successfully executed, either the `EmptyOfferSeq` predicate is used when no record is returned by the query, or the `NotEmptyOfferSeq` predicate, when it is. This, in turn, uses the `OfferSeq` concept to represent the set of records located in the query, the properties of which are the query "`id`" and the file "`uri`" where the found records are stored.

VI. Conclusion

Today, web-based EMIS greatly facilitate information search and retrieval, favoring user cooperation and decision-making. Their design requires the use of standardized methods and techniques that provide a common vocabulary to represent the knowledge in the system and a capability for mediation to allow interaction (communication, negotiation, coordination, etc.) of its components. Ontologies are able to provide that shared vocabulary, and trading systems can improve the interoperability of open and distributed system.

The present paper shows how traditional traders, properly extended to operate in WIS, are a good solution for information retrieval. For that we have introduced Ontological Web-Trading (OWT), an extension of the traditional ODP trading service to support ontological information retrieval issues on Web-based EMIS, as is the case of the SOLERES system.

Future work will focus on the implementation of SOLERES-HCI (Human-Computer Interaction). This subsystem of our EMIS is defined by means of the Computer Supported Cooperative Work (CSCW) paradigm [16] and implemented by using an innovative technology of intelligent agents and multi-agent architectures. Furthermore, we are working on this subsystem and studying how to decompose the user tasks into actions that will have to be performed by the SOLERES-KRS subsystem for retrieval of the information requested and the ontology mapping problems involved.

Finally, we would like to study, develop and incorporate new evaluation and validation techniques, such as measuring the precision of data returned to queries, response time in executing the query, usability, etc.

References

[1] Goh, D., Foo, S., 2007. Social information retrieval systems: Emerging technologies and applications for searching the web effectively. Idea Group Reference.

[2] Gama, J., May, M., 2011. Ubiquitous Knowledge Discovery. Intell. Data Anal. 15 (1), 1.

[3] Ramos, A. C., Gensel, J., Villanova-Oliver, M., and Martin, H. (2005). Adapted information retrieval in web information systems using PUMAS. In AOIS, pages 243-258.

[4] Trader, I., 1996. ISO/IEC DIS 13235-1: IT–Open Distributed Processing– ODP Trading Function–Part 1: Specification.

[5] Craske G, Tari Z, Kumar K. R. (1999) DOK-Trader: A CORBA Persistent Trader with Query Routing Facilities. Proc. of the Int. Symp. on Distributed Objects and Applications, pp. 230

[6] Merz M, Müller K, Lamersdorf W. (1994) Service Trading and Mediation in Distributed Computing Systems. Conf. on Distributed Computing Systems, pp. 450–450.

[7] Busse S, Kutsche R. D., Leser U, H Weber (1999) Federated Information Systems: Concepts, Terminology and Architectures. Technical Report 99-9, Technical University of Berlin.

[8] Lamparter S, Schnizler B. (2006) Trading Services in Ontology-driven Markets. Proc. of the 2006 ACM Symp. on Applied Computing, pp. 1679–1683.

[9] Tsai WT, Huang Q, Xu J, Chen Y, Paul R. (2007) Ontology-based Dynamic Process Collaboration in Service-Oriented Architecture. Proc. Service-Oriented Comp. & App., pp. 39–46

[10] Collins J, Ketter W, Gini M. (2009) Flexible decision control in an autonomous trading agent. Electronic Commerce Research and Applications.

[11] Ziming Z, Liyi Z (2007) An integrated approach for developing e-commerce system. Proc. Of the Wireless Communications, Networkig and Mobile Computing, pp. 3596–3599.

[12] Iribarne, L., Padilla, N., Asensio, J., Criado, J., Ayala, R., Almendros, J., Menenti, M., 2011. An Open-Environmental Ontology Modeling. IEEE Trans. on SMC Part A 41 (4), 730–745.

[13] Padilla, N., Iribarne, L., Asensio, J., Muñoz, F., Ayala, R., 2008. Modelling an Environmental Knowledge-Representation System. Lecture Notes in Computer Science (LNCS), 5288: 70-78. Springer. DOI: 10.1007/978-3-540-87781-3_8.

[14] Asensio, J., Iribarne, L., Padilla, N., Ayala, R., 2008. Implementing trading agents for adaptable and evolutive COTS components architectures. In: Proceedings of the International Conference on e-Business, Porto, Portugal. pp. 259–262.

[15] Iribarne, L., Troya, J., Vallecillo, A., 2004. A trading service for COTS components. The Computer Journal 47 (3), 342–357.

[16] Pendharkar, P., 2007. The theory and experiments of designing cooperative intelligent systems. Decision Support Systems 43 (3), 1014–1030.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 251–256

Rhetorical Browzing in Journalistic Texts: Preliminary Investigations

Patrice Enjalbert, Alexandre Labadié, Stéphane Ferrari
Laboratoire GREYC
Université de Caen & CNRS
Bd Maréchal Juin - BP 5186 F
14032 Caen Cedex, France
FirstName.Name@unicaen.fr

***Abstract*—The work presented in this paper concerns discourse structure analysis and its applications to intra- and inter-document search. In a typical application, which could be called "rhetorical browsing", the system will provide assistance to a journal reader in order to focus on texts and passages presenting certain *kind* of information and comments, according to his/her current interest: may be raw information, possibly with chronological dimension, or on contrary analyses, recommendations, debates, etc.. The discourse model can be related to Swales's "discourse moves" and the derived "argumentative zoning" procedures for scientific documents. However due to the nature of the considered texts, zones are defined in more "generalist" terms, following the classic Narration-Description-Argumentation-Prescription typology and especially C. Smith's notion of "discourse modes". The paper presents some preliminary steps performed in order to test the feasibility of the project. First of all, in order to ground our research on firm observations, we decided to build a corpus of journalistic texts, annotated according to the discourse model in view. Quantified results concerning the organization of discourse modes within texts could be obtained thanks to these annotations. In a second step, an experimental procedure for automatic tagging of text passages according to discourse modes has been designed, implemented and tested on the corpus.**

I. Introduction

ONE can currently observe an increasing interest for discourse structure analysis in the NLP community, both for applicative purposes (improvement of document indexation, summarization, document browsing, passage extraction...) and corpus-based linguistic studies. A very popular approach tries to capture text organization in terms of successive "homogeneous" blocks, representing the succession of "topics" addressed in the text. This so-called *thematic segmentation* has received many implementations and experimentations, in the line of Hearst's *Text Tiling* [1].

Rhetorical zoning is a less represented but developing matter. Notably, a number of on-going works are based on Swales' notion of "discourse moves" [2]. Attempts to automatically discover such structures by means of machine learning techniques notably count the pioneer work of [3] for scientific texts, and extensions to other kinds of texts as in [4]. In order to adapt these ideas to our journalistic corpus, we consider a refinement of the Descriptive-Argumentative-Narrative-Prescription model considered (with many variants) in literary studies [5], [6], [7]. According to this model, texts or passages of texts can be labeled by such a *discourse* (or *rhetoric*) *mode*.

Our interest is strongly related with practical concerns. As news readers we observe that, from one reading to another, we may be interested in a different kind of content: maybe raw information, with possibly strong chronological aspects, or on contrary analyses and explanations, recommendations, etc. And not only different papers will match our expectations, but even, especially in long articles, specific passages in them. Hence an interesting consequence of our work would be inter- and intra- document browsing, according to rhetorical and not only topical criteria.

In order to ground our research on firm observations, we decided to build a corpus annotated according to the discourse model in view. The corpus is composed of in-depth articles in economy and politics from the French newspaper *Le Monde*. The annotation task consisted in a labeling of texts passages with a selected set of discourse modes. Quantified results concerning the organization of discourse modes within texts could be obtained thanks to these annotations. In a second step, an experimental procedure for automatic tagging of text passages according to discourse modes has been designed, implemented and tested on the corpus.

The paper is organized as follows. We first describe the corpus, the discourse model, and the annotation procedure. Quantified results concerning the organization of discourse modes within texts are then presented, completed by the description of the automated tagging procedure.

II. Corpus, model and annotation procedure

A. Texts, annotators and tools

The corpus in view is composed of journalistic texts from *Le Monde*, year 1994. This choice is due both to applicative goals and to the linguistic quality of the journal. We randomly selected 30 texts (mainly in politics and economy) of different sizes. The corpus totalizes 46689 words and was shared out among 3 categories: *Small*: less than 1000 words (15 texts); *Medium*: between 1000 and 3000 words (10 texts); and *Large*: more than 3000 words (5 texts). Each text has been annotated by 3 different annotators from a group of 5 with a random distribution between annotators in each size categories. Our 5

978-83-60810-53-8/$25.00

annotators were students in the master degrees of Linguistic and Computer Science.

The annotation was performed under the *Glozz* platform[1]. *Glozz* is based on a generic meta-model which allows to define any specific set of units (segments) and relations with editable features. It proposes a graphical environment and an SQL export, allowing annotations mining through standard database tools [8].

B. Rhetorical and annotation model

The approach of rhetorical structure we consider is coarse-grained and segment-oriented (rather than relation-driven and bottom-up oriented as in discourse models such as RST [9] or SDRT [10]). Generally speaking, a *rhetorical segment* can be defined as filling a specific communicative function. Such segments can be defined in different ways.

One, following [2], is based on the notion of *discourse move*. Moves are conventional parts of the message, specific of a given genre[2]; for example, in scientific articles: context of the study, aim and hypotheses, experiments, results and discussion. In NLP, such a model has been notably worked out in [3] and adapted to other kinds of texts, such as administrative letters, in [4].

Another approach, in a sense more "universalist" is the classic *Narration-Description-Argumentation-Prescription* model [5], [6], [7]. Such *discourse modes* (according to Smith's denomination) may be considered as characterizations applying to full texts or, better, to parts of them. This model appeared to be well suited to our corpus and to the practical goals in view.

However, some adaptations were made. We observe that, in general, several discourse modes are simultaneously present in a same portion of text; for example description is intertwined with argumentation, or with narration. Rather than defining single characterizations of text segments (descriptive *or* argumentative *or* narrative...), discourse modes rather act as "colors" or "shades" that can combine.

Thus, the task of *rhetorical tagging* is described as follows. We make the hypothesis that paragraphs can be considered as relevant textual units: clearly, this hypothesis could be reconsidered but it seems an acceptable first approximation. Rhetorical tagging consists in identifying which discourse modes are present in a given paragraph and with which intensity. We proposed a set of seven discourse modes divided into two main dimensions, *representational* (or *ideational*) and *interpersonal*[3]. Annotators had to allocate a score to seven fields representing the intensity these seven discourse modes.

a) Representational dimension: It concerns the semantic content of the message, the representations construed by the reader. Four graded fields where proposed relative to four rhetoric modes.

Description: Indicates the weight of factual information in the paragraph.
Argumentation-Explanation: Represents to which extent the paragraph is about convincing or explaining something to the reader. We considered that the mechanisms of argumentation and explanation are the same, even if the goals are not.
Chronology: Indicates the weight of temporally marked information in the paragraph.
Prospection: Represents to which extent the paragraph projects the reader into the future.

b) Interpersonal dimension: It concerns the relation between the writer and the reader in the communicative process and includes three fields:
Personal commitment: Does the paragraph reflect the author's personal opinion or is it rather presented as objective ?
Prescription: To what extent the paragraph is about advising or instructing the reader to do something ?
Polyphony: Indicates the weight of directly or indirectly reported speech in the paragraph.
Each of the seven fields is given a score between 0 and 2. **0:** the discourse mode is absent or marginally present; **1:** it is present, but not essential to understand the paragraph; **2:** it constitutes a major key to understand the paragraph.

C. Annotators agreement

We are in the standard situation of a known set of items (for each paragraph, the seven fields corresponding to the seven discourse modes) that receive some "tags" (0, 1 or 2 reflecting the presence and intensity of the mode in this particular paragraph) so that a Kappa measure will do well. *Fleiss' kappa* [11] coefficients for each text are presented in table I. For each text, the given score is the mean value of the scores of all its segments.

TABLE I
FLEISS *kappa* ON RHETORICAL MODES

Text	κ	Text	κ
Large		Small	
0101_31	0.23	0101_1	0.42
0108_96	**0.48**	0101_14	0.41
0204_34	0.31	0101_20	0.36
0628_49	0.29	1229_33	0.39
1220_101	0.28	1230_90	0.33
Average	0.32	1231_75	0.49
Medium		1231_86	0.32
0126_121	0.26	1231_93	0.39
0718_138	−0.01	1231_70	**0.59**
0820_12	**0.41**	1231_84	0.52
0831_135	0.23	1231_89	0.30
1230_24	0.21	0101_13	0.37
0131_108	0.31	0101_19	0.36
0801_108	0.06	0101_6	0.23
0829_120	0.32	1230_3	0.28
0929_135	0.15	Average	0.39
1231_92	0.35		
Average	0.23		

The scores show an average agreement quality ranging from a "low fair" (medium texts) to "almost moderate" (small

[1] http://www.glozz.org/

[2] Where "genre" should be interpreted in a very narrow sense, and better called "micro-genre"

[3] The term "representational" is inspired by Adam's terminology and "ideational" by Halliday's one.

texts)[4]. They may look "modest", but one must keep in mind the highly interpretative nature of the task. Also remind that annotators had to choose between three possibilities; when considering only if the rhetorical color is present or absent all κ raise by, at least, 0.1 (except for one text), leading to a global factor of 0.42, "moderate". The worst score on medium texts seems - according to the post-annotation debriefing - due to a particularly open interpretation of some of these texts.

On the overall, these results seem to show that a form of "convergence" does exists, pleading for the relevance of the model. Improvements could probably be obtained thanks to a better and less ambiguous definition of the discourse modes, for which a careful analysis of the present discrepancies should be helpful. Also differentiated analyses according to the different modes could be interesting: are some of them less controversial than others?

III. The dynamic of discourse modes

All along this section we will use abbreviations indicated in table II to designate the rhetorical modes: DESC for Description, ARGU for Argumentation-Explanation, etc. The first column of the table presents a first bunch of raw observations, namely the distribution of modes along all texts and all annotators. Let us insist that it counts *scores*, i.e. the number of paragraphs annotated according to a particular mode, ponderated by the intensity factor given by the annotator. The sum of all scores for one mode is compared to the sum of all scores for all modes.

Two modes, DESC and ARGU are massively preeminent, while interpersonal ones are globally under-represented. This is clearly related to the kind of texts in the corpus, rather of "objective" nature, which do not include editorials for example, or forums. Let's consider now more elaborate questions.

A. Preferred positions of discourse modes.

Figure 1 shows the repartition of rhetorical modes (evaluated as always in terms of scores) in the beginning, middle and final parts of texts. The global impression may corroborate primary intuitions. One can see two symmetric groups: CHRONO and DESC prefer the beginning with no marked preference for the other two: they present "facts" to be discussed later; while POLY, ARGU, PROS and PRES (which constitute such discussions) rather occur in the end. COMM is more equally distributed.

B. Interdependence relations between discourse modes

Some correlations, positive or negative, between discourse modes may be conjectured from the previous table and observed on specific texts. For example a kind of contravariance between DESC and ARGU. The question arises whether such observation could be confirmed and generalized some way: are DESC and ARGU "generally" contravariant ? Are there other such pairs? In order to investigate this question we computed a statistical correlation coefficient.

[4]According to the interpretation grid of [12]: $\kappa < 0$: poor, $0 < \kappa < 0.2$: slight, $0.2 < \kappa < 0.4$: fair, $0.4 < \kappa < 0.6$: moderate, $0.6 < \kappa < 0.8$: substantial, $0.8 < \kappa < 1$: almost perfect

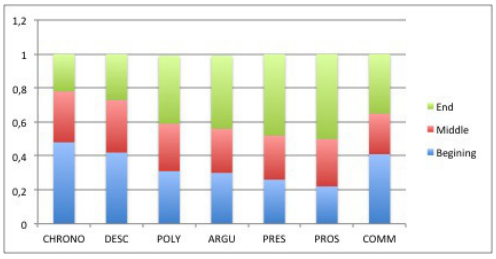


Fig. 1. Distribution of discourse modes along texts.

Results are shown in table III, limited to the four most representative rhetorical modes.

Three, low but perceptible correlations appear: a negative one between argumentation and description, as expected from our "manual" observations; a positive one between argumentation and personal commitment, which is not surprising; and a negative one again, between personal commitment and polyphony: stating other people positions is somewhat exclusive from expressing one's own. It is worth mentioning that, if the correlations are weak in value, the annotators agree on their direction, positive or negative, with one exception out of 18 pair annotator/mode: a fact that tends to strengthen the relevance of the results.

TABLE II
RHETORICAL MODES AND TOPIC TRANSITIONS. (1) THIS MODE/ALL MODES, ANYWHERE. (2) THIS MODE/ALL MODES, RESTRICTED TO TRANSITIONS. (3) THIS MODE IN TRANSITIONS / THIS MODE ANYWHERE. DESC: DESCRIPTION, ARGU: ARGUMENTATION-EXPLANATION, CHRO: CHRONOLOGY, PROS: PROSPECTIVE, POLY: POLYPHONY, PRES: PRESCRIPTIVE, COMM: PERSONAL COMMITMENT

	(1)	(2)	(3)
DESC	0.25	0.26	0.62
ARGU	0.27	0.28	0.6
PROS	0.06	0.04	0.41
CHRO	0.07	0.06	0.52
POLY	0.13	0.15	0.68
PRES	0.1	0.08	0.45
COMM	0.12	0.12	0.59
ALL	1	1	0.58

TABLE III
CORRELATION BETWEEN RHETORICAL MODES.

	ARGU	DESC	COMM	POLY
ARGU	-	-0,18	0.22	0
DESC	-	-	-0,08	0,08
COMM	-	-	-	-0,16
POLY	-	-	-	-

C. Rhetorical profiles.

One can produce graphics such as figures 2 and 3, useful to study the distribution of rhetorical modes along a given text. Rates of the 7 rhetorical modes (vertical axe) are displayed in relation with the successive paragraphs (horizontal axe)[5]. The first graphic concerns a historical-narrative text: the biography

[5]In this example, each graphic concerns a single annotator.

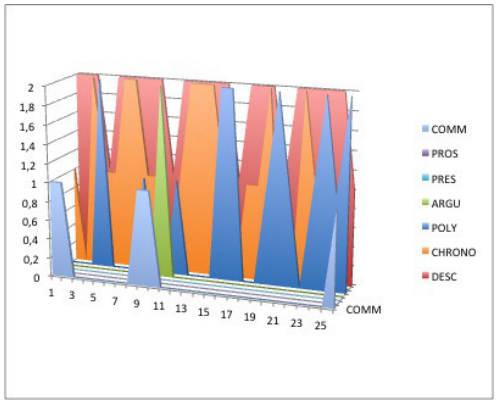


Fig. 2. The dynamic of discourse modes: Biography of an Israel spy.

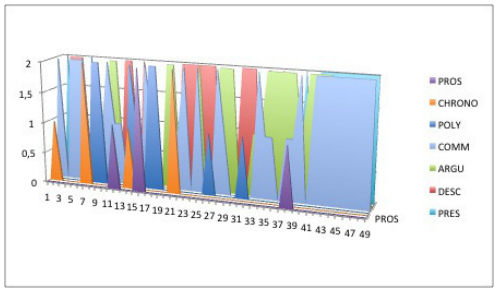


Fig. 3. The dynamic of discourse modes: "How to save Bosnia".

of an Israel spy; the second is an analytic paper about the war in Bosnia. The first thing that appears is the difference in their dominant modes: DESC, CHRONO and POLY for the first, PRES, DESC and ARGU for the second, with also COMM, which more or less coincides with PRES and is hidden by it on the figure. A closer look shows informations about the plan of the texts.

In figure 2 we have a "ground" of descriptive-chronological discourse all along the text, with mainly in the second half a strong polyphonic component (which correspond to discussions and conjectures about the "real" life and activity of this spy). In figure 3 we have a concentration of prescriptive and argumentative discourse (with strong personal commitment) in the second part, while descriptive-chronology-polyphony is rather concentrated in the beginning (stating the history and the problem). These quick observations tend to show, first that a rhetorical dynamic can be analyzed in terms of discourse modes, and second that "text profiles" can be inferred.

IV. Discourse modes and text segmentation

Going further, we can consider the question whether discourse modes by themselves may determine text segmentation i.e. allow to define a succession of segments being "rhetorically homogeneous" in some way. This question was addressed in two different assumptions: a strong one, where a segmentation would be (strictly) determined by changes in the configuration of salient discourse modes - the result seems to be negative; and a weak one, where we consider the contribution of discourse modes to some general "thematic segmentation" - and interesting correlations can be put to light.

A. Attempt towards a pure rhetorical segmentation

We imagined the following experiment. Drawing a parallel with conventional methods in thematic segmentation [1] we considered rhetorical modes as a set of descriptors: the scores given by an annotator to some paragraph defines a *vector* which represents its rhetorical orientation. We can compute angles between successive blocks and deduce continuity or discontinuity according to the angle being smaller or greater than some threshold. Unfortunately the first results are not very convincing: if some "relatively homogeneous" contiguous regions of some extent (several paragraphs) may appear, such text ranges are rather scarce. More subtle measures could probably be considered but in fact the transcription of topical segmentation techniques, based on a geometry of rhetorical descriptors, does not seem to be the good idea.

B. Topical structures and discourse modes

Another way to consider the role of discourse modes in text organization is to look for possible correlations with the topical structure of a text. Here, we took advantage of another annotation of the corpus, performed simultaneously, where annotators were asked to cut out the texts into geat "parts" according to their reader's intuition [13] . This could be called "spontaneous segmentation", as performed by an attentive reader, more or less consciously. The general question is to know what, in this operation, is relative to the "subject" (knowledge domain, discourse referents...) and what to rhetorical features. We were in this matter especially interested in the transitions between "spontaneous" segments and asked the annotators to mark sentences that, according to them, signaled such transitions[6].

Two kinds of investigations were made. The first one continues the geometrical model as above (A), searching for a possible coincidence between rhetorical gaps - measured as angles between "rhetorical vectors" - and topical changes. Unfortunately, at first sight at least, the test fails: there are cases where topical changes are accompanied with large rhetorical angles and cases where not. And the core of topical segments shows both low and hight rhetorical gaps. Again, looking for global configurations of discourse modes appears not to be the right way.

Then we decided to have a more individuated view on each discourse mode and to concentrate on annotated *transitions*: do such zones have specific rhetorical characteristics? Results are figured out in columns 2 and 3 of table II ("Transition" means "paragraph containing an annotated transition zone"). Column (2), when compared to (1), would induce a rather negative result: modes distribution does not reveal significant difference between transitions and other blocks. But (3) shows more positive results.

[6]See for instance [14], [15], [16] for studies on the linguistic characteristics of topical changes.

First we see that introductory blocks contribute for 0.58 to the total score: an important ratio since they constitute only one third of all paragraphs. Beginnings of topical segments are more strongly marked than the others on the rhetorical ground. Then we see that POLY and, with lower strength, DESC clearly prefer transitions. On the opposite side, PROS, PRES and to some extent CHRO are less represented in this position. In other words writers like to begin a new topic by the presentation of different viewpoints or descriptive information and tend to reject prospective, prescriptive or chronological considerations out of this position.

Hence, as one could expect and despite the results of our first test, there seems to be real hints for an implication of rhetorical concerns in topical organisation. On a practical ground, these results could also help in automated segmentation procedures, provided one could find reliable marks of the four distinguished modes, a question addressed in the last part of the paper.

C. Conclusion: what kind of rhetoric zoning?

Gathering the results of sections 3 and 4, some information can be synthesized concerning the organization of discourse modes along texts. The negative - but, still, informative - result is that no clear segmentation (in contiguous blocks) is likely to be based on *global* configurations of discourse modes.

Contrastively, different experiments have shown that, taken *individually*, discourse modes do determine some *zones* according to their salience - which is most important w.r.t. our targeted application. And finally we have seen that the combination of topical and rhetorical features is relevant to the spontaneous segmentation of texts by readers, which might provide hints for improvements in thematic segmentation.

V. Towards an automatic tagging of discourse modes

A. Procedure and implementation

An automatic tagging of text passages w.r.t. the given set of discourse modes appears as a necessary complement to the previous study, both in order to contribute to the validation of the model, and as the basis for the application in view. A first step was performed in this direction as follows [17]. We listed a set of features whose count allows to assign a score to each mode in each paragraph. This score is supposed to reflect the force of the considered mode. In this first experimental attempt, we considered simple features, essentially lexical and morphological ones, as illustrated by the following sample.

- Description: verb tenses representing durative processes (imparfait, présent[7]), spatial locative connectives (prepositions sur/ *on*, dans / *in*...), adjectives and relative pronouns, demonstrative pronouns, named entities.
- Chronology: verb tenses of the past (passé simple, passé composé, participe passé), temporal connectives (conjonctions and adverbials: quand / *when*, puis / *then*, ce matin / *this morning*...), dates.
- Prospection: verb tenses of future and unrealised (futur, conditionnel), cue words (à l'avenir / *in the future*, hypothèse / *hypothesis*, prévoir / *forsee*...)
- Argumentation-Explanation: logical and argumentative connectives (cependant / *however*, donc / *hence*, d'abord / *first*, ensuite / *then*, parce que / *because*...), other cue words (impliquer / *imply*, problème / *problem*, réponse / *answer*...)
- Polyphony: quotation marks, proper names and social functions (as indicating possible authors of reported speech), declarative verbs.
- Prescription: verb mode (impératif), modal verbs (pouvoir / *can*, falloir / *must*, devoir / *must*), other cue terms (important / *important*, essentiel / *essential*...)
- Personal Commitment: logical connectives, epistemic verbs at 1st person (penser / *think*, douter / *doubt*...), other cue terms (respect / *respect*, inquiétude / *concern*...).

One can remark that some features are shared by several modes: it is the co-presence of other ones of the same family that determines *in fine* their rhetorical orientation. The scoring takes this phenomenon into account as illustrated by the following example.

Text : En août, explique Hugues Portelli, qui veille sur les courbes d'opinion au sein du cabinet du premier ministre, il y a eu le consensus monétaire après la crise de juillet et, pour cette fin d'année, le premier ministre a profité, à la fois, de la gestion du conflit d'Air France, des actions menées par Charles Pasqua, notamment contre le FIS, et, enfin, du résultat obtenu sur le GATT.

In August, says Hugues Portelli, who watches over the curves of opinion within the PMO, there was consensus after the monetary crisis in July and, for this season, the Premier took the opportunity in the same time of the conflict management at Air France, the actions taken by Charles Pasqua, especially against the FIS, and finally the result of the GATT.

- Clues for Polyphony: Hugues Portelli, Charles Pasqua [named entities for persons], Expliquer [cue verb]. Score = 3.
- Description: Hugues Portelli, Charles Pasqua, Air France, FIS, GATT [named entities], qui [relative pronoun], sur, au sein de [spatial connective], premier ministre [function name], et [conjonction] (twice), Score = 11.
- Argumentation-Explanation: Expliquer [cue verb], après, et (twice), pour, à la fois, notamment, enfin [logical connectives]. Score = 8.
- Chronology: il y a eu, profité, obtenu [past tense], en août , après, juillet, fin, année, à la fois, enfin [temporal terms]. Score = 10.
- Other modes: 0 clue.

The procedure was tuned on a corpus made of some thirty texts from *Le Monde*, disjoint from the annotated ones, then tested on the later. In this first experiment we limited ourselves to identify the three modes most representative of each paragraph (Description, Argumentation-Explanation and Chronology in the example). An XML output allows to insert

[7]For obvious reasons, tenses are given their French name.

TABLE IV
AUTOMATIC AND HUMAN RHETORICAL TAGGING OF A TEXT

	Annot.1	Annot. 2	Annot. 3	*Automated Annot.*
§1	1. ARGU 2. DESC 3. POLY	1. ARGU 2. COMM 3. POLY	1. POLY 2. PRES	*1. ARGU* *2. DESC* *3. POLY*
§2	1. ARGU 2. DESC 3. COMM	1. ARGU 2. COMM 3. POLY	1. PRES 2. ARGU	*1. ARGU* *2. DESC* *3. POLY*
§3	1. ARGU 2. COMM	1. COMM 2. PRES 3. ARGU	1. PRES 2. ARGU	*1. DESC* *2. ARGU* *3. COMM*

this result in the text.

B. First results and evaluation

We compared our automatic labeling to the manual annotations of the corpus. The result, still qualitative, is that our annotation does not scatter more than what can be noted between human annotators themselves (see an example for a single text in Table IV). In other words, the automatic tagging does not seem better or worse that the manual ones, and is in fact consistent with them. If the evaluation procedure clearly requires to be refined, as well as the manual tagging itself, we believe that this first test may be considered encouraging concerning the feasibility of the task. An important issue to highlight is the underrepresentation of certain modes: Prospection and modes of the interpersonal dimension (with the exception of Polyphony). In the future we therefore need to go beyond the three prevailing modes and probably separate representational and interpersonal ones.

VI. CONCLUSION AND FUTURE WORK

In this paper we have presented a set of investigations on the rhetorical structure of journalistic texts, based on the constitution of an annotated corpus Our first positive result consists in the corpus itself since there is a recognized lack for such resources[8]. The inter-annotators agreement seems acceptable, considering the strongly interpretative nature of the task: generally speaking we believe that, in the case of discourse structure, we have to learn how to cope with this variability rather that try to reduce it to null.

The rhetorical model was designed in terms of discourse modes, due to the application in view, a specific kind of document browsing adapted to journalistic texts. It was correctly received by the annotators, which provides an encouraging hint of its relevance. In particular the idea of a combination of discourse modes in a same passage, with graduation of their salience, seems to receive confirmation.

Several quantified observations, performed thanks to the annotated corpus, give useful informations on the distribution of discourse modes and their contribution to text zoning of some kind. A notion of "rhetorical profile" emerges, combining global dominant modes with the "dynamic" of modes distribution along the text.

[8]Glozz annotations are "stand off" and may be obtained for free from the authors, provided the applicant has him/herself acquired the rights on *Le Monde* corpus.

Finally, a first step was performed towards an automatic tagging in relation to the model.

Further work should include the following questions.

1. An extension of the corpus, in order to give a firmer value to our analyses. A careful examination of the discrepancies between annotators could provide useful hints in order to tune the model and remove ambiguities in its description. Achieving a better inter-annotators agreements would be a good confirmation of these improvements. Another issue would be to split the corpus into different more homogeneous subtypes.

2. Improving the automatic labeling. Other linguistic parameters should be considered. Especially interesting would be aspectual values as described in [6]. Machine learning issues should be considered, but need clearly a great effort in corpus annotation.

3. Finally, the application itself should be considered, which implies to convert "pragmatic" requirements of readers into configurations of rhetorical modes.

REFERENCES

[1] M. A. Hearst, "Texttiling: Segmenting text into multi-paragraph subtopic passages." *Computational Linguistics.*, vol. 23, no. 1, pp. 33–64, 1997.

[2] J. Swales, *Research Genres: Exploration and Application*. Cambridge University Press, 2004.

[3] S. Teufel and M. Moens, "Discourse-level argumentation in scientific articles: human and automatic annotation." *In Proceedings of ACL-99 Workshop "Towards Standards and Tools for Discourse Tagging"*, pp. 84–93, 1999.

[4] D. Biber, U. Connot, and T. A. Upton, *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. John Benjamins Publishing Co, 2007.

[5] E. Werlich, *Typologie der texte*. Quelle and Meyer, 1975.

[6] C. S. Smith, "Discourse modes: aspectual entities and tense interpretation," *Cahiers de Grammaire*, vol. 26, pp. 183–206, 2001.

[7] J. M. Adam, *La linguistique textuelle. Introduction à l'analyse textuelle des discours*. Armand Colin, 2005.

[8] A. Widlöcher and Y. Mathet, "La plate-forme glozz: environnement d'annotation et d'exploration de corpus," *Actes de TALN'09*, 2009.

[9] W. C. Mann and S. A. Thompson, "Rhetorical structure theory: Toward a functional theory of text organization," *Text*, vol. 8, no. 3, pp. 243–281, 1988.

[10] N. Asher, *Reference to Abstract Objects in Discourse: A Philosophical Semantics for Natural Language Metaphysics*. Kluwer Academic Publishers, 1993.

[11] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[12] J. Landis and G. Koch, "A one-way components of variance model for categorical data," *Biometrics*, vol. 33, pp. 671–679, 1977.

[13] A. Labadié, P. Enjalbert, and S. Ferrari, "Transitions thématiques : Annotation d'un corpus journalistique et premières analyses (manual thematic annotation of a journalistic corpus : first observations and evaluation) [in french]," in *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2: TALN*. Grenoble, France: ATALA/AFCP, June 2012, pp. 503–510. [Online]. Available: http://www.aclweb.org/anthology/F/F12/F12-2046

[14] M. Charolles, "L'encadrement du discours : univers, champs, domaines et espaces." *Cahier de Recherche Linguistique*, vol. 6, pp. 1–73, 1997.

[15] N. Asher, P. Dennis, and B. Reese, "Names and pops and discourse structure," in *Workshop on Constraints in Discourse*, Maynooth, July 2006, pp. 11–18.

[16] S. Piérard and Y. Bestgen, "Validation d'une méthodologie pour l'étude des marqueurs de la segmentation dans un grand corpus de textes," *TAL*, vol. 2, no. 47, pp. 89–110, 2006.

[17] A. Attoumani, "étiquetage d'un texte selon différents modes rhétoriques. master's thesis." University of Caen, Tech. Rep., 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 257–260

Similarities in Spaces of Features and Concepts: Towards Semantic Evaluations

Wladyslaw Homenda
Faculty of Mathematics and Information Science
Warsaw University of Technology
ul. Koszykowa 75, 00-662, Warsaw, Poland
Web page: wwww.mini.pw.edu.pl/∼homenda

Agnieszka Jastrzebska
Faculty of Mathematics and Information Science
Warsaw University of Technology
ul. Koszykowa 75, 00-662, Warsaw, Poland
Email: A.Jastrzebska@mini.pw.edu.pl

***Abstract*—The article discusses abstract spaces of concepts and features. Concepts correspond to real-world objects. Concepts are described by their features. The study is devoted to relations in the space of concepts and in the space of features. Of greatest interest is similarity of structures in the concepts and features spaces. There is a direct link between features and concepts. Therefore, similarity may be analyzed through structures of both concepts and features. Authors propose generalized similarity relation, applicable to the developed framework. In addition, similarity of nested sets of the space of features and concepts is discussed. Authors introduce an algorithm, which calculates similarity of two structures of nested structures. Developed semantics leads to the set-theoretic model, which allows to flexibly describe abstract information.**

I. Introduction

SIMILARITY is one of the most important dependencies in our environment. In this article developed framework for constructing generalized similarity relations is presented.

Similarity estimation has to involve narrowing the focus on chosen aspects of compared objects. In our nomenclature object is called concept, object's attribute is a feature. We chose these names intentionally, to highlight that our model may be applicable to modeling in various areas of science. In sections II-A and III the core of developed framework of phenomena description is depicted. We start from the space of features - pieces of information, that describe concepts. In section III-B our own approach to similarity relations modeling in the fuzzified space of concepts and features is introduced. Similarity relations for features vectors are discussed. We introduce also an algorithm for computing similarity of linear orders in the space of features.

The goal of the paper is to present the research on concepts and features spaces descriptions and similarity modeling. Valuation mappings and similarity relations able to process fuzzified descriptions of real-world objects are introduced.

II. Preliminaries

In our approach a start point are features - descriptions of concepts. Features are gathered in vectors. We operate in the namespace of features and in the space of their evaluations rather than on the concepts (objects) themselves. We are interested in similarity of descriptions of concepts, i.e. in vectors of features' evaluations. We assume, that each hypothetical or real concept can be described with qualitatively the same set of features, but evaluated differently.

Due to space limitations we do not present literature review on this topic. Interesting research on similarity can be found not only, but also in: [2], [5], [6], [8] and [9]. It is important to mention, that approaches present in the literature are suitable for similarity based on features. Our model is concepts' oriented in its nature. Therefore, it is necessary to include relations between concepts and features.

We have developed a framework for describing the space of concepts and the space of features. We have also proposed similarity measures dedicated for this model, which we present in the next paragraphs.

A. The space of concepts and the space of features

1) The space of features: A concept corresponds to a real-world object. Usually, due to various constraints and complexity of real-world phenomena, we do not operate directly on concepts. Instead, we describe them with their features. In the developed model the space of features is defined as follows:

$$\mathcal{D} = \{(\mu_1, \ldots, \mu_n) : \mu_i \in [0,1],\ i = 1, \ldots, n,\ n \in \mathcal{N}\} \quad (1)$$

Under our assumptions features are imprecise. Concepts correspond to real-world objects. One of many imprecise information representation models known in literature may be applied, for example: [1] or [3]. Authors treat imprecise information analogically to the uncertainty in the sense of Zadeh. We are aware that there are other frameworks (i.e. probability theory) that are able to describe uncertainty, which we are not recalling here.

Features evaluations are expressed through their degree of membership as a single numerical value from the $[0,1]$ interval. Features vectors belong to the namespace of features. There is unlimited amount of features vectors evaluations, but in our application there is finite amount of features. Of interest is possibility of features space structuring by introducing certain relations, like inclusion, exclusion and overlapping.

2) The space of concepts: $\mathcal{C}$ is the set of all concepts fulfilling some logical conditions, e.g. consumers from a city, pixels from certain images, musical symbols of some score.

$$\mathcal{C} = \{c_1, c_2, \ldots, c_r\} \quad (2)$$

978-83-60810-53-8/$25.00

such that $r \in \mathcal{N}$ is the number of concepts in the space $\mathcal{C}$. The source of information about concepts may be for example measurement devices or questionnaire surveys.

The space of concepts is limited. We are interested in structuring the space of concepts through the space of features, the space of their evaluations and dependencies in these spaces as well as in their subsets. Concepts' space analysis will be performed using relations of similarity, inclusion, exclusion and others.

In the next section similarity relations customized for the developed model are proposed. Presented technique aims at mimicking human way of how similarity is estimated. Concepts are described with their features, comparison happens in a feature-wise fashion. The strength of belongingness of a particular feature to the concept (corresponding to the degree of membership) influences similarity measure.

III. Similarity in the space of concepts and in the space of features

In this section we introduce similarity measures adjusted for spaces of concepts and features.

A. The valuation mapping

Firstly, let us discuss the valuation mapping. It is a relation, that for every vector of features assigns a set of concepts.

$$V : \mathcal{D} \to 2^{\mathcal{C}} \tag{3}$$

For instance, the simplest valuation assigns all concepts with a given features' vector $(\mu_1, \ldots, \mu_n)$ to this vector of features:

$$V(\mu_1, \mu_2, \ldots, \mu_n) = \{c \in \mathcal{C} : d(c) = (\mu_1, \mu_2, \ldots, \mu_n)\} \tag{4}$$

where the mapping $d : \mathcal{C} \to \mathcal{D}$ defines features for concepts. Valuation mapping V generates a subset of the space of concepts, such that each i-th feature of selected concepts was evaluated exactly the same, as the respective i-th feature from analyzed vector $(\mu_1, \mu_2, \ldots, \mu_n)$. Valuation mapping defined in formula 4 is called *pointwise valuation mapping*.

Alternatively, we propose generalized approach to valuation mappings called *filling up valuation mapping* defined as follows:

$$V_I(\mu_1, \ldots, \mu_n) = \big\{c \in \mathcal{C} : d(c) = (\mu_{c1}, \ldots, \mu_{cn}) \text{ and } \mu_{ci} \leqslant \mu_i \text{ for } i = 1, \ldots, n\big\} \tag{5}$$

Filling up valuation mapping generates a subset of the space of concepts, that groups objects, which each i-th feature was not evaluated as greater than respective feature in given features vector. In other words, filling up approach is a generalization of pointwise valuation mapping V, which translates features vectors into groups of concepts. As a result, we extract objects, which description satisfy certain conditions to some point, but not beyond that point. In this study conditions are between 0 and 1, but they may be different if we assume other information representation model (for example: balanced fuzzy sets defined in [3] utilize $[-1, 1]$ interval, intuitionistic fuzzy sets defined [1] employs doubled unit interval $[0, 1]$).

Valuation mappings V and V_I are semantic mappings. They allow transformation from the namespace of features and the space of their evaluations to subsets of the space of concepts. In practice, we use valuation mappings to generate subsets of the space of concepts, in which we are interested in. The key, by which subsets are generated, are features - and that was our initial goal, to describe concepts and groups of concepts by their features.

B. Similarity relations

In this section we introduce similarity relation for features' vectors. Let us assume that we have two vectors of features:

$$\mu_A = (\mu_{A1}, \mu_{A2}, \ldots, \mu_{An})$$
$$\mu_B = (\mu_{B1}, \mu_{B2}, \ldots, \mu_{Bn})$$

Below we introduce a generalized similarity measure $s_{G(eneralized)}$ of two vectors of features.

$$s_G(\mu_A, \mu_B) = \frac{\big|V_I(\mu_A) \cap V_I(\mu_B)\big|}{\big|V_I(\mu_A) \cap V_I(\mu_B)\big| + \mathcal{V}_{A\setminus B} + \mathcal{V}_{B\setminus A}} \tag{6}$$

where:

$$\mathcal{V}_{A\setminus B} = \int_0^{\rho_{max}} \alpha(x) \cdot \mathcal{V}_{A\setminus B}(x)\, dx$$
$$\mathcal{V}_{B\setminus A} = \int_0^{\lambda_{max}} \beta(x) \cdot \mathcal{V}_{B\setminus A}(x)\, dx \tag{7}$$

and

α and β are real nonnegative functions

and

$$\mathcal{V}_{A\setminus B}(x) = \Big|\big\{c \in V_I(\mu_A) \setminus V_I(\mu_B) : \max_{i=1}^{n}\{\mu_{Ai} - \mu_{ci}\} = x\big\}\Big|$$
$$\mathcal{V}_{B\setminus A}(x) = \Big|\big\{c \in V_I(\mu_B) \setminus V_I(\mu_A) : \max_{i=1}^{n}\{\mu_{Bi} - \mu_{ci}\} = x\big\}\Big|$$

and

$$\rho_{max} = \min\big\{\rho \geqslant 0 : (\forall\, i = 1, 2, \ldots, n)\ \mu_{Bi} + \rho \geqslant \mu_{Ai}\big\}$$
$$\lambda_{max} = \min\big\{\lambda \geqslant 0 : (\forall\, i = 1, 2, \ldots, n)\ \mu_{Ai} + \lambda \geqslant \mu_{Bi}\big\}$$

and

$$\mu_c = \big(\mu_{c1}, \mu_{c2}, \ldots, \mu_{cn}\big)$$

is the vector of features of the concept c.

Valuation mapping V_I transforms vectors of features into subsets of the space of concepts satisfying certain conditions (as in formula 5). Generalized similarity s_G is calculated as a fraction. Similarity is enlarged as the size of intersection of compared subsets of the space of concepts grows. Similarity becomes smaller as the number of elements that do not belong to sets' intersection grow. The decreasing effect of features, which are not shared, is conditioned on the difference between these features and vectors μ_A and μ_B. We integrate on the space of features. Integration is done separately for concepts in $V_I(\mu_A) \setminus V_I(\mu_B)$ and in $V_I(\mu_B) \setminus V_I(\mu_A)$. The domain of integration spans from 0 to ρ_{max} or to λ_{max} respectively for these two cases. Functions α and β allow to introduce more

punishing effect of $V_I(\mu_A) \setminus V_I(\mu_B)$ and $V_I(\mu_B) \setminus V_I(\mu_A)$ on the similarity value. α and β may be also used to enhance nonsymmetry of relation s_G.

Let us also introduce a discretized version of similarity relation s_G (named $s_{D(iscretized)}$). For any $0 < \rho$ and $0 < \lambda$ the value of similarity is based on estimations of integrals from the formula 6 in a following manner:

$$\mathcal{V}_{A\setminus B}(x) = \sum_{j=1}^{\rho_{max}} \Big(\alpha(j) \cdot \mathcal{V}_{Aj}\Big)$$
$$\mathcal{V}_{B\setminus A}(x) = \sum_{j=1}^{\lambda_{max}} \Big(\beta(j) \cdot \mathcal{V}_{Bj}\Big) \tag{8}$$

where:

$$\mathcal{V}_{Aj} = \left|\Big\{c \in V_I(\mu_A) \setminus V_I(\mu_B) : \tau_{\rho j}\big(\max_{i=1}^{n}\{\mu_{Ai} - \mu_{ci}\}\big)\Big\}\right|$$
$$\mathcal{V}_{Bj} = \left|\Big\{c \in V_I(\mu_B) \setminus V_I(\mu_A) : \tau_{\lambda j}\big(\max_{i=1}^{n}\{\mu_{Bi} - \mu_{ci}\}\big)\Big\}\right|$$

and

$$\tau_{\rho j}(\text{exp}) \;\equiv\; \rho \cdot (j-1) < \text{exp} \leqslant \rho \cdot j$$
$$\tau_{\lambda j}(\text{exp}) \;\equiv\; \lambda \cdot (j-1) < \text{exp} \leqslant \lambda \cdot j$$

and

$$\rho_{max} = \min\Big\{j{=}1,\ldots : (\forall\, i{=}1,\ldots,n)\ \mu_{Bi} + \rho \cdot j \geqslant \mu_{Ai}\Big\}$$
$$\lambda_{max} = \min\Big\{j{=}1,\ldots : (\forall\, i{=}1,\ldots,n)\ \mu_{Ai} + \rho \cdot j \geqslant \mu_{Bi}\Big\}$$

Analogously to the formula 6, we use filling up valuation mapping V_I. It produces a subset of the space of concepts, that satisfies given conditions to the extent not greater than the conditions stated in the input features vector. By analogy, in the s_D compared sets intersection and outlying parts ($V_I(\mu_A) \setminus V_I(\mu_B)$ and $V_I(\mu_B) \setminus V_I(\mu_A)$) are accounted. Concepts not present in sets' intersection are lying in one of the two remaining parts. They decrease the value of similarity in a nonsymmetrical fashion (through functions $\alpha(j)$ and $\beta(j)$). The decreasing impact of concepts lying in sets' differences is conditioned on the difference between the particular concept μ_c and μ_A or μ_B. We may visualize such „outlying" concepts as crescent-shaped hulls around sets' intersection. These crescent-shaped hulls are divided into up to ρ_{max} and λ_{max} segments. The greater amount of concepts fall to the furthest part of such crescent, the more decreasing effect there is on the similarity value.

We may simplify formula s_D further on. Instead of functions α and β under the sum in formulas 8, we may multiply coefficients by parameter λ or ρ and by j in a following way:

$$\mathcal{V}_{A\setminus B}(x) = \sum_{j=1}^{\rho_{max}} \Big(\rho \cdot j \cdot \mathcal{V}_{Aj}\Big)$$
$$\mathcal{V}_{B\setminus A}(x) = \sum_{j=1}^{\lambda_{max}} \Big(\lambda \cdot j \cdot \mathcal{V}_{Bj}\Big) \tag{9}$$

The proposed idea relies on relation built around sets' intersection and concepts lying beyond this intersection. Alignment of the outlying concepts influences similarity value. Similarity relation's codomain is $[0, 1]$. Note, that it is reflexive. The relation s_G was also intentionally constructed to be nonsymmetric, but they can be adjusted to be symmetric. Asymmetry of similarity relation is a highly desirable property from the applicational point of view. Due to space limitations we do not elaborate on similarity relation properties.

In given definition of valuation mapping V_I, what strikes immediately, is that the similarity relation induced by this mapping may also create linear orders (chains) in the space of subsets of $\mathcal{D}$. In this article we discuss linear orders only. Of interest is similarity of such nested structures. In the next section we present this nontrivial modeling problem to a greater extent.

C. Similarity of linear orders in the space of concepts

In this paragraph we investigate such subsets of the space of features $\mathcal{D}$, that this mapping computes the same value for all features' vectors included in such subset, i.e. $\mathcal{A} \in \mathcal{D}$ is such subset if $(\forall\, \mu_1, \mu_2 \in \mathcal{A})\, V_I(\mu_1) = V_I(\mu_2)$. The structures of such subsets can be formally described as linear orders. Let us recall that linear order is a pair $(X, \leqslant)$, where X is a set of elements and $\leqslant$ is a binary relation satisfying axioms of: antisymmetry, transitivity and totality.

Let us introduce a similarity relation able to compare subsets of the space of features nested in the sense explained above, i.e. $(\forall\, \mathcal{B}, \mathcal{A} \in \mathcal{D})\ \mathcal{B} \leqslant \mathcal{A} \equiv V_I(\mathcal{B}) \subset V_I(\mathcal{A})$. Also, whenever $\mu_A = (\mu_{A1}, \ldots, \mu_{An}) \in \mathcal{A}$ and $\mu_B = (\mu_{B1}, \ldots, \mu_{Bn}) \in \mathcal{B}$ and $\mu_{Bi} \leqslant \mu_{Ai},\ i = 1, \ldots, n$, then $\mathcal{B} \leqslant \mathcal{A}$.

Comparing nested subsets of the space of features requires taking into account not only cardinality of compared subsets, but also actual elements lying inside. Therefore, measures of similarity operating only on cardinalities are not satisfactory to describe such structures. We need to compare actual content of each subset.

In order to compare nested subsets of the space of features we have developed an algorithm, which we describe here. Features nesting is understood in a following way: given features vector $\mu_A = (\mu_{A1}, \mu_{A2}, \ldots, \mu_{An})$ nests features vector $\mu_B = (\mu_{B1}, \mu_{B2}, \ldots, \mu_{Bn})$ if:

$$\mu_{Bi} \leqslant \mu_{Ai},\ i = 1, 2, \ldots, n \tag{10}$$

The space of features (see formula 1) and subsets of the space of concepts (see formula 2) generated with valuation mapping V_I (defined in formula 5) are recalled here. Nested features, through filling up type of transformation enforced by the valuation mapping, generate nested subsets of the space of concepts. Structure of nested sets in the space of concepts corresponds to features' nesting.

If sets of concepts are nested, then each bigger set contains each concept from each smaller set and optionally several more concepts. Let us analyze an exemplar order E_o, which contains following 3 subsets of some space of concepts:

$$P_1 = \{c_1, c_2\}$$
$$P_2 = \{c_1, c_2, c_4\}$$
$$P_3 = \{c_1, c_2, c_4, c_7, c_8\}$$

where P_1 is a subset generated with $V_I(\mu_{P_1})$, P_2 is a subset generated with $V_I(\mu_{P_2})$ and P_3 is a subset generated with $V_I(\mu_{P_3})$. μ_{P_1}, μ_{P_2} and μ_{P_3} are particular features vectors evaluations. For clarity of algorithm description, nomenclature used later refers only to subsets of the space of concepts named as P_m. The method of obtaining these subsets, through valuation mapping V_I, is assumed by default. In the given example of E_o following structure is observed: $P_1 \subseteq P_2 \subseteq P_3$.

In the developed algorithm, written to compare two nested structures of the space of concepts, as input data we have two such structures (denoted as $E1$ and $E2$):

$$E1 = P_{i_1}, \ldots, P_{i_k}$$
$$E2 = P_{j_1}, \ldots, P_{j_l}$$

where $P_{i_1}, \ldots, P_{i_k}, P_{j_1}, \ldots, P_{j_l}$ are subsets of the space of concepts. $E1$ and $E2$ are linear orders, so $P_{i_1} \subseteq \ldots \subseteq P_{i_k}$ and $P_{j_1} \subseteq \ldots \subseteq P_{j_l}$. We do not make any assumptions about the length of orders E1 and E2. $P_{i_1}, \ldots, P_{i_k}, P_{j_1}, \ldots, P_{j_l}$ contain concepts. $P_{i_1}, \ldots, P_{i_k}, P_{j_1}, \ldots, P_{j_l}$ are sets, so the order of appearance of concepts in each set can be omitted. Hence, we always use alphabetical order, to improve efficiency of our algorithm. Each order E can be written as a sequence of concepts, starting from the „deepest" of the nested sets.

The input data to our algorithm are two linear orders $E1$ and $E2$ written as sets in a form (convention) described above.

1) Start with order E, which has last set smaller. If cardinalities of last sets of both orders are equal, choose one order at random. Denote this order as $E_{f(irst)}$ and its last subset as P_{f_u} (it is either P_{i_k} or P_{j_l}).
2) Denote the second order as $E_{s(econd)}$. Denote last (biggest) set of order E_s as P_{s_v}. Search for such set in the order E_s, which shares the biggest number of the same elements (concepts) with P_{f_u} and is the largest. We assumed inclusion (see formula 10), so the last set of E_s, which is P_{s_v} will be always chosen. It is either P_{i_k} or P_{j_l}, the one not chosen in the above point.
3) Collate E_f and E_s in a following way: set P_{f_u} corresponds to set P_{s_v}, set $P_{f_{u-1}}$ corresponds to set $P_{s_{v-1}}$ and so on. If sets from one order run out, assume $\varnothing$.
4) For each pair of sets P_{f_i} and P_{s_i} compute $s_D(P_{f_i}, P_{s_i})$. The formula for s_D is given in 6 with redefined nonoverlapping parts defined by formulas 8 (see section III-B).
5) Similarity of linear orders E_f and E_s is equal to aggregated similarities computed as in point 4, i.e. $\text{aggr}\{s_D(P_{f_i}, P_{s_i}), i = 1, 2, \ldots, \max\{f_u, s_v\}\}$, with some aggregation operator aggr. Note that for linear orders, as in discussed case, collated are simply P_{i_k} and P_{j_l}, $P_{i_{k-1}}$ and $P_{j_{l-1}}$ etc.

Properties of the developed algorithm depend on assumed information representation model, on the similarity relation applied in step 4 and on the aggregating operator calculated in step 5. Due to space constraints we do not compare here various possibilities, which may be chosen in steps 4 and 5. In our first attempt as aggregating function we took mean, as it is very intuitive and normalized measure of dependency between real numbers (and the sum of all $s_D(P_{f_i}, P_{s_i})$ gives us a real number). To maintain comparability, aggregating function should produce normalized values of similarity.

IV. Conclusions

The article discusses developed model of features and concepts spaces. Of interest is similarity between descriptions of real-world objects, which we call concepts. Such descriptions (features vectors evaluations) through valuation mapping generate subsets of the space of concepts. Valuation mapping from the namespace of features into subsets of the space of concepts can be performed in a point-wise fashion or in an filling up way. All concepts form the universe of discourse, on which we do not directly operate. Instead, we use features, which describe concepts.

Two similarity relations developed for this model are introduced. First one is a generalized similarity relation between features vectors. Second one is a discretized version of the generalized similarity relation. Presented measures take into account fuzziness of analyzed information.

In this paper an algorithm of evaluating similarity between structures of nested features vectors based on generalized similarity relation and valuation mapping is introduced.

Acknowledgment

The research is supported by the National Science Center, grant No 2011/01/B/ST6/06478, decision no DEC-2011/01/B/ST6/06478.

A. Jastrzebska contribution is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme (2007-2013) and European Regional Development Fund.

References

[1] K. T. Atanassov: *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems 20, 1986, pp. 87-96.
[2] Belanche L. , Orozco J., *Things to Know about a (dis)similarity Measure*, in: LNAI 6881, 2011, pp. 100-109.
[3] Homenda W., *Balanced Fuzzy Sets*, Information Sciences 176, 2006, pp. 2467-2506.
[4] Hung W., Yang M., *Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance*, in: Pattern Recognition Letters 25, 2004, pp. 16031611.
[5] Julian-Iranzo P., *A procedure for the construction of a similarity relation*, in: proc. of IPMU'08, 2008, pp. 489 - 496.
[6] Klawonn F., Kruse R., *Similarity Relations and Independence Concepts*, in: G. Della Riccia, D. Dubois, R. Kruse, H.-J. Lenz (eds.): Preferences and Similarities, Springer, Wien, 2008, pp. 179- 196.
[7] Orozco J., Belanche L. ,*On Aggregation Operators of Transitive Similarity and Dissimilarity Relations*, w: FUZZ-IEEE, 2004, pp. 1373-1377.
[8] Szmidt E., Kacprzyk J., *A Similarity Measure for Intuitionistic Fuzzy Sets and Its Application in Supporting Medical Diagnostic Reasoning*, in: LNAI 3070, pp. 388-393, 2004.
[9] Tversky, A., *Features of Similarity*, in: Psychological Reviews 84 (4), 1977, pp. 327-352.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 261–265

Antisocial Behavior Corpus for Harmful Language Detection

Myriam Munezero[1] Maxim Mozgovoy[2] Tuomo Kakkonen[1] Vitaly Klyuev[2] Erkki Sutinen[1]

[1]School of Computing
University of Eastern Finland
P.O.Box 111, FI-80101
Joensuu, Finland
Email: {mmunez, tkakkone, sutinen}@cs.joensuu.fi

[2]The University of Aizu
Tsuruga, Ikki-machi,
Aizu-Wakamatsu, Fukushima
965-8580 Japan
Email: {mozgovoy, vkluev}@u-aizu.ac.jp

***Abstract*—We report on experiments that demonstrate the relevance of our *AntiSocial Behavior* (ASB) corpus as a machine learning resource to detect antisocial behavior from text. We first describe the corpus and then, by using the corpus for training machine learning algorithms, we build a set of binary classifiers. Experimental evaluations revealed that classifiers built based on the ASB corpus produce reliable classification results with up to 98% accuracy. We believe that the dataset will be valuable to researchers and practitioners working in preventing, controlling and diagnosing antisocial behavior and related problems.**

I. Introduction

'WHAT is said' is important and can reveal a lot about a person's thoughts, emotions and behavior. It particularly is important, when what is said, expresses feelings or thoughts of harming another. As Biber [1] points out, a writer's thoughts, opinions and attitudes about a topic can be explicitly or implicitly expressed through the choice of word and grammatical constructions. Due to the proliferation of the Internet and Web 2.0, written information on how people feel and their plans and interests is more readily available to researchers studying natural language.

The feelings and actions of harming other human beings can be considered as manifestations of *antisocial behavior* (ASB). ASB is broadly defined as any unconsidered action taken against individuals or groups of individuals that may cause harm or distress to society [2]. Often individuals involved in ASB have disclosed in advance their emotions and plans through oral or written language [3]. Reputedly, the Internet has been used as the outlet for the expression of such emotional states and / or plans of violent acts through the use of blogs or video sites [4]. Moreover, online communication is often used as a way of shouting out their intentions before engaging in their acts of violence [5].

The wealth of antisocial and criminal activity taking place on the Web has resulted in a surge of research interest in the automatic detection of this negative and destructive content. Being able to automatically detect negative content is beneficial, for instance, to managers of websites that allow users to post content or as part of an early warning system to authorities on possible threats to public safety. The automatic detection of ASB could also give rise to self-awareness systems for the individuals that are expressing thoughts or emotions related to ASB.

Identifying the individuals who pose danger to a community involves collecting and analyzing information pertaining to their attitude, thoughts on violence, descriptions of criminal activity and threats among others, including information about homicidal or suicidal ideation [6]. However this information is often difficult to obtain. Reasons such as privacy, legality of the often sensitive information, affect its availability to researchers for analysis.

Hence, albeit the problems antisocial behavior causes, there still does not exist a publicly available corpus of ASB texts. However, research projects that focus on ASB and that have been motivated, for instance by the occurrence of school shootings, require a domain relevant corpus for learning linguistic features that may be used for recognizing future risks of antisocial and destructive behavior from texts.

In this paper, we present such a collection of documents, aimed to remedy the situation. To our knowledge, this is the first attempt to build a corpus with a wide variety of types of antisocial, criminal and extremist content; the previous works have concentrated on a single type of antisocial content such as cyberbullying [7] or forms of extremism [8].

Furthermore, we use our corpus to address the problem of detecting ASB for texts by applying *machine learning* (ML) and text mining techniques. We train ML algorithms with positive examples obtained from the ASB corpus, and with negative examples of antisocial behavior collected from the ISEAR [9], Movie reviews [10] and Wikipedia [11] corpora.

Our experimental results show that classification based on content features discriminates ASB texts from non-ASB texts with accuracy up to 98%. Thus we demon-

This work was supported by: 1) "Detecting and visualizing emotions and their changes in Text" project, No.14166, funded by the Academy of Finland; 2) JSPS KAKENHI Grant Number 25330410.

strate that the ASB corpus can serve as a valuable resource for an ongoing antisocial behavior research.

II. Related Work

While the detection of spam in e-mail messages and web content dates back to the early days of the Internet, detection of antisocial content is a new and emerging area of research interest. The methods applied in the detection of antisocial content draw from the ones developed for detecting spam. In discussing related work, as no previous general models for detecting antisocial behavior from text exist, we provide an overview of the work done in the context of detecting cyberbullying, terrorism and criminal behavior.

Perhaps the most notable related work has been carried out in a research project entitled "Intelligent information system supporting observation, searching and detection for security of citizens in urban environment" (INDECT) [12]. The project aimed at automatic detection of terroristic threats and recognition of serious criminal behavior or violence based on multi-media content. Within the context of INDECT, criminal behavior as "behavior related to terrorist acts, serious criminal activities or criminal activities in the Internet".

Our work differs from the one done in the INDECT project in the focus of the research. While INDECT aims at using the analysis of images, video, and text, our focus is on the analysis of text data.

In their cyberbullying study, Dinakar et al. [7] made use of YouTube comments that involved sensitive topics related to race and culture, sexuality and intelligence. Moreover, Yin et al. [13] in their research made use of online forums for detecting online harassment. Bogdanova et al. [14] in their cyberpedophilia research made use of online perverted journal texts on which to learn models to discriminated pedophiles from non-pedophiles.

Thus, although the corpora used in the studies reported above contain negative behaviors, no corpus has yet addressed the more broad antisocial behavior which as Hanrahan [15] explains is characterized by covert and overt hostility and intentional aggression toward others.

III. Corpora

Textual data is required for analyzing what is said, thought or felt in texts. Unfortunately, when it comes to analyzing antisocial behavior, a suitable text collection is difficult to find. Many of the document collections, for example, those from YouTube and MySpace are generic collections and need to be filtered according to the research area.

It was because of the difficulty and lack of a domain-relevant corpus that we sought to create our own. The corpus can further drive the study of linguistic patterns and emotional content present in ASB texts.

The following subsections describe each corpus used in the experimental study. As we are firstly concerned with the binary classification analysis (that is either a document is deemed as being antisocial behavior or it is not), we therefore collected both positive (Subsection A) and negative (Subsections B-D) examples of non-antisocial behavior texts. To obtain the negative examples of antisocial behavioral, we used popular sentiment corpus (movie reviews [10]), emotion annotated corpus (ISEAR [9]) and factual Wikipedia texts extracts [11]. Table 1 summarizes the documents collected.

A. Antisocial Behavior Corpus

As part of a bigger project that involves detecting antisocial behavior from text, we have created a corpus of aggressive, violent, and hostile texts[1]. Two researchers searched online content in order to collect the documents from various blog posts and news-websites which they could conclusively identify as being ASB. In total 148 documents were identified as ASB. The collection is all English texts, having topics such as: serial killer manifestos, antisocial texts, terrorism, violence-based texts, and suicide notes.

Importantly, the messages in these documents are reflective of the author's thoughts and emotions. The corpus was collected specifically for the purpose of detecting antisocial behavior, conflict, crime and violence behavior from text documents. The collection is based on the research on antisocial behavior that has shown that aggression, violence, hostility, and lack of empathy are among the traits that are most directly associated with ASB [16], [17]. Antisocial behavior also has strong links to negative emotions, such as anger, frustration, arrogance, shame, anxiety, depression, sadness and fear [18]. The link of emotions to antisocial behavior will guide our future research.

B. International Survey on Emotion Antecedents and Reactions (ISEAR)

The ISEAR corpus is a collection of student reports on situations in which the respondents felt any of the seven major emotions: joy, fear, anger, sadness, disgust, shame, and guilt. The responses include descriptions of how they appraised the situation and how they reacted [9].

C. Movie Reviews

This collection consists of 2000 movie reviews. They are labeled in respect to their polarity: negative and positive. The corpus was first used in [10], and now is often applied in sentiment analysis and opinion mining research as a standard development and test set.

D. Wikipedia Text Extracts

We searched and collected Wikipedia articles by using similar concepts such as those we found to be characteristic ASB: killing, terror, violence, aggression, and frustration. The aim of including these texts was to observe how

[1] The current work-in-progress version of the corpus is available upon request.

well our classification algorithms could distinguish between antisocial behavior texts and informative texts containing similar keywords.

TABLE I
CORPORA DESCRIPTION WITH SOURCE, NUMBER OF FILES AND AVERAGE FILES SIZE

Corpus	Source	Documents	Avg. File Size (characters)
ASB	blog posts	148	680
ISEAR	[9]	265	110
Movie reviews	[19]	178	390
Wikipedia extracts	[11]	212	680
Total		803	

IV. EXPERIMENTAL SETUP

In order to test the corpus, we approached the ASB detection problem as a classification task. We performed the step-by-step process outlined in Figure 1.

A. Preprocessing Data

We processed each collected online entry or blog post as a whole. That is we assigned the whole text or message as being antisocial behavior or not. From the corpora, we have two fields; a text field consisting of the message and a binary class label (1 = antisocial behavior, 0 = non-antisocial behavior).

The message field needs to be preprocessed because it contains unstructured text. We applied further preprocessing using WEKA utility (StringToWordVector) that performs tokenization, stemming, and stop/frequent word removal.

B. Machine Learning-Based Classification

For classifying the documents into the two classes, we experimented with three supervised ML classifiers: Naïve Bayes Multinomial, SMO for the implementation of Support Vector Machines, and J48 for Decision Trees. The three selected algorithms have shown to be effective in various text classification studies. We made use of the WEKA tool for the above classifiers.

As a first experiment with the corpus, we used a Vector Space Model approach so as to consider the words as independent entities. The model makes an implicit assumption that the order of words in document does not matter, also called the Bag-of-Words (BoW) assumption [20]. The approach is sufficient for the classification task, as the collection of words appearing in the document (in any order) is usually sufficient to differentiate between semantic concepts [20]. Each document in the corpora was represented as a feature vector composed of binary attributes for each word that occurs in the file.

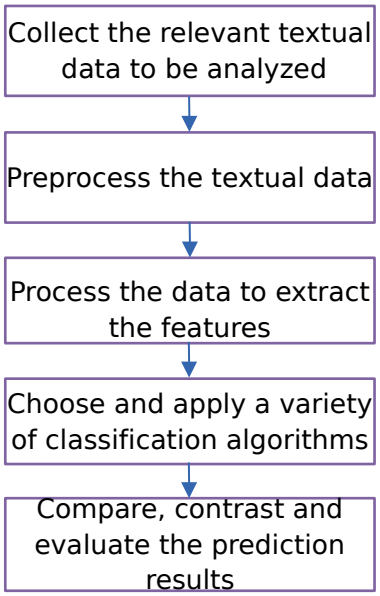


Fig 1. Process map. (Adapted from [20])

Let $\{f_1, \ldots, f_m\}$ be a predefined set of m features that can appear in a document. Let $n_i(d)$ be the number of times f_i occurs in a document d. Then each document d is represented by the document vector $d:=(n_1(d), n_2(d), \ldots, n_m(d))$ [10]. If a word appears in a given file, its corresponding attribute is set to 1, otherwise it is set to 0. Generally, the BoW approach works well for text classification. However, it does not take into consideration any semantic and contextual information.

Moreover, in order to reduce the number of words in the BOW representation we used the LovinsStemmer in order to replace each word by its stem.

We experimented with the three classifiers:

Multinomial Näive Bayes (NBM). With the Näive Bayes classifier, the input is assumed to be independent. The NB classifier, given the data estimates the probability of a class which is proportional to the probability of the class times the probability the data given the class [20]. In other words, the NB classifier assigns a given document d the class $c^* = \text{argmax}_c\ P(c|d)$ [10]. We used the Multinomial Näive Bayes classifier implemented in WEKA, which uses a multinomial distribution for each of the features.

Support Vector Machine (SVM). The classification method of SVM is based on the maximum margin hyperplane rather than probabilities as the Naïve Bayes [20]. In particular, the SVM classifier in a binary classification case aims to find a hyperplane, represented by a vector that maximally separates the document vectors in one class from those in the other [10].

J48 Decision Tree (J48). This classifier is an implementation of the C4.5 decision tree in WEKA. Decision trees are predictive machine models that are used for classification tasks by starting at the root of tree and moving through it until a leaf is encountered [21]. The decision

tree is built from the input training data using the property of information gain or entropy to build and divide nodes of the decision tree in a manner that best represents the training data and the feature vector [7].

The evaluation of the classifiers is discussed in the next section.

V. Results

For an exploratory purpose, we conducted four experiments using the ASB corpus for classifying emotional sentences.

We made use of three corpora as negative examples of ASB: ISEAR, Movie reviews, and Wikipedia extracts as described in Subsection B together with the positive examples of ASB to train supervised ML algorithms. In the first experiment, binary classifiers using the three algorithms were trained on ASB+ISEAR, in the second on ASB+Movie reviews, and in the third on ASB+Wikipedia extracts. Finally, all the corpora were combined.

The performance of the classifiers was then compared in terms of accuracy, precision, recall amd F-measure. For baseline values, we made use of the ZeroR classifier from WEKA which classifies data into the most frequent class in the training set. We made use of ten-fold cross validation whereby samples of data are randomly drawn for analysis and the classification algorithm then computes predicted values [20]. Table 2 displays the average of the ten-fold cross validation results on the corpora for each of the ML techniques.

Based on the results shown in Table 2, the ML algorithms NBM, SMO, and J48 clearly surpass the baseline performance. They further show that for our experiments the NBM and SMO algorithms have the highest accuracy rates. The use of the global corpus (All) also resulted in high accuracy results, as it contains heterogeneous data, however, the difference between the SMO accuracy results and the baseline is much lower. With the global corpus, SMO is statistically better than the next-best classifier (NBM) with a confidence level of about 96% based on the accuracy rate. Its F-measure (0.96), a function of both precision and recall, further indicates a high accuracy.

The experimental results illustrate that from our collected corpus, we can successfully classify antisocial behavior type of texts.

VI. Conclusion and Future Work

In this paper, we applied text classification techniques for the detection of antisocial behavior. In order to accomplish our task we applied various classification algorithms.

Our experimental results show that the task can be successfully accomplished. Experiments show that we achieve high accuracy using Naïve Bayes Multinomial and SMO.

In this paper we have used individual words as features without any additional syntactic or semantic knowledge. In future we are planning to incorporate emotion related information that may positively affect the accuracy of the task.

Ideally, text mining techniques are applied to corpora containing thousands or even millions of documents. In

Table II.
Results for the ASB corpora using the accuracy rate (%)

Corpora	Classifier	Accuracy	Precision	Recall	F-measure
ASB + ISEAR	NBM	**94.91**	**0.95**	**0.94**	**0.95**
	SMO	93.94	0.94	0.93	0.93
	J48	87.89	0.87	0.87	0.87
	Baseline	64.16	0.41	0.64	0.50
ASB + Movie Review	NBM	**98.61**	**0.98**	**0.98**	**0.98**
	SMO	95.83	0.95	0.95	0.95
	J48	90.27	0.90	0.90	0.90
	Baseline	58.88	0.34	0.58	0.43
ASB + Wikipedia	NBM	95.15	0.95	0.95	0,95
	SMO	**95.64**	**0.95**	**0.95**	**0.95**
	J48	88.13	0.88	0.88	0.88
	Baseline	64.16	0.41	0.64	0.50
All	NBM	94.82	0.81	0.93	0.87
	SMO	**96.46**	**0.96**	**0.96**	**0.96**
	J48	92.92	0.92	0.92	0.92
	Baseline	81.31	0.66	0.81	0.72

this case, fewer than 200 records were used that could be confidently identified as antisocial behavior. For further linguistic pattern analysis, a larger corpus will need to be attained. In order to attain a larger corpus, we will incorporate semi-automated methods that will ensure that each topic in the corpus is sufficiently represented.

With the larger corpus, researchers can identify features such as the presence of emotions, causal events or linguistic patterns that pertain to ASB which can be used to train ML algorithms. The main purpose of the corpus is for it to be used as a ML resource.

However, despite these limitations, the created corpus proved to be effective in training ML algorithms.

REFERENCES

[1] D. Biber, *University language: A corpus-based study of spoken and written registers*: John Benjamins Publishing Company, 2006.

[2] R. Card and R. Ward, *The Crime and Disorder Act 1998*. Bristol, England: Jordans, 1998.

[3] M. E. O'Toole, *The school shooter: A threat assessment perspective*. Quantico, Va: Critical Incident Response Group (CIRG), National Center for the Analysis of Violent Crime (NCAVC), FBI Academy, 2000.

[4] S. Crowley, *Finland shocked at fatal shooting.* Available: http://news.bbc.co.uk/1/hi/world/europe/7084045.stm (2013, Mar. 15).

[5] N. Böckler, T. Seeger, P. Sitzer, and W. Heitmeyer, *School Shootings: International Research, Case Studies, and Concepts for Prevention*. Dordrecht: Springer, 2012.

[6] M. Logan, *Case Study: No More Bagpipes. The Threat of the Psychopath.* Available: http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/july-2012/case-study (2013, Mar. 15).

[7] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying," in *International Conference on Weblog and Social Media-Social Mobile Web Workshop*, 2011.

[8] A. Abbasi, "Affect intensity analysis of dark web forums," in *Intelligence and Security Informatics, 2007 IEEE*: IEEE, 2007, pp. 282–288.

[9] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning," *Journal of personality and social psychology*, vol. 66, p. 310, 1994.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*: Association for Computational Linguistics, 2002, pp. 79–86.

[11] Wikimedia Foundation, *Wikipedia: The Free Encyclopedia* (2013, May. 08).

[12] The INDECT Consortium, *XML Data Corpus: Report on Methodology for Collection, Cleaning and Unified Representation of Large Textual Data from Various Sources: News Reports Weblogs Chat.* Available: http://www.indect-project.eu/files/deliverables/public/INDECT_Deliverable_4.1_v20090630a.pdf (2010, Dec. 10).

[13] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection of harassment on web 2.0," *Proceedings of the Content Analysis in the WEB*, vol. 2, 2009.

[14] D. Bogdanova, P. Rosso, and T. Solorio, "Modelling fixated discourse in chats with cyberpedophiles," in *Proceedings of the Workshop on Computational Approaches to Deception Detection*: Association for Computational Linguistics, 2012, pp. 86–90.

[15] C. Hanrahan, "Antisocial Behavior," in *The Gale encyclopedia of children's health: Infancy through adolescence*, K. M. Krapp and J. Wilson, Eds, Detroit: Thomson Gale, 2005.

[16] D. Clarke, *Pro-Social and Anti-Social Behaviour*. Abingdon, UK: Taylor & Francis, 2003.

[17] W. G. Parrott, *Emotions in social psychology: Essential readings*. Philadelphia: Psychology Press, 2001.

[18] L. J. Cohen, "Neurobiology of Antisociality," in *Neurobiology of exceptionality*, C. Stough, Ed, New York: Kluwer Academic/Plenum Publishers, 2005.

[19] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*: Association for Computational Linguistics, 2004, p. 271.

[20] G. Miner, J. Elder, T. Hill, R. Nisbet, D. Delen, and A. Fast, *Practical text mining and statistical analysis for non-structured text data applications,* 1st ed. Waltham, MA: Academic Press, 2012.

[21] J. R. Quinlan, *C4.5: Programs for machine learning*. San Mateo, Calif: Morgan Kaufmann Publishers, 1993.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 267–273

An Approach for Developing a Mobile Accessed Music Search Integration Platform

Marina Purgina
Institute of Computing and Control
St. Petersburg State
Polytechnical University
St. Petersburg, Russia, 194021
Email: mapurgina@gmail.com

Andrey Kuznetsov
Institute of Computing and Control
St. Petersburg State
Polytechnical University
St. Petersburg, Russia, 194021
Email: andrei.n.kuznetsov@gmail.com

Evgeny Pyshkin
Institute of Computing and Control
St. Petersburg State
Polytechnical University
St. Petersburg, Russia, 194021
Email: pyshkin@icc.spbstu.ru

***Abstract*—We introduce the architecture and the data model of the software for integrated access to music searching web services. We illustrate our approach by developing a mobile accessed application which allows users of Android running touch screen devices accessing several music searchers including *Musipedia*, *Music Ngram Viewer*, and *FolkTuneFinder*. The application supports various styles of music input query. We pay special attention to query style transformation aimed to fit well the requirements of the supported searching services. By examples of using developed tools we show how they are helpful while discovering citations and similarity in music compositions.**

I. Introduction

A VARIETY of multimedia resources constitutes considerable part of the present-day Web information content. The searching services usually provide special features to deal with different types of media such as books, maps, images, audio and video recordings, software, etc. Together with general-purpose searching systems, there are solutions using specialized interfaces adopted to the subject domains. Truly, quality of a searching service depends both on the efficiency of algorithms it relies on, and on user interface facilities. As shown in our previous work, such interfaces include special syntax forms, user query visualization facilities, interactive assisting tools, components for non-textual query input, interactive and "clickable" concept clouds, and so on [1]. Depending on searching tasks, specialized user interfaces may support different kinds of input like mathematical equations or chemical changes, geographic maps, XML-based resource descriptions, software source code fragments, editable graphs, etc.

In text searching such aspects as morphological and synonymic variations, malapropisms, spelling errors, and time dependency condition particular difficulties of a searching process. In the music searching domain there are specific complications like tonality changes, omitted or incorrectly played notes or intervals, time and rhythmic errors. Thus, although there are eventual similarities between text and music information retrieval, they differ significantly [2].

In our previous work (see [3]) we analyzed and developed an improved EMD algorithm used to compare single voice and polyphonic music fragments represented in symbolic form.

Human ability to recognize music is strongly interrelated to listener's experience which may be considered itself to be a product of music intelligent perception [4], [5]. Recently (see [6]) we also analyzed internal models of music representation (with most attention to a function-based representation) being the foundation of various algorithms for melody extraction, main voice recognition, authorship attribution, etc. Music processing algorithms use the previous user experience implicitly. As examples, we could cite the *Skyline* melody extraction algorithm [7] based on the empirical principle that the melody is often in the upper voice, or *Melody Lines* algorithms based on the idea of grouping notes with closer pitches [8].

The remaining text of the article is organized as follows. In section II we review music searching systems and approaches of the day. We also introduce our experience in the domain of human centric computing and refer to some recent related works. In section III we describe music query input styles and analyze possible transformations of music input forms so as to fit the requirements of searching services. Section IV contains the description of the developed Android application architecture. We show how it works and make an attempt to analyze the searching output from the point of view of a musicologist.

II. Background and Related Works

Apart from searching media by metadata descriptions (like author, title, genre, or production year information), main scenarios of music searching are the following:

1) Searching music information by existing audio fragment considered as an input.
2) Searching music by the written note score.
3) Searching compositions by human remembrance represented in a form of singed, hummed, tapped or anyhow else defined melody or rhythm fragment.

Searching by given audio fragments is supported by many specialized search engines such as *Audiotag*, *Tunatic* or *Shazam*[9]. As a rule, it is implemented on the basis of so called audio fingerprinting technique. The idea of this approach is to convert an audio fragment of fixed length to a low-dimensional vector by extracting certain spectral features from the input signal. Then this vector (being a kind of audio spectral fingerprint) is compared to fingerprints stored in some database [10], [11].

978-83-60810-53-8/$25.00

TABLE I
ACCESSING MUSIC SEARCHING WEB SERVICES

Name	*Access*				
	GUI	*Micro*	*Text*	*Soft*	*Web*
Audiotag	+	–	–	–	+
Tunatic	+	+	–	–	–
Shazam	+	+	Partially	–	+
Midomi	+	+	–	–	+
Musipedia	+	+	Partially	SOAP	+
Ritmoteka	+	–	–	–	+
Songtapper	+	–	–	–	+
Music Ngram Viewer	+	–	+	REST/JSON	+
FolkTuneFinder	+	–	+	REST/JSON	+

The second scenario *Searching by note score* implicates two possibilities. The first one is to look for a given music fragment in the note sheet databases (this is beyond the scope of this paper). The main difficulty demanding using special music oriented algorithms is that the note score may contain user errors. The second possibility is to convert the melody into one of forms discussed hereafter so as to use existing searching facilities.

Finally, in the case of *Searching by human remembrance*, a system deals with main voice, rhythm, melody contour or interval sequence which is usually not perfectly defined. Hence, it's impossible to search directly within the binary contents of audio resources: we have no faithful audio fragment.

Thus, there are numerous ways to provide music input for a searching system. The extensive description of input styles used by music search engines may be found in [12]. Presently there are many searching web services allowing customers using one of several possible styles to input a music query including such services like *Midomi*, *Musipedia*, *Ritmoteka*, *Songtapper*, *Music Ngram Viewer*, and *FolkTuneFinder*.

Table I represents possible ways to access different music web searching services and pays attention to the following facilities:

- **GUI** Graphical user interface
- **Micro** Using microphone
- **Text** Text mode
- **Soft** Access for software applications by SOAP or similar software interaction protocols
- **Web** Web interface

As you can see, nowadays many services are accessible via browsers since they support Web interface features. Another important issue is the possibility to access some services from inside the software applications by using open protocols. It gives the way to create tools which allow users not to be limited by only one service at a time. Due to such tools we are able to use integrally different features of different searchers, to access additional resources such as *youtube* clips, and to deal with additional search attributes like genres, styles, time periods, etc.

Particularly, *Musipedia* service uses SOAP protocol described in [13]. *FolkTuneFinder* and *Music Ngram Viewer*

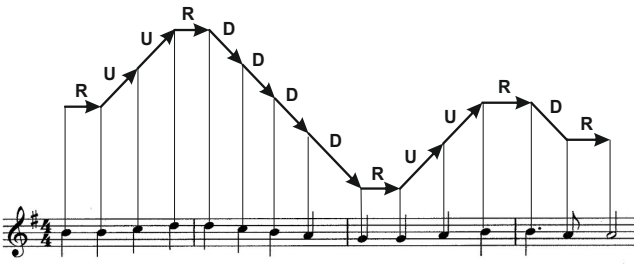


Fig. 1. Beethoven's "Ode to Joe" fragment represented in Parsons code

(both are also used in our work as target searching services) are based on the REST architectural style and their responses are wrapped in JSON format. Detailed description of the API usage rules and examples for *Music Ngram Viewer* service may be found in [14].

In respect to music inputs styles, existing tools support the following opportunities to define a music fragment:

- To sing or to hum the theme and to transfer the recording to the music search engine.
- To write notes by using one of known music notations directly (e.g. music score, Helmholtz or American pitch notation, MIDI notation, etc.).
- To tap the rhythm.
- To play the melody with the use of a virtual keyboard.
- To use MIDI-compatible instrument or it's software model.
- To enter Parsons code, or to set the melody contour by using "U, R, D" instructions [1] as shown in Figure 1.
- To define keywords or enter text query.

Table II lists some known music web searching services together with information about supported music query input styles, namely:

- **Audio** Audio fragment
- **Notes** Music score or pitch notation
- **Hum** Singing or humming
- **Rhythm** Tapping the rhythm
- **VKB** Virtual keyboard generating note sequence with rhythm
- **URD** Parsons code
- **MIDI** MIDI-piano
- **Text** Keywords or text query

III. MUSIC QUERY INPUT STYLES

As shown in the above section, we may define the music query by using different input styles. For a searching frame-

[1]Each pair of consecutive notes is coded as **U** ("sound goes Up") if the second note is higher than the first note, **R** ("Repeat") if the consecutive pitches are equal, and **D** ("Down") otherwise. Some systems use **S** ("the Same") instead of **R** to designate pitch repetition. Rhythm is completely ignored.

TABLE II
MUSIC SEARCHING WEB SERVICES INPUT STYLES

Name	*Web link*	*Input style*							
		Audio	*Notes*	*Hum*	*Rhythm*	*VKB*	*URD*	*MIDI*	*Text*
Audiotag	http://www.audiotag.info	+	–	–	–	–	–	–	–
Tunatic	http://www.wildbits.com/tunatic/	+	–	–	–	–	–	–	–
Shazam	http://www.shazam.com	+	–	–	–	–	–	–	+
Midomi	http://www.midomi.com	–	–	+	–	–	–	–	+
Musipedia	http://www.musipedia.org	–	+	+	+	+	+	+	+
Ritmoteka	http://www.ritmoteka.ru	–	–	–	+	–	–	–	–
Songtapper	http://www.bored.com/songtapper	–	–	–	+	–	–	–	–
Music Ngram Viewer	http://www.peachnote.com	–	–	–	–	+	–	–	–
FolkTuneFinder	http://www.folktunefinder.com	–	–	–	+	+	+	+	+

work, the important issue is not only featuring different input interfaces but transforming one query form to the another depending on searching service availability and it's communication schema.

Different input styles are useful since the user music qualification differs. Melody definition by using a virtual or real keyboard is one of the most exact ways to represent the query, since it accumulates most melody components. However it is not common that users are skilled enough to use the piano keyboard as well as to write adequate note score.

Contrariwise, tapping a rhythm seems to be relatively simple way to define music searching query. The problem is that the number of possible rhythm patterns is evidently less than the number of compositions. It means that even if we succeed to tap the rhythm correctly, we may apparently have a list of thousands titles in return [6].

For the melody contour schema it is possible to choose pitch and time quantization so that the pitch-time representation is equivalent to piano-roll notation [15].

Symbolic pitch notations may be useful if we are limited by only text user controls.

The development of mobile devices with touch screens affects strongly the usage aspects of music searching interfaces. Such devices make possible simulating many kinds of music instruments, although the virtual piano-style keyboard remains the most popular interface. MIDI standards supports transferring of the simulated playing signals between applications regardless the kind of simulated musical instrument. In contrast to the work [16] we don't propose a novel music search engine based on improved music input technologies. Our research is focused on creating middleware application which helps to communicate with existing searchers.

A. Input Styles Transformations

We represent relationships between music query input styles in form of an oriented graph shown in Figure 2. Every transition arc shows the possible transformation from one input style to another, together with indication which note information has to be extracted.

Since the virtual keyboard based query implicitly includes such note attributes as it's duration and it's pitch, there is no much difficulty to transform the keyboard input into the

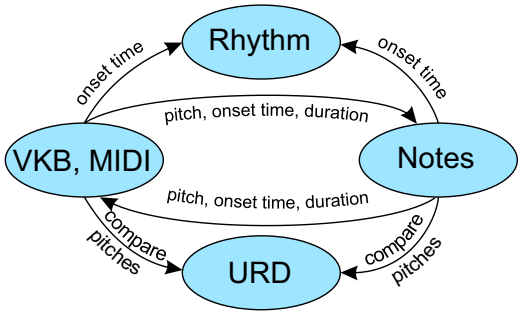


Fig. 2. Graph of music query transformations

rhythm or pitch notation. Next we are able to get the music contour from the sequence of sound pitches.

Clear that in such a way we restrict the user query, and therefore it seems we couldn't expect better searching results. However such transformations may have sense for at least two reasons:

- We attempt to emphasize the meaning of special melody attributes.
- We would like to try to connect a searching service which probably uses quite different music database (e.g. specialized on some music genre[2]) although it supports only restricted input methods (e.g. rhythm or pitch notation).

Regarding to the user interface issues, the ability to move from one input style to another renders possible to switch easily between different searching systems within the framework of one mobile or web application without re-entering the query.

B. Models We Use to Represent Queries

Despite the fact that in our earlier works we argued for the function based representation as one of the best way to describe music, this form seems to be too complicated to be employed directly at a user interface level. Usually user queries are relatively short (and it is true not only for the case of

[2]We turn our attention to the example of such a case in the following section of this paper.

music [17]), so we use sequence of *Note* objects to represent the searching query.

The attributes of a *Note* object are the following:

- *Note name* according to the American pitch notation
- Its *Octave number*
- Its *Onset time*
- Its *End time*

Values necessary for different input styles (such as a note duration or its MIDI value) may be computed on the base of above information. A series of onset time values may be used to generate a rhythm sequence.

IV. Introducing Android Application for Access to Music Searching Services

Nowadays, people are happy to use their mobile devices to access different searching services at any time from any place. They use different types of such devices which may have different input mechanisms like phone keys, *qwerty*-keyboards, touch screens, voice recognition devices, and so on. The variety of devices running on Android operating system is rapidly increasing during last years, so we decided to use Android platform for our music searching application.

For our implementation we selected some music searchers which may be accessed programmatically, particularly: *Musipedia*, *Music Ngram Viewer*, and *FolkTuneFinder*. For three searching systems we implemented four user query input styles:

- Note score editor supporting one voice definition
- Parsons code
- Rhythm tapping
- Piano style virtual keyboard with additional representation of American pitch notation[3]

A. Application Architecture

Despite general application construction ideas are common for various operating platforms, there is obvious specificity of the Android applications: the application architecture and its activities life cycle is governed by the operating system and the Android API. So the solution being discussed in this article isn't platform independent. However let us note that the Android application can serve as a model for implementing flexible human centric interface which is oriented to present-day style of using hardware and software facilities of various mobile devices.

That's why we skip the description of the specific Android implementation details like how to define resources, or how to support different screen resolutions and orientations. We don't discuss component layout control and internal data models either. Being limited by the paper scope, we only describe the principal Android application activities, their interaction and their connection to user interface components.

Figure 3 represents main components of our music searching helper application.

[3]In fact, it means that we support symbolic pitch notation input too.

The main activity *InputStyleSelection* provides the interface for input style choice. According to the selected input style the respective activity (*MelodyContour*, *MusicScore*, *VirtualPiano*, or *RhythmTapper*) opens and provides the corresponding input interface. The user input is stored as a list of *Note* objects used to construct the query as required. Classes *FolkTuneFinder*, *PeachNote* and *Musipedia* transfer the user input to one of supported search engines and filter their outputs.

B. Web Protocol Adapters

With respect to searching services' application interfaces mentioned in section II, the web information exchange protocol adapters have been implemented as Figure 4 illustrates.

The SOAP protocol is not recommended for mobile devices since it uses verbose XML format and may be considerably slower in comparison with other middleware technologies. Unfortunately it is the only way to communicate with the *Musipedia* system. In our case, the mentioned SOAP disadvantages shouldn't case concern since the exchange occurs relatively rarely, only when the respective button is pressed by a user, and there is small amount of information being transferred. We use *org.ksoap2* Java package [18] containing classes required for handling SOAP envelopes and literal XML content. To implement interaction with other searching services (based on the REST architecture and wrapping their responses in JSON format which is typically more compact in comparison with XML) we use *Google Gson* Java library [19]. It allows converting Java objects into their JSON representation as well as backward converting JSON strings to equivalent Java objects.

C. Usage Example

The application starts with a welcome screen for the preferred input style selection (Figure 5).

Then the respective activity starts as shown in Figure 6 representing the example of a virtual keyboard interface. As described earlier (see section III) the user actions are being stored in form of a note sequence with respect to the following related data:

- A *pitch* represented in the American pitch notation (note name and octave number
- The pitch *onset time*
- The pitch *end time*

Other properties may be computed depending on the requirements of a music searcher. Let us illustrate this by the input represented in form of a simplified timing chart (with respect to the note names rather than sound frequencies). The chart in Figure 7 represents some first notes of the well known Russian folk song "Birch Tree".

For the reason that *Musipedia* searcher requires a sequence of triplets containing an onset time, a MIDI pitch and its duration, the user input shown in Figure 7 is converted to the following query data:

0.0, 76, 0.54; 0.66, 76, 0.47; 1.21, 76, 1.43; 1.72, 76, 0.50; 2.41, 74, 0.98; 3.57, 72, 0.27; 3.89, 72, 0.52; 4.62, 71, 0.81; 5.57, 69, 0.75;

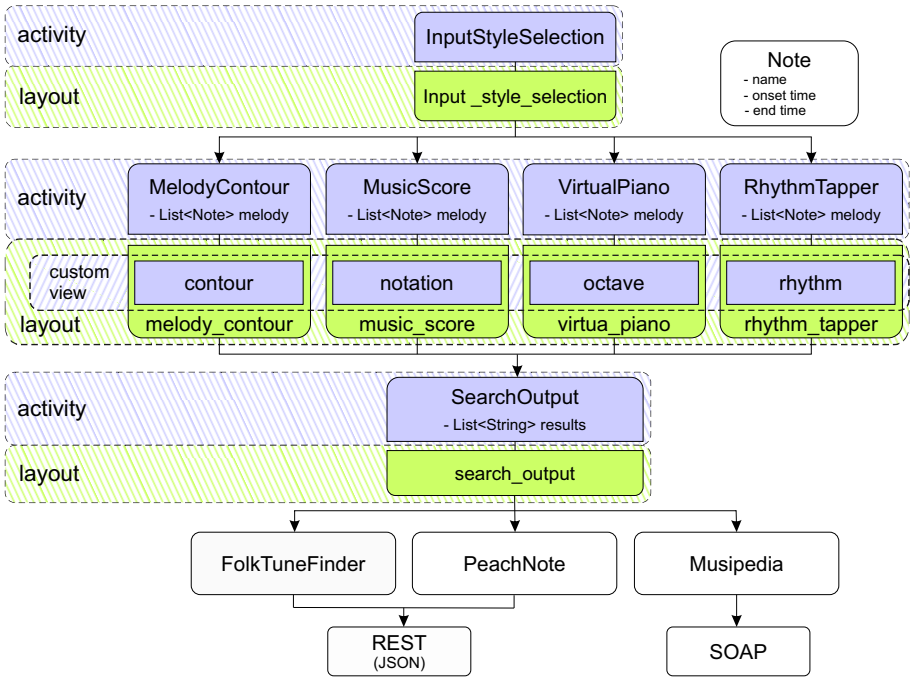


Fig. 3. Android music searching application architecture

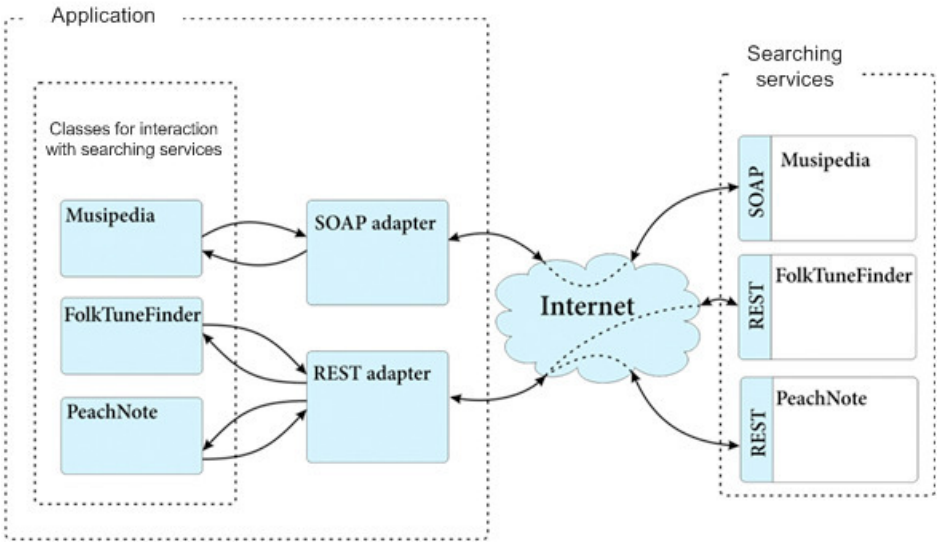


Fig. 4. Web protocol adapters

After this the respective information is included to the SOAP request which is subsequently sent to the *Musipedia* server. As a result, the searching system returns the list of retrieved compositions as shown in Figure 8 [4]. We see the confirmation of the known fact that this melody was used by Piotr Tchaikovsky in the 4th movement of his Symphony No.4 in F-moll (compare with the fragment of the symphony note score shown in Figure 9).

Using other searching engines may enhance searching results by taking into account other music genres. Let us take the *FolkTuneFinder* service which requires a sequence of MIDI pitches. Hence the user input is transformed into the sequence of MIDI pitches as follows:

76, 76, 76, 76, 74, 72, 72, 71, 69

[4]We selected Tchaikovsky's work, but as you can see, the similar theme may also be recognized in some other known compositions.

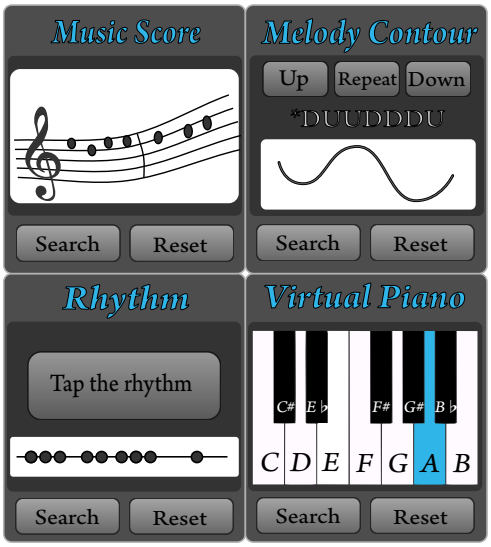


Fig. 5. Main activity: input style selection

Fig. 6. Virtual keyboard input

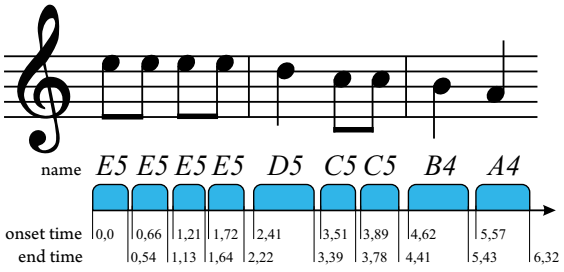


Fig. 7. Test melody: note score representation and timing chart

For the case of the melody contour defined with using Parsons code, the user input is the following *"RRRDDRDD"*.

Fig. 8. Result retrieved by *Musipedia* searcher: *Symphony No. 4*

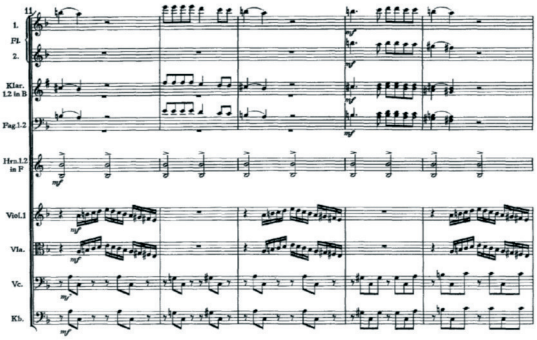


Fig. 9. Birch Tree song cited by Tchaikovsky in his 4th symphony

We implemented the interface component which allows constructing the URD-query by pushing buttons *Up*, *Down* and *Repeat* with synchronous demonstration of the respective graphical contour which is being generated automatically[5]. As you see in Figure 10 the resulting output also contains the "Birch Tree" among other compositions. Note that since the melody contour is a less exact input method (comparing to direct melody definition), it is normal that we don't have the desired melody in the very first lines.

The example we selected for the illustration shows well one important aspect of music searching process, although in a slightly simplified manner. When we discover the composition corresponding to the given request, we may expect obtaining even more information than simply a desired piece of music. Fast every Russian knows the "Birch Tree" song since the early childhood years. But only those who listen to the classical music discover this theme in one motive of Tchaikovsky's symphony. In contrast to this, western music lovers may listen this motive first just in the Tchaikovsky's work, and after a while recognize it as a citation of the Russian folk song. Isn't it a kind of process similar to a music perception in terms of musicology?

[5]We consider to investigate possibility to support a melody contour drawing interface in future implementations.

Fig. 10. Results retrieved by *FolkTuneFinder* searcher

V. Conclusion and Future Work

In the domain of human-centric computing much attention is paid to the facilitating user interface features in relation with a kind of data being processed. As a special type of information retrieval systems, music retrieval systems demand special ways to interact with users. They include not only traditional text or media based queries but specific forms of user input facilities such as note score representations, virtual or MIDI-compatible instruments, as well as composing queries based on melody humming or rhythm tapping which may contain errors of human interpretation. Such approaches may help to overcome limitations of fingerprinting techniques which require exact or nearly exact audio fragments to proceed with searching in the databases of stored music compositions. In our work we investigated styles of user inputs used in various music searching services and applications. We applied transformation rules of query conversion from one input style to another to a software tool communicating with programmatically accessible music searching services from mobile devices running on the Android operating system.

In the current implementation we supported only those music queries which are representable in symbolic form (e.g. note score, pitch notation, note sequences, or contour symbolic description). User interface facilities may be improved if we consider other ways to interact with the user having a touch screen device. It may include, for example, melody contour or rhythm drawing facilities. Even for the searching services that we used currently, there are input styles which are still not incorporated into the existing software prototype. We investigate possibilities to support interfaces for melody singing or humming. Actually we faced the problem to pass the audio query to the searching engines via existing data transfer protocols that we are allowed to use. Ways to extend the interface may also include a support for connected MIDI-compatible devices and text-based searching facilities aimed to explore music metadata information. The another interesting improvement which could fit well especially mobile equipment interfaces is to support music tagging as described for example in [20]. Hence the key idea is to connect different kinds of searching services with rich user input facilities so as to follow better the usage style of modern mobile devices.

Acknowledgment

The authors would like to thank Joe, creator of the *FolkTuneFinder* service, for the opportunity to access his software from our application and for his kind help in organizing software interface with the *FolkTuneFinder*.

References

[1] E. Pyshkin and A. Kuznetsov, "Approaches for web search user interfaces," *Journal of Convergence*, vol. 1, no. 1, 2010.
[2] Z. Mazur and K. Wiklak, "Music information retrieval on the internet," in *Advances in Multimedia and Network Information System Technologies*. Springer, 2010, pp. 229–243.
[3] A. Kuznetsov and E. Pyshkin, "Searching for music: from melodies in mind to the resources on the web," in *Proceedings of the 13th international conference on humans and computers*. University of Aizu Press, 2010, pp. 152–158.
[4] B. Snyder, *Music and Memory: An Introduction*. Cambridge, Mass. [u.a.]: MIT Press, 2000.
[5] D. Deutch, "Music perception," *Frontiers in Bioscience*, 2007.
[6] A. Kuznetsov and E. Pyshkin, "Function-based and circuit-based symbolic music representation, or back to Beethoven," in *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments*. ACM, 2012, pp. 171–177.
[7] A. L. Uitdenbogerd and J. Zobel, "Manipulation of music for melody matching," in *Proceedings of the sixth ACM international conference on Multimedia*, Bristol, United Kngdm, September 1998.
[8] C. Isikhan and G. Ozcan, "A survey of melody extraction techniques for music information retrieval," in *Proceedings of 4th Conference on Interdisciplinary Musicology (SIM'08)*, Thessaloniki, Greece, July 2008.
[9] A. Wang, "The shazam music recognition service," *Communications of the ACM*, vol. 49, no. 8, pp. 44–48, 2006.
[10] W. Hatch, "A quick review of audio fingerprinting," McGill University, Tech. Rep., March 2003. [Online]. Available: http://www.music.mcgill.ca/~wes/docs/finger2.pdf
[11] P. Cano, T. Kalker, E. Batlle, and J. Haitsma, "A review of algorithms for audio fingerprinting," *The Journal of VLSI Signal Processing*, vol. 41, no. 3, pp. 271–284, 2005.
[12] A. Nanopoulos, D. Rafailidis, M. M. Ruxanda, and Y. Manolopoulos, "Music search engines: Specifications and challenges," *Information Processing & Management*, vol. 45, no. 3, pp. 392–396, 2009.
[13] "Musipedia SOAP interface." [Online]. Available: http://www.musipedia.org/soap_interface.html
[14] "Music ngram viewer API." [Online]. Available: http://www.peachnote.com/api.html
[15] D. Shasha and Y. Zhu, *High Performance Discovery in Time Series: Techniques and Case Studies*. Springer Verlag, New York, 2004.
[16] M. Wang, W. Mao, and H.-K. Goh, "Music search engine with virtual musical instruments playing interface," in *Advances in Multimedia Modeling*. Springer, 2013, pp. 502–504.
[17] V. Klyuev and Y. Haralambous, "A query expansion technique using the ewc semantic relatedness measure," *Informatica: An International Journal of Computing and Informatics*, vol. 35, no. 4, pp. 401–406, 2011.
[18] "Package org.ksoap2." [Online]. Available: http://ksoap2.sourceforge.net/doc/api/org/ksoap2/package-summary.html
[19] I. Singh, J. Leitch, and J. Wilson, "Gson user guide." [Online]. Available: https://sites.google.com/site/gson/gson-user-guide
[20] K. Bischoff, C. S. Firan, and R. Paiu, "Deriving music theme annotations from user tags," in *Proceedings of the WWW 2009*, 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 275–278

Evaluation of beef production and consumption ontology and presentation of its actual and potential applications

Rafał Trójczak, Robert Trypuz and
Przemysław Grądzki
The John Paul II Catholic University of Lublin
Faculty of Philosophy
Al. Racławickie 14, Lublin, Poland
Email: trypuz@kul.pl

Jerzy Wierzbicki and Alicja Woźniak
Polish Beef Association
Ul. Kruczkowskiego 3, 00-380 Warszawa
Email: jerzy.wierzbicki@pzpbm.pl

***Abstract*—The paper concerns beef production and consumption ontology (*OntoBeef*) and its applications. It is presented the three-stage *OntoBeef* evaluation process with a special focus on description of interaction of ontologists with domain experts. We also describe Linked Open Data (LOD) philosophy and show how links between *OntoBeef* and four other ontologies were established. We also present the components of *OntoBeef*-driven information system, a technology used to its creation and its functionalities. In particular we describe thesaurus component of the information system incorporating LOD connections.**

I. Introduction

THIS paper is a continuation of the one presented a year ago during WEO-DIA (FedCSIS) 2012 workshop (see [1]). In WEO-DIA 2012 paper we have described motivations to create a beef production and consumption ontology—which we call *OntoBeef* in short—and a project ProOptiBeef[1] within our research was carried out. In particular the paper contained information about the methodology of building the ontology, its content and possible applications. We have presented there ontological choices made while building the ontology and their justification. We have also shown how *OntoBeef* is used for browsing a database of articles (being indexed by the ontology concepts). In this paper we shall describe the way in which *OntoBeef* was further validated by the domain experts. We shall also present the initial stage of *OntoBeef*-driven application built by us—in particular we shall focus on its technological and functional sides.

The structure of the paper is the following. In Section II we present the three-stage *OntoBeef* evaluation process. In Section III Linked Open Data philosophy is introduced. In that section we show how links between *OntoBeef* concepts and the concepts of four other ontologies were established. Finally, in Section IV we present the components of a built by us ontology-driven information system, a technology used to its creation and its functionalities.

[1]http://www.prooptibeef.pl/

II. *OntoBeef* evaluation

OntoBeef ontology has been evaluated by seven experts: Prof. Krystyna Gutkowska (Institute of Rural and Agricultural Development PAS), Prof. Zenon Nogalski and MSc eng. Maciej Borzyszkowski (University of Warmia and Mazury), Prof. Agnieszka Wierzbicka, Dr eng. Marcin Gołębiewski, Dr eng. Eliza Kostyra and MSc eng. Rita Rakowska (Warsaw University of Life Sciences). The experts has been invited to three-stage evaluation process. In the first stage of the evaluation the experts have been choosing the concepts to be validated, in the second stage of it they have been evaluating labels assigned to each concept and finally they have been assessing the ontological relations between concepts. In what follows, we shall describe in details each of the three stages.

A. Concepts choosing

At the first stage the experts have been asked to choose among all 2344 concepts these which belong to their domains of interest. In figure 1 we can see the screen of the application used to support the process.

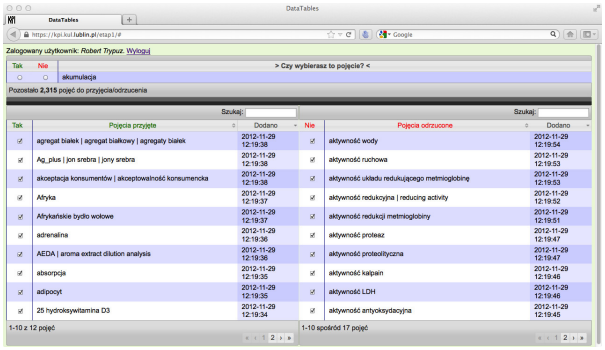


Fig. 1. First stage of *OntoBeef* evaluation

Concepts were displayed one by one. Each of them was represented by the sequence of synonymous names. An expert could choose between "Yes" (I do accept the concept and

978-83-60810-53-8/$25.00

want to take care of it later on) and “No” (I don’t accept the concept). Accepted concepts have been gathered in the table on the left and rejected concepts in the table on the right (see figure 1). An expert could always change his mind by unmarking the checkbox close to the chosen or rejected concept. It is worth noting that one and the same concept could be chosen by two or more experts (in fact there were many concepts validated by more than one expert; e.g. two experts shared even 675 concepts).

B. Labels evaluation

When all the concepts have been distributed to the experts, the second stage of the evaluations process began. During the stage the experts were asked to make an order in the labels assigned to each concept. In figure 2 we can see the screen of the application used to support the process.

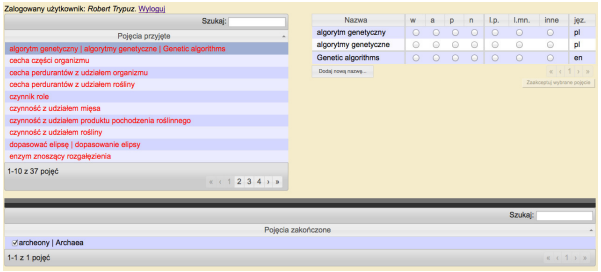


Fig. 2. Second stage of *OntoBeef* evaluation

Experts assessed each label by assigning to it one of the four label properties: “proper” (w), “adequate” (a), “common” (p) and “wrong/hidden” (n). They identify also the grammatical category of labels (“singular” or “plural”) and validated their language. Experts might also add a new synonymous label to the concept in any language. Aforementioned label properties: `proper_label`, `adequate_label`, `common_label` and `wrong_label` are instances of `owl:AnnotationProperty` and stay in relation `rdfs:subPropertyOf` to `rdfs:label`. Additionally proper label is a sub-property of adequate label. It is assumed the each class has to have exactly one proper label per language[2]. Since many concepts have been chosen by two or more experts during the first stage, we could expect that the second stage of the evaluation will lead to the emergence of conflicts. And in fact after the work of the experts has been done we found that 654 concepts meet a conflict of labels, e.g. that the same label for the concept had different characteristics (from the set of properties: proper, adequate, common and wrong) or more than one label has been recognized by experts as “proper”. Most of the conflicts were solved by using the following criteria:

1) in the case of labels in the singular and the plural forms determined as “proper”, the singular remained “proper” whereas the plural one(s) became “adequate”;
2) in case of the full name and its acronym were determined as “proper”, the full name was selected as “proper” and it acronym as “adequate”;
3) in the conflict between more than two experts, the choice proposed by majority was accepted;
4) in case the conflict could not be solved otherwise, the label which is more common in the Google search resources was chosen as “proper”;
5) indication of experts, which is not consistent with the original meaning of the term definition has been ignored.

[2]Thus `proper_label` is much like `skos:prefLabel`. Similarly `adequate_label` and `common_label` correspond to `skos:altLabel` and `wrong_label` to `skos:hiddenLabel`.

C. Evaluation of the ontological relations

The final stage of the evaluation process was analysis of the ontological relations in *OntoBeef*. In figure 3 one can see the screen of the application of this stage.

Fig. 3. The final stage of *OntoBeef* evaluation

An expert had an access to all the previously chosen concepts of the ontology and to the labels proposed by all experts during the second stage. For each concept the following information has been provided: its labels, ancestors, children, siblings and some ontological connections (e.g. “participation” or “parthood”) with other concepts. An expert could submit a comment or suggestion on concept’s labels, ancestors, etc, by clicking on “Zgłoś sugestię” button (see figure 3) – after clicking it a new window with a space for typing comment appears. All visited concepts have been colored yellow and unvisited yet—red. The result of this evaluation stage was 494 received submissions, which have been then analyzed and applied by ontologists.

D. Lesson learned

The evaluation process of *OntoBeef* by domain experts was only partially successful. During the second stage of the process (see section II-B) experts added many correct labels to the concepts, what helped us then to establish connections between *OntoBeef* concepts and LOD ontologies (see section III). But it is also true that many added labels were simply wrong, mostly because misunderstanding of the real concepts references (e.g. to the concept possessing a label “child” some expert added “calf”). Most of the remarks submitted during the last evaluation stage (344 out of 492) concerned labels.

They were constructive and improved *OntoBeef* quality. The rest of them (i.e., concerning ancestors, children, siblings and some ontological connections) were rather missing the point. One of the reasons of this state of affairs is that domain experts supporting ontologist in the project were not trained in ontological thinking and did not feel competent enough to suggest changes in the *OntoBeef* structure. Our earlier experience shows that much better results gives direct face-to-face cooperation of ontologists with domain experts. But this way of processing engages more people, is more time-consuming and as such is obviously more expensive.

III. *OntoBeef* AND LINKED OPEN DATA

After *OntoBeef* was finally validated, its concepts have been linked with other (lightweight) open ontologies. At least two meanings of "Linked Data" are known. In the first one the phrase means a method of knowledge creation and sharing[3]. In the second meaning "Linked Data" refers to "collection of interrelated datasets on the Web"[4]. Of course both definitions are compatible; Linked Data as a collection of interrelated datasets is brought about by many working agents acting according to Linked Data as a method. Linked Open Data is Linked Data which is released under an open license. (Semantic) Web visionary Tim Berners-Lee provided the following set of requirements which a data should possess to be called Linked Data[5]: 1) to be available on the web; 2)to be available as machine-readable structured data; 3) to be coded in some of open standards from W3C (e.g. RDF) to identify things; 4) to be linked to other people's data. It is also strongly suggested to register data at some open data catalogue (e.g. The Data Hub[6]), what in practice leads to a few more technical requirements (e.g. that HTTP URI of a piece of data should be dereferenceable – see [2]). *OntoBeef* has been linked with four thesauri: AGROVOC, General Multilingual Environmental Thesaurus (GEMET), National Agricultural Library's Agricultural Thesaurus (NAL), and STW Thesaurus for Economics (STW). Interlinking process has been done in two steps. In the first step for each thesaurus we have created database with two column table "concept number – label" . The same representation has been created for *OntoBeef*. Then by SQL query we have selected the concepts from thesauri and *OntoBeef* which have the same labels in common. In the second step ontology experts have validated the quality of the automatic connection of concepts and removed the wrong connections where needed. Finally *OntoBeef* has 797 links with AGROVOC, 211 links with GEMET, 546 links with NAL and 119 links with STW.

IV. APPLICATIONS

OntoBeef and its connections to other ontologies are a good starting point for building an ontology-driven information system (IS, in short). An ontology-driven IS is IS in which "ontology profitably "drives" all aspects and all components" of it [3, section 3].

In figure 4 we find four components of IS currently being developed by us within ProOptiBeef project. Semantic Oxpecker was described in the proceedings of FedCSIS 2012 (see [1, section V]). Components: "theses representation and search" and "interface to the database of results of experiments" are under development. Thesaurus component has been already created and will be described in section IV-B.

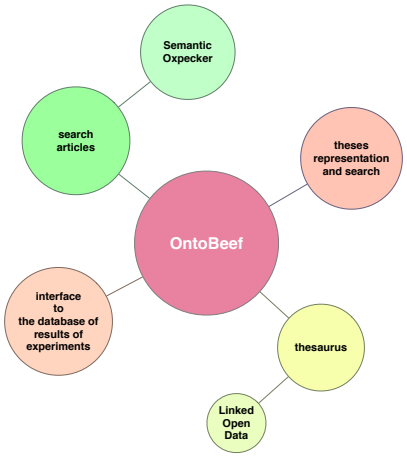


Fig. 4. Four components of information system based on *OntoBeef*

Before we started developing our IS, we have formulated a number of requirements a technology used to create the application should satisfy. The four most important of these conditions are: a) to run as the web application; b) to support OWL 2 (in which *OntoBeef* is formulated); c) to guarantee smooth application performance; d) to be flexible enough to accommodate new functionalities.

Based on our experience we chose three options initially: 1) JavaScript with jOWL framework; 2) Java Enterprise Edition ("Java EE" or "JEE", in short) with Jena framework; 3) Java EE with OWL-API framework [4]. All these technologies satisfy the first condition. Jena framework does not support OWL 2. jOWL framework does not satisfy the third condition, because it requires downloading ontology each time user's computer reloads application page. This takes time and distracts smooth application performance. Finally only OWL API framework meets all the requirements.

A. *OWL API*

In [4] we read that OWL-API is "a high level Application Programming Interface (API) that supports the creation and manipulation of OWL Ontologies". Its first version has been released in 2003. The last version 3.4.3 (which we are currently using) has been released in 2013. OWL-API is open source project managed by people from University of Manchester, written in Java programming language. It is worth noting that OWL-API was used for the development of components of a widely used ontology editor Protégé.

[3]http://aims.fao.org/standards/agrovoc/linked-open-data
[4]http://www.w3.org/standards/semanticweb/data
[5]http://www.w3.org/DesignIssues/LinkedData.html
[6]http://datahub.io

OWL-API allows to use a variety of notation: RDF/XML, OWL/XML, Turtle, Manchester and others. It supports the use of reasoners. It also "includes validators for the OWL 2 profiles – OWL 2 QL, OWL 2 EL and OWL 2 RL" [4].

B. Thesaurus component

Thesaurus component was developed within Java EE plus OWL-API framework. We shall now present how the web part of thesaurus component is running. Java EE consists of a large number of elements. In our application we use only a few selected, namely: servlets technology, JavaServer Pages (JSP) and JDBC technology. The first one handles the low-level web operations such as handling request and response, reading and writing HTTP headers. JSP allows to create the HTML pages in Java with great ease (in comparison with servlet technology). Access to the database is implemented through JDBC technology. It enables to abstract away from particular database technology (in our project MySQL RDBMS was adopted).

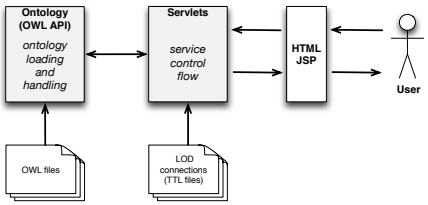


Fig. 5. The control flow diagram for thesaurus component of application

In figure 5 we can see the control flow diagram for the thesaurus component. When a user invokes the web application by writing URL address or by clicking any link on thesaurus component's page, a user's browser is sending HTTP request to a server where application is running. The server (in our case Apache Tomcat web container) receives request and runs appropriate servlet. If necessary the servlet retrieves information about the concept from OWL file by OWL-API. When the servlet obtains data about the concept it also checks Linked Open Data connections stored in TTL files. After obtaining all the necessary information, HTML page is sent to the end user. From the end user perspective the application looks as presented in figure 6. For each concept – in this case "beef"[7]– in ontology the component displays its labels: common and adequate (among them the proper ones indicated by the green color), the list of domain experts who validated the concept, the ancestors, the children, the sibling concepts and some other ontological relations as for instance parthood and participation. An end user can search for a concept. It is worth noting that a concept can be found also by typing a wrong label which are assigned to the concept (however they are invisible for the user). Application enables also registration and after log in allows reporting suggestions and comments considering concept labels, ancestors, children, siblings and other ontological properties. In the top application bar (see figure 6) there is LOD part, which displays LOD connections. In figure 6 we can see that beef class is linked with three thesauri. By linking *OntoBeef* with other resources we get for instance translations of labels to 22 languages, definitions and some related (RT), border (BT) and narrower (NT) terms to the searched one.

[7]See: http://onto.beef.org.pl/domain/concept/201

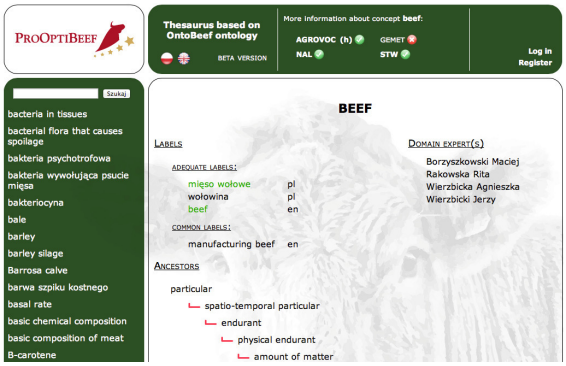


Fig. 6. *OntoBeef* thesaurus component

V. Conclusion and Perspectives

In this paper we described how *OntoBeef* ontology was validated by the domain experts. We presented the thesaurus component of *OntoBeef*-driven application. We also described the technological and functional aspects of our application. Finally the ongoing work was also described. We are very happy to notice that the researchers and practitioners working in the domain of beef production and consumption in Poland recognize the impact an ontology (as an artifact and as a methodology) had on the way they think about their domain and on the quality of their communication. There is still a lot of ontological work to be done in the field. For instance the beef sector has a lot of local carcass cuts systems which are in part mutually incompatible. We believe that their ontological implementation in *OntoBeef* will be the first step towards their comparison and integration

Acknowledgment

Research was realized within the Project no. WND-POIG.01.03.01-00-204/09 Optimizing of Beef Production in Poland According to "from Fork to Farm" Strategy co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme 2007 – 2013.

References

[1] P. Kulicki, R. Trypuz, and J. Wierzbicki, "Towards beef production and consumption ontology and its application," in *Proceedings of the Federated Conference on Computer Science and Information Systems*, 2012, pp. 483—488.

[2] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, ser. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, 2011.

[3] N. Guarino, "Formal ontology in information systems," in *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June*, I. P. Amsterdam, Ed., 1998, pp. 3–15.

[4] M. Horridge and S. Bechhofer, "The OWL API: A Java API for OWL Ontologies," *Semantic Web Journal*, no. 2(1), pp. 11–21, 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 279–286

Query Construction for Related Document Search Based on User Annotations

Jakub Ševcech, Mária Bieliková
Faculty of Informatics and Information Technologies,
Slovak University of Technology,
Ilkovičova, 842 16 Bratislava, Slovakia
Email: {name.surname}@stuba.sk

***Abstract*—We often use various services for creating bookmarks, tags, highlights and other types of annotations while surfing the Web or just reading electronic documents. These annotations represent additional information on particular information source. We proposed a method for query construction to search for related documents to currently studied document. We use the document content where we concentrate on user created annotations as indicators of user's interest in particular parts of the document. Our method for query construction is based on spreading activation in a graph created from the document content. We evaluated proposed method within a service called Annota, which allows users to insert various types of annotations into web pages and PDF documents displayed in the web browser. We analyzed properties of various types of annotations inserted by users of Annota into documents. Based on these properties, we also performed a simulation to determine optimal parameters and compare proposed method against commonly used tf-idf based method.**

I. Introduction

WE OFTEN use various services for creating bookmarks, tags, highlights and other types of annotations while surfing the Web or when reading electronic documents. We use these annotations as means to store our thoughts or to organize personal collections of documents using methods such as tag-cloud. Many services supporting document bookmarking and manual annotation of documents provides us the possibility to create various types of annotation by simulating the process of annotation creation in printed documents. These services do not provide us with new types of annotations in addition to annotations we have been already creating in printed documents. They rather provide us new possibilities for annotation utilization. There is active research in the field of utilization of annotation [1], for example in support of navigation between documents. In the work presented in [2] the authors use the term social document to represent document enhanced by the user generated content such as annotations. They used the user generated content similarly to anchor texts while indexing documents. This representation of documents proved to provide improved performance in content-based mining applications on the Web such as search engines, recommendation systems etc.

User created annotations can be considered a form of user's context he creates while reading documents and traveling in digital space [3]. Great many applications use annotations as means for navigation between documents and for organizing content. For example, in [4] the authors describe an organization of learning materials and collaboration of students while learning using an educational system that provides students the possibility to attach various types of annotations to learning objects. The study of various search tasks supported by a social bookmarking service deployed in a large enterprise is presented in [5]. The authors concluded that bookmarking services and annotations attached to documents can enhance document organization and social navigation.

User generated tags are one of the most commonly used methods for organizing content. Tags are used for organizing bookmarks in services such as Diigo[1] or Delicious[2], but they are also used to organize notes[3], in various blogs and many other applications. Tags and other types of annotations are means document visitors can use to create custom navigation. They can categorize or describe resources and by this way create navigation that fits their needs without relying on navigation provided by document author.

User created annotations can be used not only to support navigation, but there are many other possible applications. Tags are used for folksonomy construction [6], annotations can play a great role for example in content enrichment and content quality improvement such as in an education system presented in [4]. In this system the authors use content error reports, user generated comments and questions, to improve course content and other types of annotations such as tags and highlights for the navigation and even the content summarization [7].

Currently, there are many services allowing users to annotate the documents. Annotations are used to support the navigation in users' collection of documents, they allow users to create their own organization of documents via tags and they help in search for documents. All of these applications motivate users to create annotations by a prospect of future improvement in inter or intra document navigation. Users benefit created annotations only after there is enough annotated documents, or when returning to once annotated document.

[1] Diigo, http://www.diigo.com/
[2] Delicious, https://delicious.com/
[3] Evernote, https://www.evernote.com/

978-83-60810-53-8/$25.00

Problem with this approach is that there is lack of immediate reward for the annotation creation.

In this paper we propose a method for query construction from currently studied document and attached annotations. This method produces a query that can be used in related document retrieval where the query is taking into account user's interest provided by created annotations. The query is created in time the user is reading the documents and it is used to search for further documents related to the currently studied document. The reward for user creating annotations is thus provided in time of annotation creation.

II. Related Work

One of possible employment of annotations in information processing is the document search. There are two possible approaches for exploitation of annotations in search. One is to use annotations while indexing documents by expanding documents in a similar way anchor texts are used [2] or by ranking document quality using bookmarks and annotations as document quality indicators [8].

The second possible application of annotations in document search is in query expansion or query construction process. An example of annotations used for query expansion is presented in [9], where tags attached to search results are used to expand initial query similarly to pseudo-relevance feedback based query expansion. Multiple methods for query expansion in folksonomies are presented in [10]. Of particular interest are methods expanding queries by tags from folksonomies on the basis of semantic similarity between words of the query and these tags.

An example of annotations used as queries to retrieve related documents is presented in [11]. The authors asked users to read a set of documents and to create annotations into documents using a tablet. They used these annotations as queries in related document search. They used different weights for different types of annotations in query construction and they compared search precision of these queries with relevance feedback expanded queries. Queries derived from user's annotations produced significantly better results than relevance feedback queries. Whereas query expansion requires that users create an initial query, query composition using annotations does not require additional activity of readers instead it reuses annotations created with other purposes such as better understanding of the document.

In experiment presented in [11] authors let users to create annotations into documents for evaluation of their applicability in related document search. More often in search for related documents the content of source document is used to create queries. In [12] authors used the most important phrases from the source document as queries for document retrieval. Extracting most important phrases is similar to document summarization. However they used extracted phrases as queries in related documents retrieval.

Another work concerning search for related documents is described in [13]. The authors use related document search as a mean for recommendation of citations into unpublished manuscripts. They use text-based features of the document to retrieve similar documents and citation features to establish authority of documents.

Similar document retrieval has its application in document recommendation. In work presented in [14] they used list of documents similar to users visited documents to recommend related documents. To compute document similarity they used document representation based on word vector extracted from its content and similarity metric based on cosine similarity.

Searching for related documents can be useful also in the domain of plagiarism detection. In [15] query construction from the source document is used for retrieval of documents the suspicious document may be plagiarized from. In the query construction process the most frequent words from the document are used.

Document term frequency for query construction from document content is used also in popular content-based search engines ElasticSearch[4] and Apache Solr[5]. They provide special type of query interface called "more like this" query, which processes source text and returns list of similar documents. Internally, the search engine extracts the most important words using tf-idf metric from source text and it uses the most important words as a query for related documents search. By comparison to previous described method, tf-idf based method uses additionally to in-document term frequency also information about terms from the collection related documents are searched in.

In multiple works authors showed that annotations represent important source of information for document retrieval. Methods for query construction for document retrieval are however using only document content and information about document collection in query construction process. They are not using user created annotations as user's interest indicators when creating query for document retrieval. In our work we proposed and evaluated a method for query construction from the document content enhanced by user created annotations. Annotations are used as interest indicators to determine parts of the document user is mostly interested in. Using user created annotations our method creates a keyword query for related document search taking into account the user interests. Annotations are used in time of their creation and they provide immediate motivation in form of related document search.

III. Method for Query Construction

Currently the most common form of query used when searching for documents on the Web is the list of keywords. To retrieve words from the document to be used as query for related document search it is possible to use multiple different approaches. It is possible to extract most frequent terms, use tf-idf metric or various ATR algorithms [16] to extract keywords and so on. The tf-idf based method provides rather straightforward possibility to incorporate user created annotations: the source text of the document is extended by the content of created annotations possibly with various weights for different types of annotations.

[4] ElasticSearch, http://www.elasticsearch.org/
[5] Apache Solr, http://lucene.apache.org/solr/

However, the method using the tf-idf for query word extraction takes into account only the number of occurrences of words in the source document and in the document collection. We believe that not only the number of word occurrences but also the structure of the source text is important in a search query construction for related document retrieval. Especially, if we suppose that while reading the document users are most commonly interested in only a portion of the document, the portion where they attach annotations.

We use user created annotations to increase weights of annotated parts of the document in query construction process and to attach additional content to the document. We proposed a method based on spreading activation in text of studied document transformed to a graph. The method uses annotations as interest indicators to extract parts of documents the user is most interested in.

The proposed method is composed of two phases:

1. Text to graph transformation that conserves word occurrence frequency in node degree and text structure in graph edges structure.
2. Graph nodes activation introduced by annotations attached to the document and query word extraction using spreading activation algorithm in created graph.

The text to graph transformation is based on word neighborhood in the text. The graph created from text using words neighborhood conserves words importance in node degree but it also reflects the structure of the source text in the structure of edges [17]. Using various graph algorithms such as community detection, various node and edge weightings or spreading activation we can extract properties such as related words, most important terms, topics etc. We use this graph to extract words that can form queries to retrieve similar documents using spreading activation algorithm.

Text to graph transformation

To transform document text to a graph, it is firstly preprocessed in several steps: segmentation, tokenization, stop-words removal and stemming. After these steps the initial text is transformed into list of words. Every unique word from this list is transformed into single node of the graph. The edges of the graph are created between two nodes if corresponding words in the text are neighbors or they are in the predefined maximal distance. The text to graph transformation is described by the following pseudocode:

```
words=text.split.removeStopwords.stem
length=words.size
nodes=words.uniq
edges=[]
for(i=0;i<length;i++){
  for(j=i;i<min(i+dist,length-1);j++){
    edges.add(words[i],words[j])
  }    }
graph=Graph.new(nodes,edges)
```

As settings for maximal distance between words we used options described in [17], where they used two passages through the text with maximal distance set to two words and five words. By using these setting, the words with greater distance were connected and close words have more common edges at the same time.

All created edges have the same weight but by using two passages through the text, more edges are created between close words than between farther words. For the purpose of speeding up the spreading activation in the next step, we connected multiple edges between the same nodes and we set weight of the resulting edge as number of connected edges.

Query word extraction

In the text transformed to the graph we use spreading activation algorithm to find the most important nodes/words. This algorithm is commonly used for example to find most related nodes in the graph to the initially activated node. The activation introduced into the initial node is spreading through the edges and after the change in nodes activation is smaller than specified threshold, the most related nodes have the greatest amount of activation concentrated.

It is possible to use this algorithm for related nodes search but also for other application such as keyword extraction [18]. We use this algorithm to find the most important words in the graph created from the text. The initial activation is introduced to nodes, annotations are attached to. The initial activation is propagating through the graph and it is concentrating in most important words of the text. When user created annotations are used to insert initial activation, user's interest are reflected in the most important words extracted after spreading activation.

When using annotations to insert initial activation into the document graph we consider separately annotations that:

- highlighting parts of the document and
- inserting additional content into the document.

The proposed method takes into account both types. Those, which highlight parts of the document, contribute by activation to nodes representing words of highlighted part of the document. Annotations enriching content of the document are extending the document graph by adding new nodes and edges and they are inserting activation to this extended part of the graph. When inserting activation to extended parts of the document we assume that some portion of the words used in the annotation content are located in the document text as well. The activation from the extended part of the graph can then pass to the rest of the graph through common nodes. This assumption may be violated in the case where the document and the associated comments are in different languages. Therefore, in performed experiments we translated the content of every annotation using Google Translate service.

When initial activation is spreading through the created graph, the nodes where activation is concentrating are the most important words of the graph and are considered words fit into the query. In our case the activation is inserted into the graph through annotations attached to document by its reader. As we use annotations as user's interest indicators, the activation is spreading from document parts, user is most

interested in and words with highest activation level are reflecting user's interests.

The proposed method is able to extract words, which are important for annotated part of the document, but it is also able to extract globally important words, that are important for document as a whole. The portion of locally and globally important words can be controlled by number of iteration of the algorithm. With increasing number of iterations the activation is spreading from activated part of the document and extracted locally important words are changed to globally important words. When using this method it is thus important to determine when to stop the algorithm to find the best portion of globally and locally important words. It is also important to determine the right amount of activation inserted into the graph by various types of annotations.

The method for query word extraction uses annotations to insert initial activation into text transformed to graph. In case when no annotations are attached to the document, it is possible to extract globally important words from the document by activating whole document's text.

IV. Creation of the Web Page Annotation

The key element in document annotation is the selection of a method to link documents and created annotations. Multiple systems supporting annotation creation assume that documents will not change after annotations are inserted. This is very strong assumption we cannot make in a domain such as web pages. We have to use method for annotation interlinking with document content with regard to documents which may change over time. In [19] multiple criteria, which must meet the robust method for locating annotations into documents, are defined. Some of the criteria are:

- The method has to be robust to common changes in the referenced document.
- Has to be based on document content.
- Has to work with uncooperative servers.
- The information necessary to locate annotation have to be relatively small compared to the document content.

At the same time in this work they suggest several approaches that meet these criteria. One of them is to use annotation context in form of surrounding text to place the annotation into the document. The method using document content to place annotations is defined also in Open Annotation Model [20]. It is tolerant to changes in the document content and when using approximate matching of strings it as also to some extent tolerant to changes in annotation context as well.

We developed a service called Annota[6] [21], which allows users to attach annotations to arbitrary web pages or PDF documents displayed in a web browser. Annota service is realized as a browser extension through which user can create various types of annotations such as:

- tags,
- highlights,
- comments attached to text selections and
- notes attached to the document as a whole.

The service is focused on supporting visitors of digital libraries as we collect metadata on articles from selected digital libraries (ACM DL, SpringerLink, IEEE Xplore). We realized the possibility to insert annotations into arbitrary web pages and articles in digital libraries, by bookmarking and sharing documents and annotations in user groups.

The Annota service allows users to organize documents by tags or folders. It is possible to search in document's texts in user's library or library of bookmarked documents of all users. An example of web page annotated using Annota is displayed at Figure 1. The figure shows a sidebar, where it is possible to bookmark displayed page, insert tags, edit note and share bookmark with groups user is member of. Users are able to highlight text fragments of the web page and to attach comments to these text selections.

The basic scenario of the service usage follows user studying a document. The user has following possibilities of particular activities:

- Bookmarking documents.
- Highlighting parts of the text and creating other types of annotations.
- Sharing bookmarked document via group sharing.
- Collaborative annotation of documents.

The browser extension allows users to create annotations that link to document as a whole (tags, note) or to particular parts of the document (highlight). As the extension is inserting annotations into web pages and they change frequently and without notification, we had to use a method for annotation linking to specified parts of the document that is robust to changes in annotated document.

To attach annotations to document parts we use redundant representation of annotation location to support linking annotations into changing documents. To locate annotation in the text, we store highlighted text with order of its in-text occurrence together with surrounding text. The combination of selected text and text occurrence order is tolerant to changes in the document content except changes in selected text and most changes before annotation location. With usage of approximate matching this method is to some extent tolerant to changes in selected text as well.

We analyzed behavior of users of Annota while annotating documents using developed browser extension. Our experiments are based on usage data of 82 users who created 1 416 bookmarks and 399 in-text annotations during 4 months time period. They used Annota on day-to-day basis to bookmark interesting documents, to summarize them, to write down their thoughts about the document content and to highlight important parts of the document. We studied multiple parameters of created annotations and notes and we derived probabilistic distributions of these parameters. We studied properties such as the note length, number of highlights per user and per document, highlighted text length or probability of comment to be attached to highlighted text. All observed parameters were following logarithmic or geo-

[6] Annota, http://annota.fiit.stuba.sk/

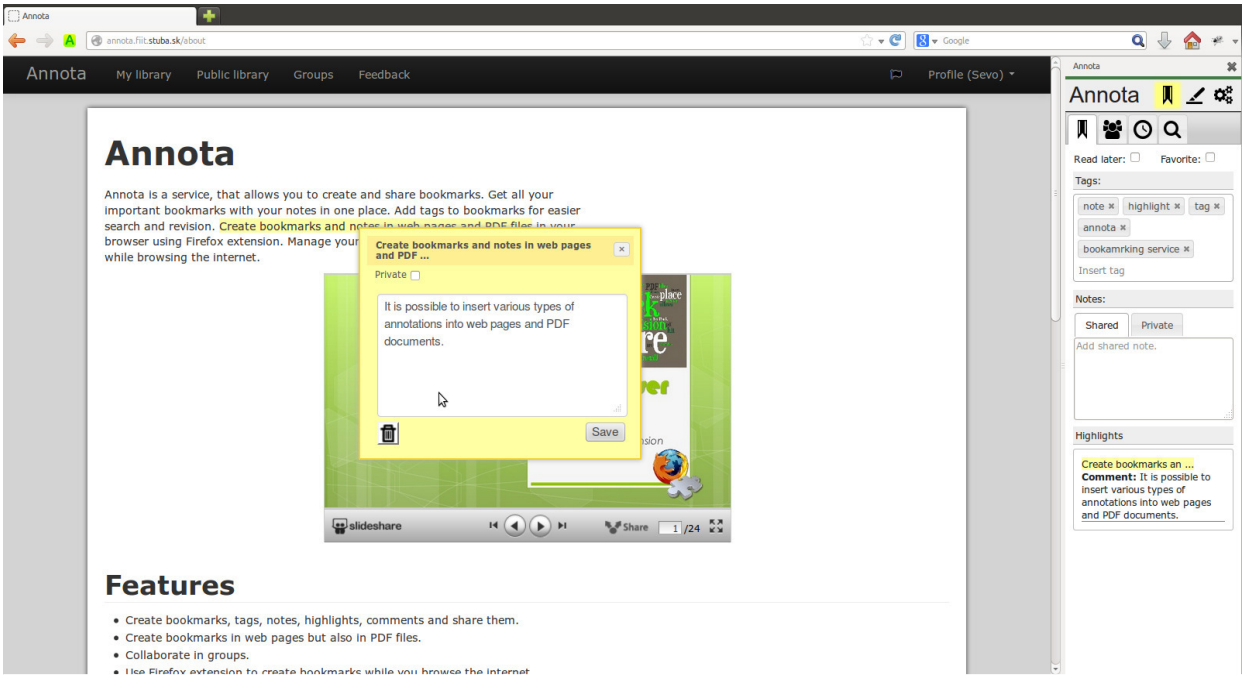


Figure 1 Web page annotated using bookmarking service Annota

metric distributions. Figure 2 represents an example of derived distribution for number of highlighted texts per document that follows logarithmic distribution.

V. Evaluation

Using various attributes of annotations and their probabilistic distributions described in previous section, we created a simulation, to find optimal weights for various types of annotations and number of iterations of proposed method for query construction from document text and attached annotations. We optimized query construction for document search precision.

The simulation was performed on the dataset we created by extracting documents from Wikipedia. We constructed

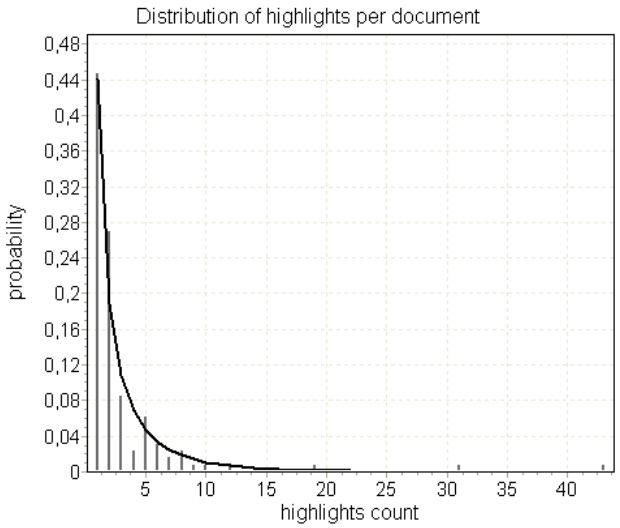


Figure 2 Logarithmic distribution of highlighted texts number per document

the source documents with aim to create documents containing several similar sections (from the point of view of used words) and with different topics. These generated documents simulate documents, where the user is interested in only a fraction of the content. To create such documents we used disambiguation pages in Wikipedia. The disambiguation page disambiguates multiple meanings of the same word and contains links to pages describing each of these meanings.

By using abstracts of pages describing different meanings of synonyms we simulate sections of the text describing multiple topics. We downloaded all disambiguation pages and we selected random subset of these pages for which we downloaded pages they are linking to. Along with these disambiguated documents we downloaded all documents, having common category with at least one of disambiguated documents.

We used search engine ElasticSearch to create an index of all downloaded documents and to search within this index. The parameters of created dataset are summarized in Table 1.

Table 1
Parameters of dataset used in simulation

Attribute	Number
All disambiguation pages	226 363
Selected disambiguation pages	86
Pages disambiguation pages are linking to	629
Categories	2 654
All downloaded pages	232 642

In the simulation we generated annotations in a way to correspond with probabilistic distributions extracted from the annotations created by users of the Annota service. From every disambiguation page and pages it was linking to, we

created one source document by combining abstracts of all pages in random order. For every source document we selected one of composing abstracts which simulated one topic user is most interested in. Into the selected abstract we generated various types of annotations, both annotations highlighting parts of the document and annotations inserting additional content. Annotations highlighting parts of the document were randomly distributed over the whole abstract. To simulate content of annotations extending content of the document (note, comments) we used random parts of the page annotated abstract was extracted from.

Generated annotations along with source document content were used to create query using proposed method based on text to graph transformation and spreading activation. Created query was used for related documents search in the index of all downloaded documents. When evaluating precision of search for related documents, we considered document to be relevant if it was from the same category as the page of annotated abstract.

We performed a simulation with several combinations of parameters and we implemented hill climbing algorithm to optimize parameter combination for the highest precision. Single iteration of performed simulation is described by following pseudocode:

```
for disambig in disambiguations do
  abstracts = disambig.pages.abstracts
  for abstract in abstracts do
    text = abstracts.shuffle.join(" ")
    graph = Graph.new(text)
    annot = Annotation.generate(abstract)
    graph.activate(annot, weights)
    graph.spread_activation
    query = graph.top_nodes
    results = ElasticSearch(query)
    cat = abstract.page.categories
    relevant = results.with_category(cat)
  end
end
```

We compared search precision for proposed method and for tf-idf based method ("more like this" query) provided by ElasticSearch when searching for 10 most relevant documents. For the purpose of comparison of proposed method with method based on tf-idf when using annotations in the query construction process, we extended rather straightforwardly the tf-idf based method to use annotations in query word extraction process. This method uses document word frequency to find most important words in the text. We extended the text of the document by text annotations were attached to and annotations content. We provided different weights for different annotations types by repeated extension of document by highlighted text and annotations content. We determined the optimal number of repetitions using parameter optimization with hill climbing algorithm.

Along with simulation using generated annotations for methods comparison, we performed two experiments to determine retrieval precision with no annotations and when whole abstract of the source document was highlighted. These experiments aimed to determine precision of compared methods when no annotations are available and when we have complete information about user's interest.

Results for simulations with generated annotations along with experiments with no annotations and with whole document fragment annotated are summarized in Table 2.

Table 2 Simulation results for spreading activation based method and tf-idf based method

Method	Precision
Tf-idf based with no annotations	21.32%
Proposed with no annotations	21.96%
Tf-idf based with generated annotations	33.64%
Proposed with generated annotations	37.07%
Tf-idf based with whole fragment annotated	43.20%
Proposed with whole fragment annotated	53.34%

Proposed method based on spreading activation obtained similar or better results to tf-idf based method in all performed experiments. The results of experiments with no annotations, where only the content of the document was used to create query, suggests that proposed method provides similar, even better results for query word extraction. These results were achieved despite the fact that proposed method is using only information from the document content and not the information about other documents in the collection by contrast to tf-idf based method. The proposed method can thus be used as an alternative to tf-idf based method when creating query from document content.

The comparison of both methods without using annotations and using generated annotations in query construction process proved that annotations are increasing precision of related documents retrieval.

The experiment with whole document fragments annotated suggests that with increasing number of annotations the precision of generated queries increases for both used methods for query word extraction.

We performed a Student's t-test on 5% level of significance for pairs of proposed method and tf-idf based method for every performed experiment to determine if we obtained statistically significant differences in mean precision for compared methods. As the computed p-value was less than 0.01% for every performed experiment, we can reject null hypothesis that the mean precisions of compared methods are equal. We obtained significant differences in mean precision for proposed method and tf-idf method for experiments using annotations in query construction process as well as for experiment with whole document fragments annotated.

To compare a real increase of precision of related document retrieval using annotations in query construction process and without using annotations, we performed a qualitative user study, where in sequence 8 volunteers were asked to annotate documents of their choice stored in the Annota service. After they annotated these documents, we generated two queries using proposed method, one using annotations and one without using annotations in query construction process. We retrieved two lists of documents using

these queries and we presented them to volunteers in random order. Volunteers were then asked to select documents describing topic related to the topic of source document from displayed lists and to select better from two presented lists.

The volunteers annotated 11 unique documents. In 9 cases they selected for more relevant the list created by method using annotations. In one case method using annotations created query in Slovak and document search returned no documents. This was caused by the fact, that in this document all annotations were written in Slovak and all documents we searched in were in English. In one case the method not using annotations obtained better results. By method taking into account annotations we obtained 34 relevant documents in total and with method not using annotations only 15.

Part of volunteers were writing annotations in Slovak, but to keep conditions the same as during document annotation out of the experiment, we allowed them to write annotations the same way they are used to. We asked one user to repeat the experiment on one document after he translated created annotations written in Slovak to English. When translated annotations were used in query construction all retrieved results were related to the source document.

In one case we asked the volunteer to repeat the experiment with increased number of annotations attached to the document. During this experiment, the volunteer doubled the number of attached annotations. In the second retrieved list of documents, the number of relevant documents retrieved increased and included one exact match with the topic user was most interested in while annotating source document. With increasing number of annotations attached to document the precision of related document retrieval is increasing.

When using annotations to create a query, the proposed method obtained better results than in the case when annotations were not used in the query construction process. When using annotations, created query retrieves more documents that describe the same topic as the source documents and more documents that describe related topics.

We used a questionnaire about user's habits when annotating documents to determine how users of Annota are creating annotations into studied documents. The majority of participants are using annotations while reading printed or electronic documents. When annotating electronic documents, they use various tools to create bookmarks, to-do lists, saving documents for later, to insert highlights, comments and other types of annotations into documents. The most frequently used types of annotations are tags and in-text highlights. The purpose for creating annotations such as notes, comments and highlights is to summarize studied documents, describe documents, highlight most important sections, to store their thoughts about studied documents and as a form of in-document navigation to support fast recollection of document when returning to previously studied document. The distribution of created in-text annotation was uniform over the whole text. In this study interviewed volunteers confirmed our assumption that using annotations users are indicating those parts of the document they are most interested in.

VI. Conclusion and future work

Annotations represent important source of information on interesting or important parts of documents. Its importance increases by possibilities of manipulating documents on the Web by the way we want to do commonly with paper documents. We studied user behavior while annotating documents on the Web and proposed a method for query construction from document content and attached annotations. In the process of query construction we considered document content and its structure by using text to graph transformation and query terms extraction using spreading activation in created graph. We used user created annotations as user's interest indicators to insert initial activation into graph created from document content.

We have developed a bookmarking service called Annota and a browser extension allowing users to insert various types of annotations into web pages and PDF documents displayed in web browser. The simulation based on probabilistic distributions of various parameters of annotations created by users of Annota proved, that annotations used when creating queries for related document retrieval can increase retrieval precision and with increasing number of attached annotations the precision rises.

We compared two methods for query word extraction. The method based on spreading activation in document text transformed to graph outperforms tf-idf based method when creating query for related documents search from source document and attached annotations. The proposed method achieved comparable results to tf-idf based method when no annotations were used in query construction. It is thus possible to use it even when no annotations are attached into the document with comparable precision as commonly used method when extracting words fit into query for related document retrieval from document content. The spreading activation based method outperformed compared method when document attached annotations were used in query construction process. The proposed method does not use information from other documents, only information from source document content and attached annotations. It is thus search engine independent and can be used to create queries for any search engine accepting queries in form of a list of keywords.

Performed user study showed that users insert annotations into document sections they are most interested in and they are use annotations to summarize documents, highlight most important parts of documents and to store their thoughts.

We evaluated proposed method for increasing related document retrieval precision of created query when using user created annotations in query construction process.

We plan to use annotations created not only by single user but also annotations created by other users when creating query for related document search. We see the potential in use of social relations such as group membership in weighting of annotations created by other users in query construction process.

In the described work, we were using annotations attached into document along with document content and we have not used user's annotations attached to other documents. By using annotations from other documents, we plan to model user's interests. Such annotation based user model can be used for further improvement of query construction process.

Moreover there are several possible enhancements related to document search process from the point of view of search engine. We plan to use annotations to enrich document content while creating index of annotated documents and we will compare performance of search in annotation enriched index against related document search in index created only from document content.

ACKNOWLEDGEMENT

This work was partially supported by the projects VG1/0675/11, VG1/0971/11 and APVV-0208-10. The authors wish to thank members of Annota team Michal Holub, Róbert Móro, Roman Burger, Martin Lipták, Juraj Kostolanský, Peter Macko and Samuel Molnár for their contribution to Annota design and implementation.

REFERENCES

[1] M. Agosti, N. Ferro, "A formal model of annotations of digital content." ACM Trans. Inf. Syst., Nov. 2007, vol. 26, no. 1.

[2] X. Zhang, L. Yang, X. Wu, et al., "sDoc: exploring social wisdom for document enhancement in web mining." In Proc. of the 18th ACM conf. on Inf. and knowledge management, ACM, 2009, pp. 395–404.

[3] P. Návrat, "Cognitive traveling in digital space: from keyword search through exploratory information seeking." Central European Journal of Computer Science, vol. 2, no. 3, pp. 170-182.

[4] M. Šimko, M. Barla, V. Mihál, M. Unčík, M. Bieliková, "Supporting Collaborative Web-based Education via Annotations." In World Conf. on Educational Multimedia, AACE, 2011, pp.2576-2585.

[5] D. Millen, M. Yang, S. Whittaker, J. Feinberg, "Social bookmarking and exploratory search." ECSCW 2007, Springer, London, 2007, pp. 21–40.

[6] C. Cattuto, C. Schmitz, A. Baldassarri, et al. "Network properties of folksonomies." AI Comm., 2007, vol. 20, no. 4, pp. 245-262.

[7] R. Moro, M. Bieliková, "Personalized Text Summarization Based on Important Terms Identification." 23rd Int. Workshop on Database and Expert Systems Applications, IEEE, 2012, pp. 131–135.

[8] Y. Yanbe, A. Jatowt, S. Nakamura, K. Tanaka, "Can social bookmarking enhance search in the web?" In Proc. of the 7th ACM/IEEE-CS joint conf. on Digital libraries, ACM, 2007, pp. 107–116.

[9] C. Biancalana, A. Micarelli, "Social tagging in query expansion: A new way for personalized web search." Computational Science and Engineering, 2009. vol. 4. IEEE, pp. 1060-1065.

[10] R. Abbasi, "Query expansion in folksonomies." In Semantic Multimedia, Springer Berlin Heidelberg, 2011, pp. 1-16.

[11] G. Golovchinsky, M. N. Price, B. N. Schilit, "From reading to retrieval: freeform ink annotations as queries." SIGCHI Bulletin. ACM Press, 1999, pp. 19–25.

[12] Y. Yang, N. Bansal, W. Dakka, et al., "Query by document." In Proc. of the Second ACM International Conf. on Web Search and Data Mining, ACM, 2009 , pp. 34–43.

[13] T. Strohman, W. B. Croft, D. Jensen, "Recommending Citations for Academic Papers." In Proc. of the 30th annual int. SIGIR conf. on Research and development in inf. retrieval, ACM, 2007, pp. 5–6.

[14] M. Kompan, M. Bieliková, "Content-based News Recommendation." In E-Commerce and Web Technologies, Lecture Notes in Business Information Processing, vol. 61, part 2, Springer, pp.61-72.

[15] Á. R. Pereira, N. Ziviani, "Retrieving similar documents from the web." Journal of Web Engineering, vol. 2, no. 4, 2003, pp. 247-261.

[16] Z. Zhang, J. Iria, C. A. Brewster, F. Ciravegna, "A comparative evaluation of term recognition algorithms." In Proc. of 6th Int. Conf. on Language Resources and Evaluation, Marrakech Morocco, 2008.

[17] D. Paranyushkin, "Visualization of Text's Polysingularity Using Network Analysis." Prototype Letters, 2011, vol. 2, no. 3, pp. 256–278.

[18] G. K. Palshikar, "Keyword extraction from a single document using centrality measures." Pattern Recognition and Machine Intelligence, Springer Berlin Heidelberg, 2007, pp. 503-510.

[19] T. A. Phelps, R. Wilensky, "Robust intra-document locations." Computer Networks, 2000, vol. 33, no. 1, pp. 105-118.

[20] P. Ciccarese, M. Ocana, L. J. Garcia Castro, S. Das, T. Clark, "An open annotation ontology for science on Web 3.0." J Biomed Semantics, 2011, vol. 2, no. 2.

[21] J. Ševcech, M. Bieliková, R. Burger, M. Barla, "Logging activity of researchers in digital library enhanced by annotations." In Proc. of 7th Workshop on Int. and Knowledge oriented Tech., 2012, pp. 197-200. (in Slovak)

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 287–290

Ontology of architectural decisions supporting ATAM based assessment of SOA architectures

Piotr Szwed*, Paweł Skrzynski*, Grzegorz Rogus* and Jan Werewka*
*AGH University of Science and Technology
Department of Applied Computer Science
Email: {pszwed,skrzynia,rogus,werewka}@agh.edu.pl

***Abstract*—Nowadays, Service Oriented Architecture (SOA) might be treated as a state of the art approach to the design and implementation of enterprise software. Contemporary software developed according to SOA paradigm is a complex structure, often integrating various platforms, technologies, products and design patterns. Hence, it arises a problem of early evaluation of a software architecture to detect design flaws that might compromise expected system qualities. Such assessment requires extensive knowledge gathering information on various types of architectural decisions, their relations and influences on quality attributes. In this paper we describe SOAROAD (SOA Related Ontology of Architectural Decisions), which was developed to support the evaluation of architectures of information systems using SOA technologies. The main goal of the ontology is to provide constructs for documenting SOA. However, it is designed to support future reasoning about architecture quality and for building a common knowledge base. When building the ontology we focused on the requirements of Architecture Tradeoff Analysis Method (ATAM) which was chosen as a reference methodology of architecture evaluation.**

***Index Terms*—software architecture, ontology, SOA, ATAM, architecture assessment, architecture evaluation, enterprise architecture**

I. Introduction

NOWADAYS, Service Oriented Architecture (SOA) might be treated as a state of the art approach to the design and implementation of enterprise software, which is driven by business requirements. Within the last decade a number of concepts related to SOA have been developed, including ESB (Enterprise Service Bus), web services, design patterns, service orchestration and choreography and various security standards. Due to the fact that there are many technologies that cover the area of SOA, the development and evaluation of SOA compliant architectures is especially interesting.

SOAROAD has been designed as a methodology for the assessment of software architectures developed according to SOA principles. It is based on the Architecture Tradeoff Analysis Method (ATAM) [11], [5], which is a mature, scenario-based, early method for architecture assessment. ATAM defines a quality model, an organizational framework for evaluation process and expected results: sensitivity points, tradeoffs and risks. A limitation of the ATAM method is that it depends on experts knowledge, perception and previous experience. It may easily happen, that an inexperienced evaluator overlooks some implicit decisions and risks introduced by them.

In the SOAROAD approach the very basic set of ATAM terms used to describe architecture is enriched by including common terminology and relationships between concepts related to various aspects of service oriented architecture design and development. The gathered knowledge, formalized as an ontology, facilitates performing an assessment in more exhaustive manner, helping to ask questions, revealing implicit design decisions and obtaining more reliable results.

The contribution of the paper is a proposal of a SOAROAD ontology as a tool supporting scenario based assessment of systems following a service-orientation paradigm and service design, development and deployment.

II. Related works

Architecture evaluation has attracted many researchers and practitioners during the last 20 years. A survey paper on this topic [18] lists 37 methods of architecture evaluation, classifying them according to two dimensions: location in the software lifecycle (early vs. late) and element being analyzed (system architecture, isolated architectural style or a design pattern). The paper suggests that scenario-based methods, including SAAM [12] and ATAM [11], [5] can be considered as a mature, reliable and easy to implement in practical situations. There are several reports on successful applications of ATAM for assessment of a battlefield control system [13], wargame simulation [10], product line architecture [8], control of a transportation system [3], credit card transactions system[16] and a dynamic map system [21]. Recently, a few extensions of ATAM were proposed, including a combination with the Analytical Hierarchy Process [24] and APTIA [14].

The application of ontologies to provide a systematic and formal description of architectural decisions was first proposed by Kruchten in [15]. The ontology distinguished several types of decisions that can be applied to software architecture and its development process. Main categories included: Existence, Ban, Property and Executive decisions. The ontology defined also attributes, which were used to describe decisions, including states (Idea, Tentative, Decided, Rejected, etc.). In [7] an ontology supporting ATAM based evaluation was proposed. The ontology specified concepts covering the ATAM model of architecture, quality attributes, architectural styles and decisions, as well as influence relations between elements of

This work was supported from AGH University of Science and Technology under Grant No. 11.11.120.859

978-83-60810-53-8/$25.00

architectural style and quality attributes. The effort to structure the knowledge about architectural decisions, was accompanied by works aimed at a development of tools enabling the edition and graphical visualization of design decisions, often in a collaborative mode, e.g. [4], [6], [17].

III. A CONCEPT OF APPLICATION OF SOAROAD ONTOLOGY

The SOAROAD (SOA Related Ontology for Architectural Decisions) has four main goals, it should: (1) provide a comprehensive description of architectural views, i.e. components and their connections; (2) gather a domain knowledge providing a unified vocabulary related to SOA and enterprise architecture; (3) help to ask question about various properties of architectural design and decisions; (4) be capable of representing assignments of properties relevant to SOA compliant technologies to elements of system architecture.

It was assumed that the ontology would follow a foundational model (ontology skeleton) defining various properties corresponding to design decisions that can be attributed to components, connections, interfaces and compositions. If applicable, these design decisions can be supplemented by additional relations. The ontology would also specify design patterns.

Another assumption is related to a distribution of the knowledge between ontology TBox (set of classes, their attributes and relations) and ABox (individuals, values of their attributes and relationships). The types of elements appearing in architectural views are classified in the TBox. Concrete elements, e.g. those appearing in the diagrams of architectural views, are represented as individuals in an ABox. The ontology describes types of design decisions (properties) as classes, whereas their values as individuals that can be directly assigned to elements of architectural views or linked to form trees.

The concept of the ontology application is presented in the Fig. 1 The process of building an architecture description starts with eliciting *Architecture views ABox*, i.e. a set of linked

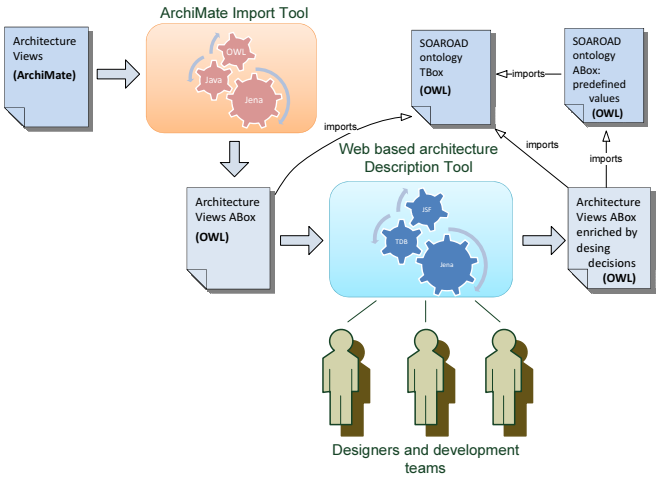


Fig. 1. A concept of application of SOAROAD ontology

components, interfaces and connections. This model can be prepared either manually or with the support of dedicated import tools converting ArchiMate [22] models of Archi editor [1] or UML [19], e.g. from VisualParadigm.

A web based tool supporting architecture description uses the classes and individuals defined in the *SOAROAD ontology Domain Description TBox* and *SOAROAD Architectural decisions ABox* to generate forms or questionnaires in which software architects or members of development teams can make assignments of property values to elements of architecture views.

The resulting *Detailed Architecture ABox* refers elements of *Architecture views ABox* and individuals defined in SOAROAD ontology (merging two input ontologies and asserting additional relations). This ontology serves as a detailed architecture documentation within a software development project. It can be examined either manually or with use of automated tools.

IV. ONTOLOGY DESCRIPTION

The SOAROAD ontology was built in three steps. Firstly, a foundational model serving as ontology skeleton was proposed. Then we manually gathered and analyzed information related to service oriented architectures, technologies, architectural approaches, design patterns, etc. originating from various sources: books, technical papers, reference manuals and Internet resources. Finally, the ontology was populated with this information by translating intermediate textual description into OWL constructs. At present the ontology consists of 110 classes, 9 object properties and 105 individuals.

The basic model of software architecture used in ATAM [2] defines it after [20] as a set of components and linking them connections. We extend this simplistic model by defining *Interfaces* and *Functions* of components as presented in Fig. 2. A connection links a component having the caller role with an interface (calee). Components, connections and interfaces can be attributed with: *ComponentProperties*, *ConnectionProperties* and *InterfaceProperties* respectively (Fig. 3). Examples of such properties are: platform, web service type, communication type, queueing and query granularity.

Composition is a coherent set of components and connectors. System architecture is itself a composition. For the purpose of analysis we may focus on a particular subset of components and connectors and describe their properties, e.g. a distribution of queries among several databases building up a composition or realization of a design pattern.

During the ATAM based evaluation the overall system architecture and properties of its parts are analyzed to establish scenario responses and achievements of corresponding quality attributes. It may be, however, observed that some architecture properties or their combinations have known influence on quality attributes, e.g. a use of asynchronous web services or applying MVC design pattern, which increases modifiability and a granularity of queries, has an impact on performance. This kind of knowledge can be expressed via *influences* relations.

Architectural decision is an assignment of a property value to a component, interface, connection or a composition. In this context the terms *property* and *architectural decision* can be used to some extent interchangeably. However, it may happen that certain decisions or components are dependent on previously assigned properties. An example of such a dependency is the composition type – a property assigned to a set (composition) of web service components. Selecting orchestration as the composition type requires that an orchestration component, e.g. BPEL capable module is used. The *required* relation or its subproperties in the ontological model express this dependency.

The assumed foundational model adopts a reification strategy while modeling various properties of an architectural design. Properties are defined as classes, whose individuals can be linked by additional relations indicating specific roles. An example of such a property is MVC design pattern, which requires the identification of components playing the roles of a Model (typically a database), a Controller (e.g. an EJB) and a View (e.g. a set of HTML pages produced by JSP scripts).

For each property, that can be treated as a class of design decision, a number of individuals (corresponding to decision values) is defined. They can be selected in assignments, e.g. *JavaEECompliantAS* (a subclass of *ComponentProperty*) has several predefined individuals: *JBoss*, *Glassfish*, *WebLogic*, *Web-Sphere*, *ColdFusion*, etc.

Example ontology assertions related to component properties are presented in Table I and Table II. A property (an ontology class) is followed by property values (individuals in the ontology) put in parentheses.

Apart from defining design decisions, the ontology specifies functions of components. Class Function contains classes of entities such as: *Routing*, *MessageMapping*, *ProtocolSwitch*, *MediationService*, *MessageValidation*, *AuditFunction*, *DatbaseIntegration*, etc.

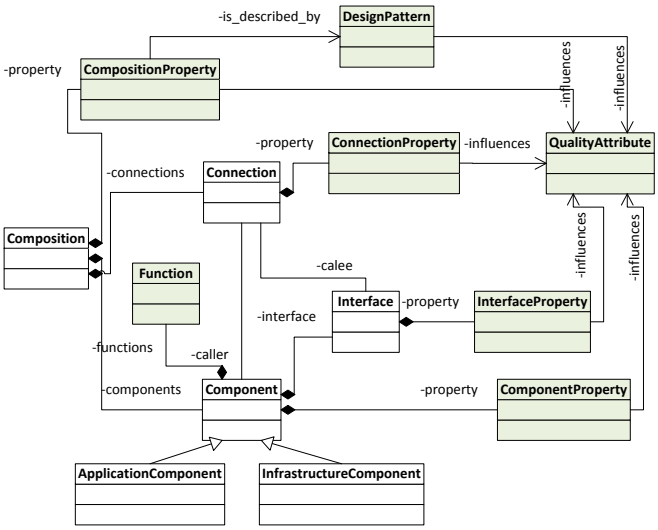


Fig. 2. Foundational model of software architecture and its properties

TABLE I
COMPONENT PROPERTIES

Property (values)	Description
PlatformTechnology (CORBA, EJB, JINI, RMI)	Set of technologies used on the platform.
ComponentLogic (flexible, fixed, rulebased)	Specifies an approach the component logic implementation.
Platform	Defines the component platform. Has several subclasses: ApplicationServer, Hardware, OperatingSystem and VirtualServer
ProgrammingLanguage (Cpp, Java, Ruby, PHP, Erlang, Python, C, C_sharp)	Define programming language used to implement a component.
StatePersistence (Stateless, Statefull)	Specifies whether a component saves internal data during and in between calls of operations on the client's behalf.

TABLE II
PROPERTIES DESCRIBING PLATFORM (SUBCLASSES OF *Platform*).

Property (values)	Description
ApplicationServer	Subclass of Platform. Defines an application server on which a component is deployed, can have such attributes, as: version (string), vendor (string)
JEECompliantAS (TomEE, Glassfish, JBoss, Interstage, JOnAS, Geronimo, SAPNeatWeaver, WebSphere, Resin, ColdFusion, WebLogic)	Subclass of ApplicationServer dedicated to JEE compliant components.
DotNetCompliantAS (AppFabric, IIS, TNAPS, Base4, Mono)	Subclass of ApplicationServer; its individuals define products for .NET enviroment
JavaAS (Jetty, Enhydra, iPlanet)	Application servers for Java environment
Hardware	Subclass of Platform. Used to specify a hardware configuration on which the component is deployed. Attributes: memory (double), processor (string), number_of_cores (int)
OperatingSystem (Windows, Unix, Linux, iOS, Android, Bada, Blackberry)	Subclass of Platform. Defines types of operating systems on which a component is executed. Attributes: version (string), vendor (string), product (string)
VirtualServer (no, yes)	Subclass of Platform. Specifies whether a component is deployed on a virtual server

The ontology provides also a taxonomy of quality attributes. A quality attribute is a nonfunctional characteristic of a component or a system. It represents the degree to which software possesses a desired combination of properties, which are defined by means of externally observable features. Some of the attributes are related to the overall system design, while others are specific to run-time or design time.

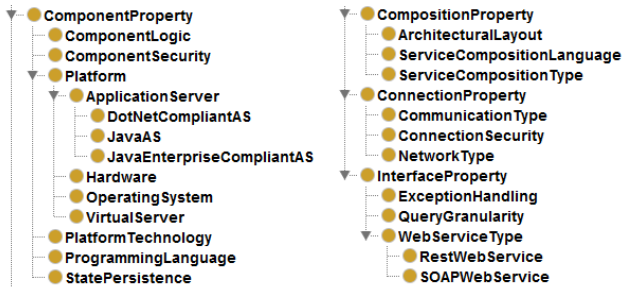


Fig. 3. Classes of properties

SOAROAD ontology defines 30 quality attributes including both terms defined in software quality model by the ISO/IEC 9126-1 norm [9] and those arising directly from requirements to architectures formulated in the SOA manifesto [1]. Examples of classes belonging to the first group are: *Functionality*, *Reliability*, *Usability*, *Efficiency*, *Maintainability* and *Portability*. The example of classes originating from SOA manifesto are *ServiceAutonomy*, *PlatformIndependency*, *LooseCoupling*, *Modularity*, *OpenStandardAdoptation*, *BusinessAgility*, etc.

When designing an applications to meet quality requirements, it is necessary to consider a potential impact of design properties on various quality attributes. SOAROAD ontology defines *influences* object property to this kind of relation.

A design pattern can be seen as a structure build of components of particular types, defining their roles and relations among them together with a set of restrictions on their usage. Design patterns do not change the functionalities of a system but only their organization or structure. One of the most important benefits of using design patterns is that they constitute standardized software building blocks with a well defined influence on quality attributes. In SOAROAD ontology the class *DesignPattern* has 56 subclasses representing patterns dedicated to SOA architecture. The examples of subclasses are: *db.EnterpriseServiceBus*, *db.EventDrivenMessaging*, *db.Orchestration*. The relation *is_described_by* links a particular *CompositionProperty* to one of the defined design patterns.

V. Conclusion

This paper describes the SOAROAD ontology and the concept of a tool supporting documentation of architectures of SOA-based systems. The proposed approach addresses the problem that can be encountered during architecture assessment: to be reliable, a reasoning about architecture qualities, must have solid foundations in a knowledge related to a particular domain: architectural styles, design patterns, used technologies and products. The idea behind SOAROAD ontology is to gather experts knowledge to enable even inexperienced users performing ATAM-based architecture evaluation. An advantage of the presented approach is that its result is a joint representation of architecture views and properties attributed to design elements formalized in OWL language.

From a software engineering perspective, such centralized information resource maintained during the software lifecycle may represent a valuable artifact, which can provide reference to design decisions throughout integration, testing and deployment phases.

On the other hand, a machine interpretable representation, constituting a graph of interconnected objects (individuals), can be processed automatically to check consistency, detect potential flaws and calculate metrics. An extensive list of metrics related to architectural design was defined in [23]. We plan to adapt them to match the structural relations in the SOAROAD ontology, as well to develop new ones.

[1] http://www.soa-manifesto.org/

References

[1] "Archi, archimate modelling tool," 2011, [Online; accessed 23-June-2012]. [Online]. Available: http://archi.cetis.ac.uk/download.html

[2] P. Bianco, R. Kotermanski, and P. Merson, "Evaluating a service-oriented architecture," Carnegie Mellon, Technical Report CMU/SEI-2007-TR-015, September 2007.

[3] N. Bouck'e, D. Weyns, K. Schelfthout, and T. Holvoet, *Applying the ATAM to an Architecture for Decentralized Control of a Transportation System*. Springer, 2006, vol. 4214, pp. 180–198.

[4] R. Capilla, F. Nava, S. Pérez, and J. C. Dueñas, "A web-based tool for managing architectural design decisions," *ACM SIGSOFT Software Engineering Notes*, vol. 31, no. 5, 2006.

[5] P. Clements, R. Kazman, and M. Klein, *Evaluating Software Architectures: Methods and Case Studies*. Addison-Wesley Professional, 2001.

[6] R. C. de Boer, P. Lago, A. Telea, and H. van Vliet, "Ontology-driven visualization of architectural design decisions," in *WICSA/ECSA*. IEEE, 2009, pp. 51–60.

[7] A. Erfanian and F. S. Aliee, "An ontology-driven software architecture evaluation method," in *Proceedings of the 3rd international workshop on Sharing and reusing architectural knowledge*, ser. SHARK '08. New York, NY, USA: ACM, 2008, pp. 79–86.

[8] S. Ferber, P. Heidl, and P. Lutz, *Reviewing product line architectures: Experience report of ATAM in an automotive context*. Springer, 2001, vol. 2290, pp. 364–382.

[9] ISO/IEC, "Software engineering – product quality, ISO/IEC 9126-1," International Organization for Standardization, Tech. Rep., 2001.

[10] L. G. Jones and A. J. Lattanze, "Using the architecture tradeoff analysis method to evaluate a wargame simulation system: A case study," *Technical Report CMUSEI2001TN022 Software Engineering Institute Carnegie Mellon University Pittsburgh PA*, no. December, p. 33, 2001.

[11] Kazman, "Atam:method for architecture evaluation," *CMU-SEI2000TR004*, 2000.

[12] R. Kazman, L. Bass, G. Abowd, and M. Webb, *SAAM: a method for analyzing the properties of software architectures*. IEEE Comput. Soc. Press, 1994, vol. 16pp, no. 5/11/2011, pp. 81–90.

[13] R. Kazman, M. Barbacci, M. Klein, J. Carriere, and S. G. Woods, "Experience with performing architecture tradeoff analysis," *Proceedings of the 21st international conference on Software engineering ICSE 99*, pp. 54–63, 1999.

[14] R. Kazman, L. Bass, and M. Klein, "The essential components of software architecture design and analysis," *Journal of Systems and Software*, vol. 79, no. 8, pp. 1207–1216, 2006.

[15] P. Kruchten, *An ontology of architectural design decisions in software intensive systems*. Citeseer, 2004, pp. 54–61.

[16] J. Lee, S. Kang, H. Chun, B. Park, and C. Lim, "Analysis of VAN-core system architecture- a case study of applying the ATAM," in *Proceedings of the 2009 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing*, ser. SNPD '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 358–363.

[17] L. Lee and P. Kruchten, *Visualizing Software Architectural Design Decisions*. Springer-Verlag, 2008, vol. 5292, pp. 359–362.

[18] B. Roy and T. C. N. Graham, "Methods for evaluating software architecture : A survey," *Computing*, vol. 545, no. 2008-545, p. 82, 2008.

[19] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual, The (2nd Edition)*. Pearson Higher Education, 2004.

[20] M. Shaw and D. Garlan, *Software Architecture: Perspectives on an Emerging Discipline*, M. Shaw and D. Garlan, Eds. Prentice Hall, 1996, vol. 123.

[21] P. Szwed, I. Wojnicki, S. Ernst, and A. Glowacz, "Application of new ATAM tools to evaluation of the dynamic map architecture," in *Multimedia Communications, Services and Security*, ser. Communications in Computer and Information Science, A. Dziech and A. Czyżewski, Eds. Springer Berlin Heidelberg, 2013, vol. 368, pp. 248–261.

[22] The Open Group, "Archimate 1.0 specificattion," 2009. [Online]. Available: http://www.opengroup.org

[23] A. Vasconcelos, P. Sousa, and J. Tribolet, "Information system architecture metrics: an enterprise engineering evaluation approach," *The Electronic Journal Information Systems Evaluation*, vol. 10, no. 1, pp. 91–122, 2007.

[24] P. Wallin, J. Froberg, and J. Axelsson, "Making decisions in integration of automotive software and electronics: A method based on ATAM and AHP," *Fourth International Workshop on Software Engineering for Automotive Systems SEAS 07*, pp. 5–5, 2007.

Workshop on Computational Optimization

Many real world problems arising in engineering, economics, medicine and other domains can be formulated as optimization tasks. These problems are frequently characterized by non-convex, non-differentiable, discontinuous, noisy or dynamic objective functions and constraints which ask for adequate computational methods.

The aim of this workshop is to stimulate the communication between researchers working on different fields of optimization and practitioners who need reliable and efficient computational optimization methods.

We invite original contributions related to both theoretical and practical aspects of optimization methods.

Topics

The list of topics includes, but is not limited to:

- unconstrained and constrained optimization
- combinatorial optimization
- global optimization
- multiobjective optimization
- optimization in dynamic and/or noisy environments
- large scale optimization
- parallel and distributed approaches in optimization
- random search algorithms, simulated annealing, tabu search and other derivative free optimization methods
- nature inspired optimization methods (evolutionary algorithms, ant colony optimization, particle swarm optimization, immune artificial systems etc)
- hybrid optimization algorithms involving natural computing techniques and other global and local optimization methods
- optimization methods for learning processes and data mining
- computational optimization methods in statistics, econometrics, finance, physics, medicine, biology, engineering etc

Event Chairs

Fidanova, Stefka, Academy of Sciences, Bulgaria
Mucherino, Antonio, IRISA, France
Zaharie, Daniela, West University of Timisoara, Romania

Program Committee

Andonov, Rumen, IRISA and University of Rennes 1, Rennes, France
Bartl, David, University of Ostrava, Czech Republic
Brest, Janez, University of Maribor, Slovenia
Goncalves, Douglas, IRISA, University of Rennes 1, France
Hosobe, Hiroshi, Hosei University, Japan
Iiduka, Hideaki, Kyushu Institute of Technology, Japan
Lampinen, Jouni, University of Vaasa, Finland
Lavor, Carlile, IMECC-UNICAMP, Brazil
Marinov, Pencho, Bulgarian Academy of Science, Bulgaria
Miettinen, Kaisa, University of Jyvaskyla, Finland
Pardalos, Panos, University of Florida, United States
Siarry, Patrick, Universite Paris XII Val de Marne, France
Stefanov, Stefan, South-West University ""Neofit Rilski, Bulgaria
Stuetzle, Thomas, Université Libre de Bruxelles (ULB) , Belgium
Suganthan, Ponnuthurai Nagaratnam, Nanyang Technological University, Singapore
Tvrdik, Josef, University of Ostrava, Czech Republic
Vrahatis, Michael, University of Patras, Greece

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 293–297

A quasi self-stabilizing algorithm for detecting fundamental cycles in a graph with DFS spanning tree given

Halina Bielak*, Michał Pańczyk†
* Institute of Mathematics, Maria Curie-Skłodowska University,
Lublin, Poland,
Email: hbiel@hektor.umcs.lublin.pl
† Institute of Computer Science, Maria Curie-Skłodowska University,
Lublin, Poland,
Email: mjpanczyk@gmail.com

***Abstract*—This paper presents a linear time quasi self-stabilizing algorithm for detecting the set of fundamental cycles on an undirected connected graph modeling asynchronous distributed system. Previous known algorithm has $\mathcal{O}(n^2)$ time complexity, whereas we prove that our stabilizes after $\mathcal{O}(n)$ moves. Distributed adversarial scheduler is considered. Both algorithms assume that the depth-first search (DFS) spanning tree (DFST) of the graph is given. The output is given in a distributed manner as a state of variables in the nodes.**

***Index Terms*—Self-stabilization, fundamental cycles, fault tolerance.**

I. Introduction

A NOTION of self-stabilizing algorithms on distributed systems was introduced by Dijkstra [1] in 1974. They can model distributed systems with message passing or shared memory. A survey in the topic can be found in the paper by Schneider [2], and more details in the book by Dolev [3]. The notions from the graph theory not defined in this paper can be found in the book by Harary [4].

A distributed self-stabilizing system consists of a set of processes (i.e. computing nodes) and communication links between them. Every node in the system runs the same algorithm and can change state of local variables. These variables determine *local state* of a node. Nodes can observe the state of variables on themselves and their neighbor nodes. The state of all the nodes in the system determines the *global state*.

Every self-stabilizing algorithm should have a class of global states called *legitimate state* defined, when the system is stable and no action can be done by the algorithm itself. Every other global state is called *illegitimate* and for the algorithm to be correct there has to be some possibility to make a move if the state is illegitimate. The aim of the self-stabilizing algorithm is to bring the legitimate (desirable) state of the whole system after some alteration (from the outside of the system) of variables in the nodes or after the system has been started.

Our model of quasi self-stabilization is a modification of the classic self-stabilization. We allow the algorithm to assume that some variables have fixed value at the beginning of the computation. Similar model has been considered by Arora and Gouda [7] and Schneider [2]. It is motivated by the fact that hardware may have the so-called protected memory. This kind of memory is resistant to changes from the outside of the system.

In this paper we present an $\mathcal{O}(n)$ time modification (we call it ASFC II) of Chaudhuri's [5] algorithm (ASFC I) for detecting fundamental cycles in a graph. A set of fundamental cycles (SFC) is a collection of cycles that can be used to construct any other cycle in the graph. Given any cycle C from graph G, it can be constructed as $C = C_1 \oplus C_2 \oplus \ldots \oplus C_k$, where $C_i \in SFC(G)$ and $G_1 \oplus G_2$ is such a subgraph constructed from subgraphs G_1 and G_2 of the G graph, that each of its edges exists either in G_1 or in G_2, so it is symmetric difference.

ASFC I requires $\mathcal{O}(n^2)$ time to stabilize, where n is the number of processes in a system. It is able to start at arbitrary configuration, given that DFST is known. In fact $\mathcal{O}(n^2)$ time complexity was assumed when no information about spanning tree structure can be altered. The information about spanning tree is meant to be stored in a protected memory.

ASFC II requires additional 2-bit variable q in each process. The initial state of q should be set right after (or during) finding the DFST. It must be also stored in a protected memory. After finding DFST and setting q, the essential part of the algorithm can be started (see chapter 3 for the details).

The effects of running both algorithms are stored in a distributed manner as a state of local variables in the processing nodes; i.e. when a system running the algorithm stabilizes, each of its nodes knows how many fundamental cycles passes through it and what are their identifiers.

The rest of this article is organized as follows. Section II introduces the notation and model of computation used in next sections. Section III presents the original Chaudhuri algorithm

978-83-60810-53-8/$25.00

and our improved version. In section IV we prove convergence and time characteristic of our algorithm, and in the last section V we conclude the paper.

II. Notation and computational model

We consider a self-stabilizing system modeled by a finite, undirected graph $G = (V, E)$. Let the number of vertices be $n = |V|$ and the number of edges $e = |E|$. We think of every process as a node in the graph, whereas connection links between them are edges. There are some variables in each node. Names and types of the variables are set during design of an algorithm. Every node can also look up for the state of the variables at its neighbors.

Each node runs the same algorithm. The algorithm consists of a set of rules. A rule has the form as below.

An example rule
label: **if** *guard* **then** assignment instructions

A *guard* is a logic predicate which can refer to variables in the node itself and its neighbors. We say that a rule is *active* if its guard is evaluated to be true. A node is *active* if it contains any active rule. If there is no active node in the graph, we say that the system is *stabilized*.

A self-stabilizing system contains also a *scheduler*. Its task is to choose one process from the set of active processes and to trigger an active rule in it. Such an action we call a *move*. In this article we assume the scheduler is distributed and adversarial, so the order of activating nodes is nondeterministic. As a consequence, while describing pessimistic complexity of the algorithm (number of moves), we must take the worst case scenario of triggering actions in particular nodes.

Now we present some notation used in the paper. Most of the following notation is the same as in [5]. As mentioned before, we consider connected, undirected graph $G = (V, E)$ with the vertex set V and the edge set E. We also assume that the DFST $T(r) = (V, E')$ with the root node r is given. Note that the set of nontree edges is $E - E'$. So we have the following notation:

- $n(i)$ the set of neighbor nodes of the node i,
- $p(i)$ the parent node of i in $T(r)$ (we assume that $p(r) = null$),
- $c(i)$ the set of children of the node i in $T(r)$ (leaf nodes have $c(i) = \emptyset$),
- $nt(i)$ the set of nontree edges incident to the node i ($nt(i) = \{\{i, k\} : \{i, k\} \notin E'\}$),
- $C(i, j)$ the fundamental cycle created by the nontree edge $(i, j) \in E - E'$ together with the path between i and j in the tree $T(r)$,
- $a(i)$ the set of ancestors of the node i (we assume that $i \in a(i)$),
- $d(i)$ the set of descendants of the node i (we assume that $i \in d(i)$),
- $s(i)$ the set of all nontree edges such that each of them connects a descendant and a proper ancestor of i (more precisely $s(i) = \{\{x, y\} : \{x, y\} \in E - E', x \in d(i), y \in a(i) - i\}$),
- $su(i)$ $\bigcup_{j \in c(i)} s(j)$,
- $fc(i)$ the set of nontree edges such that the fundamental cycle created by each of them passes through i (more precisely $fc(i) = \{\{x, y\} : \{x, y\} \in E - E', x \in a(i), y \in d(i)\}$),
- $A \triangle B$ symmetric difference of the sets A and B ($A \triangle B = (A \cup B) - (A \cap B)$).

Some examples are presented in Fig. 1.

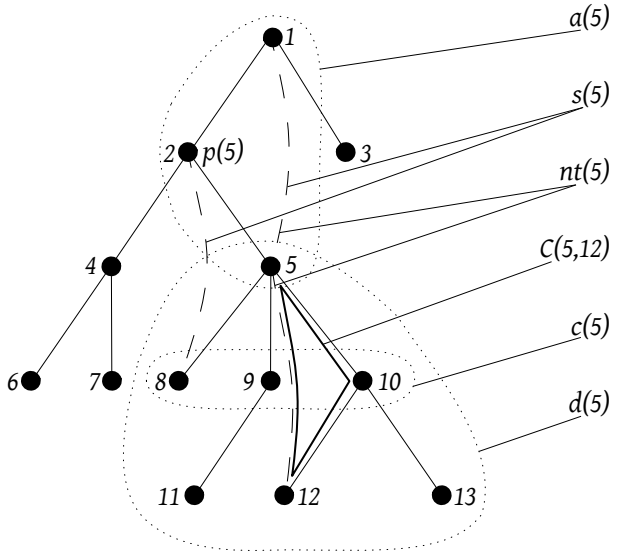


Fig. 1. Illustration of the notation for node $i = 5$: $n(5) = \{1, 2, 8, 9, 10, 12\}$, $su(5) = \{\{2, 8\}, \{5, 12\}\}$, $fc(5) = \{\{1, 5\}, \{2, 8\}, \{5, 12\}\}$. Dashed lines indicate nontree edges. Solid lines indicate tree edges.

III. The algorithm

As it was mentioned before, our algorithm (ASFC II) is a modification of Chaudhuri's original one (ASFC I). Our change alters only the starting precondition for the system and the order in which processing nodes are triggered (so the complexity changes). The difference is that ASFC I runs in $\mathcal{O}(n^2)$ time if there is DFST of the system given, and in $\mathcal{O}(n^3)$ time if there is not DFST detected yet. This situation requires running an algorithm for finding DFST. Algorithm developed by Colin and Dolev [6] can be used for that.

In contrast our ASFC II algorithm runs in $\mathcal{O}(n)$ time, but there have to be DFST given for the system and all the nodes must have the special variable $q(i)$ set to the **Stationary** state. If the DFST was not given or the information was corrupted, complexity is also $\mathcal{O}(n^3)$ because of need to run the algorithm finding DFST.

When the DFST is correct, but $q(i)$ in some nodes is wrong, complexity is $\mathcal{O}(n^2)$. That is because pessimistic scenario is quite the same as in ASFC I, and $q(i)$ can be ignored while estimating the number of moves.

All the semantics connected essentially with SFC, i.e. states of variables representing SFC in the distributed manner in ASFC II, are unaltered when compared to ASFC I. Thus we present here summary of the original algorithm; complete version can be found in [5].

Let us assume a DFST of the system is given, otherwise both algorithms must first calculate it. Two variables $p(i)$, $c(i)$ in each node state the structure of the tree and inferred from these two the value $nt(i)$.

Actually the main objective is to find $fc(i)$ for every node i in the system. The correct state of $fc(i)$ propagates along with $s(i)$ from leaves of the tree through internal nodes up to the root. The original Chaudhuri's algorithm is given as Alg. 1.

Algorithm 1: The ASFC I algorithm

leaf: **if** $c(i) = \emptyset \wedge (s(i) \neq nt(i) \vee fc(i) \neq nt(i))$ **then**
  $s(i) := nt(i)$
  $fc(i) := nt(i)$
internal: **if**
$c(i) \neq \emptyset \wedge (s(i) \neq nt(i) \triangle su(i) \vee fc(i) \neq su(i))$ **then**
  $s(i) := nt(i) \triangle su(i)$
  $fc(i) := su(i)$

Now we state the theorem, whose proof can be found in [5].

Theorem 1. *The ASFC I algorithm stabilizes the system after at most $\mathcal{O}(n^2)$ moves, under the condition that DFST is given.*

The meaning of symmetric difference in the internal rule is that as the computation passes along the treepath from leaves to the root the rule first incorporates the nontree edge to the set $s(i)$. The next time the same nontree edge is encountered (but the other end vertex of it), it is deleted from the $s(i)$ set.

A simple example of computation of the ASFC I algorithm, which exploits fully $\mathcal{O}(n^2)$ time is when DFS search gives simple path and each node is active but the scheduler every time chooses a node nearest to the root. First move would be done by the root node. Second move would be done by the unique child of the root, and then third by the root, because it was made active by triggering of its child. So in this way, the last phase would start from the unique leaf and finish in the root. There would be n phases, every i-th phase would involve i moves, so it gives $(n^2 + n)/2$ moves.

In the ASFC II algorithm the idea is not to make correction of the node's variables $fc(i)$ or $s(i)$ immediately after the fault has occurred. It would cause situation where ancestor nodes get "repaired" according to the incorrect information (not updated yet) from its children, as in the example above. Instead of this, every node is marked for updating and the update is done only if all the descendants are updated so they have correct values of $s(i)$ and $fc(i)$. A node cannot get back to **Stationary** state right after updating correct values of $s(i)$ or $fc(i)$, because it has to wait for all the nodes to be corrected (see rule 2.b). After all the nodes get updated, every node is being marked as almost stationary (root node first and then all its descendants recursively). Then all nodes can turn into stationary state — starting from leaves up to the root. An additional variable $q(i)$ for each node i is used

$$q(i) \in \{\textbf{Stationary}, \textbf{NeedUpdate}, \textbf{Updated}, \textbf{NearStationary}\}$$

(shortly: S, NU, U, NS), which represents the state of the node.

As it was mentioned above, we assume the DFST for the system is given and represented in a distributed way. According to Arora and Gouda [7] and Schneider [2], it is possible to make some precondition for a self-stabilizing system, therefore we assume a precondition such that the value of variable $q(i)$ for every node i is **Stationary**.

One can fulfill this requirement for example by putting the $q(i)$ variable in some kind of protected memory, e.g. such that this variable can be altered only by the algorithm — not by a change from outside of the system.

The ASFC II algorithm is given as Alg. 2.

Algorithm 2: The ASFC II algorithm

1.a): **if** $q(i) = \textbf{\textit{Stationary}} \wedge c(i) = \emptyset \wedge (s(i) \neq nt(i) \vee fc(i) \neq nt(i))$ **then**
  $q(i) :=$ **NeedUpdate**
1.b): **if** $q(i) = \textbf{\textit{Stationary}} \wedge c(i) \neq \emptyset \wedge (s(i) \neq nt(i) \triangle su(i) \vee fc(i) \neq su(i))$ **then**
  $q(i) :=$ **NeedUpdate**
1.c): **if** $q(i) = \textbf{\textit{Stationary}} \wedge q(p(i)) = \textbf{\textit{NeedUpdate}}$ **then**
  $q(i) :=$ **NeedUpdate**
1.d): **if** $q(i) = \textbf{\textit{Stationary}} \wedge c(i) \neq \emptyset \wedge \exists_{k \in c(i)}(q(k) \in \{\textbf{\textit{NeedUpdate}}, \textbf{\textit{Updated}}\})$ **then**
  $q(i) :=$ **NeedUpdate**
2.a): **if** $q(i) = \textbf{\textit{NeedUpdate}} \wedge c(i) = \emptyset$ **then**
  $q(i) :=$ **Updated**
  $s(i) := nt(i)$
  $fc(i) := nt(i)$
2.b): **if**
$q(i) = \textbf{\textit{NeedUpdate}} \wedge c(i) \neq \emptyset \wedge \forall_{k \in c(i)}(q(k) = \textbf{\textit{Updated}})$
then
  $q(i) :=$ **Updated**
  $s(i) := nt(i) \triangle su(i)$
  $fc(i) := su(i)$
3.a): **if**
$q(i) = \textbf{\textit{Updated}} \wedge p(i) = null \wedge \forall_{k \in c(i)}(q(k) = \textbf{\textit{Updated}})$
then
  $q(i) :=$ **NearStationary**
3.b): **if** $q(i) = \textbf{\textit{Updated}} \wedge q(p(i)) = \textbf{\textit{NearStationary}}$ **then**
  $q(i) :=$ **NearStationary**
4): **if**
$q(i) = \textbf{\textit{NearStationary}} \wedge \forall_{k \in c(i)}(q(k) = \textbf{\textit{Stationary}})$ **then**
  $q(i) :=$ **Stationary**

An illustration of the algorithm rules is presented in Fig. 2, where only tree edges are shown. The Fig. 3 presents an example execution of the algorithm.

IV. Convergence and complexity

We will show that the system stabilizes after exactly $4n$ moves if there were any faults regardless of their number.

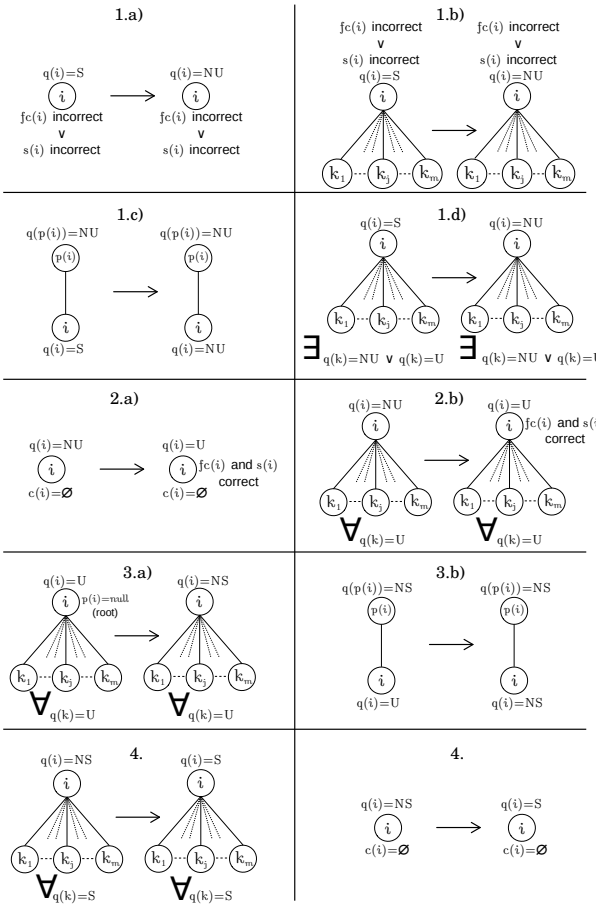


Fig. 2. An illustration of the ASFC II algorithm rules, where $k \in c(i)$.

All the time we assume the precondition $\forall_{i \in V(G)} q(i) =$ **Stationary** is fulfilled after any transient fault has occurred and a DFST of the graph is also given.

Lemma 1. *Each node i changes its state $q(i)$ cyclically from* ***Stationary*** *through* ***NeedUpdate***, ***Updated***, ***NearStationary***, *back to* ***Stationary***.

Proof: It follows from the rules of the algorithm. Only the rules 1.a)–1.d) change the state from $q(i) =$ **Stationary** and set it into state $q(i) =$ **NeedUpdate**. Next, only rules 2.a) and 2.b) change the state from $q(i) =$ **NeedUpdate** and set it into state $q(i) =$ **Updated**. Analogously, there are rules 3.a) and 3.b) changing $q(i) =$ **Updated** only into $q(i) =$ **NearStationary** and the rule 4. changing $q(i) =$ **NearStationary** into $q(i) =$ **Stationary**. ■

Lemma 2. *If a node i changes its state $q(i)$ from* ***NeedUpdate*** *to* ***Updated***, *then every its descendant has correct state of $s(i)$, $fc(i)$ and $q(i) =$* ***Updated***.

Proof: According to the rule 2.b), the non-leaf node i can change its state $q(i)$ from **NeedUpdate** to **Updated** only if every its child has the state $q(i) =$ **Updated**. For every child we can repeat our reasoning as for their parent. Going this way down to the leaves, we can see that every descendant has been in the state $q(i) =$ **Updated** at least for a while between the node i was **Stationary** and became **Updated**.

To end the proof it is now sufficient to show that a node cannot switch itself from $q(i) =$ **Updated** if its parent also has state $q(p(i)) =$ **Updated**. In other words, the node is inactive until its parent's state switches into $q(p(i)) =$ **NearStationary**. It follows from the fact that rule 3.b) is the unique rule that could change the state of the node i. But this rule cannot be active, since $q(p(i)) =$ **Updated**.

From the fact that values of $s(i)$ and $fc(i)$ are set during switching to state $q(i) =$ **Updated**, and this state propagates from leaves to the up, follows that $s(i)$ and $fc(i)$ are correct. ■

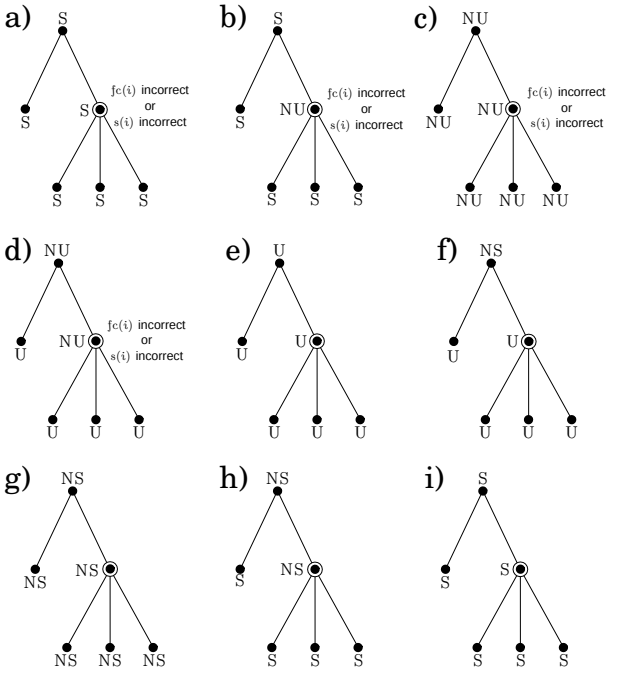


Fig. 3. An example of execution of the ASFC II algorithm:
a) there is one node i (circled in the figure) with incorrect value of $fc(i)$ or $s(i)$,
b) its state $q(i)$ changes to **NeedUpdate** (rule 1.b),
c) its ancestors (rule 1.d) and descendants (rule 1.c) also get the **NeedUpdate** state,
d) leaves get the **Updated** state (rule 2.a),
e) and it propagates through internal nodes up to the root (rule 2.b),
f) then root gets **NearStationary** state (rule 3.a),
g) which propagates down to the leaves (rule 3.b),
h) then leaves get **Stationary** (rule 4),
i) and it propagates up to the root, now system is stabilized (rule 4).

From Lemma 2 we have that if the root node changes its state to **Updated**, every other node i in the tree has updated its states $s(i)$ and $fc(i)$ to its proper values and has set the state $q(i)$ to **Updated**.

Lemma 3. *If a node is changing its state into $q(i) =$* ***NearStationary***, *then every its ascendant also has state* ***NearStationary***.

Proof: A node i must have parent's state

$q(p(i)) =$ **NearStationary**, to change its state into $q(i) =$ **NearStationary** (rule 3.b). Repeating this reasoning inductively, we have that every ascendant had state **NearStationary** at least in the past.

Now it is sufficient to show, that it could not have changed it. For change from **NearStationary** into **Stationary**, every child of a node has to be also in the **Stationary** state (rule 4). Again, repeating this reasoning down to the node i we have contrary to the fact that the node i is right about to change into state **NearStationary** from **Updated**.

■

In particular by Lemma 3 we have that a leaf gets into **NearStationary** state if every its ascendant also has **NearStationary** state. Rule 4 makes a stoppage — a node has to wait for all its children to change from **NearStationary** into **Stationary**, before it changes its state into **Stationary**. By induction we have that it holds not only for children, but for all descendants.

Once a node goes through whole sequence of changing of the own state from **Stationary**, through **NeedUpdate**, **Updated**, **NearStationary**, back to **Stationary**, it has correct values of $fc(i)$ and $s(i)$. From the Lemmas 1–3 we have that after a fault has occurred in at least one node, every node in the system has to go through all the sequence of changing. So we have the following theorem:

Theorem 2. *If there is DFST given in a system, every node has state* ***Stationary*** *and there is any number of nodes with incorrect value of fc or s, the ASFC II algorithm stabilizes the system after $4n$ moves.*

V. Conclusion

We have presented the modification of Chaudhuri's algorithm for finding a set of fundamental cycles in a graph. It demands particular precondition on the state of a system, but thanks to this, it offers linear time of the stabilization. No matter how many faults have occurred in the system, the ASFC II algorithm always performs $4n$ moves.

Motivation for the problem, apart from theoretical point of view, is applicable in computer networks. If we model a computer network as a graph, then the number of cycles passing through a node is a measure of its reliability. The more cycles go through a node, the more link failures (incident to the node) can occur whereas the graph is still connected.

References

[1] Dijkstra E. W., Self-stabilizing in spite of distributed control, *Communications of the ACM*, 17 (1974), 643–644.
[2] Schneider M., Self-Stabilization, *ACM Computing Surveys*, Vol. 25, No 1, March 1993.
[3] Dolev S., Self-stabilization, The MIT Press, 2000.
[4] Harary F., Graph Theory, Addison-Wesley, 1972.
[5] Chaudhuri P., A Self-Stabilizing Algorithm for Detecting Fundamental Cycles in a Graph, *Journal of Computer and System Sciences* 59, (1999) 84–93.
[6] Collin Z., Dolev S., Self-stabilizing depth-first search, *Information Processing Letters* 49 (1994), 297–301.
[7] Arora A., Gouda M. G., Closure and convergence: A foundation for fault-tolerant computing, *Proceedings of the 22nd International Conference on Fault-Tolerant Computing Systems* 1992.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 299–305

Anticipation in the Dial-a-Ride Problem: an introduction to the robustness

Samuel Deleplanque
Blaise Pascal University
LIMOS CNRS Laboratory
LABEX IMOBS3
Clermont-Ferrand 63000, France
Email: deleplan@isima.fr

Jean-Pierre Derutin
Blaise Pascal University
Institut Pascal CNRS Laboratory
LABEX IMOBS3
Clermont-Ferrand 63000, France
Email: derutin@univ-bpclermont.fr

Alain Quilliot
Blaise Pascal University
LIMOS CNRS Laboratory
LABEX IMOBS3
Clermont-Ferrand 63000, France
Email: quilliot@isima.fr

***Abstract*—The Dial-a-Ride Problem (DARP) models an operation research problem related to the on demand transport. This paper introduces one of the fundamental features of this type of transport: the robustness. This paper solves the Dial-a-Ride Problem by integrating a measure of insertion capacity called Insertability. The technique used is a greedy insertion algorithm based on time constraint propagation (time windows, maximum ride time and maximum route time). In the present work, we integrate a new way to measure the impact of each insertion on the other not inserted demands. We propose its calculation, study its behavior, discuss the transition to dynamic context and present a way to make the system more robust.**

I. Introduction

TODAY, the Dial-a-Ride Problems are used in transportation services for elderly or disabled people. Also, the recent evolution in the transport field such as connected cars, autonomous transportation, and the emergence of the shared service might need to use this type of problem at much larger scales. But this type of transport is expensive and the management of the vehicles requires as much efficiency as possible, however the number of requests included in the vehicles planning can vary depending on the resolution used.

In [1] we solve the DARP by using constraint propagation in a greedy insertion heuristic. This technique obtains good results, especially in a reactive context, and is easily adaptable to a dynamic context. But, each demand is inserted one after another and the process doesn't take into account the impact of each insertion on the other not inserted demands, and so, in a dynamic context, the future demands. In this work, we present a measure of an insertion capacity named *Insertability*. We introduce its calculation by integrating the impact of an insertion on the time constraints (time windows, maximum route time and maximum ride time).

This measure may be used in different ways: selection of the demand to insert, selection of the insertion parameters, and exclusion of a demand. These three uses may be related to static as well as dynamic contexts by anticipating the future demands. The goal is to insert the current demand in order to build flexible routes for the future ones.

This paper is organized in the following manner: after a literature review, the next section will propose a model of the classic DARP. Then, we will review how to handle the temporal constraints with a heuristic solution based on insertion techniques using propagation constraints. We will continue by explaining the way to measure the *Insertability*, a calculation based on the evolution of the time windows after an insertion. Then, we will give some uses of this measure including making an appointment which minimize the time windows (cf. Figure 1). In the last part of the paper, the computational results will show the efficiency of our *Insertability*'s measure and we will report the evolution of the number of demands inserted in a resolution of some instances' sets.

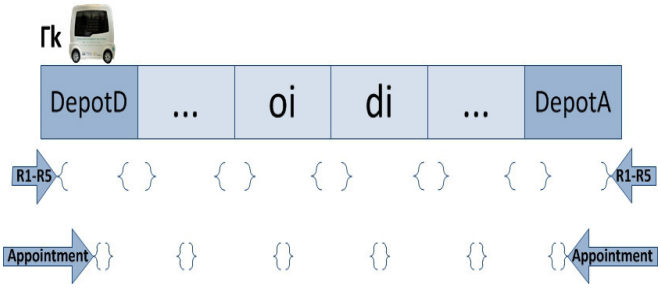


Fig. 1. Times windows' contraction

II. Literature review

The first works of the transportation optimization problem are related to the Traveling Salesman Problem ([2]). Since that time, other transportation problems have emerged as the vehicle routing and scheduling problems, and the Pick-up and Delivery Problem (PDP). The PDP is the ancestor of the problem of the Dial-a-ride problem which has been studied since the 1970's. DARP can be modeled in different ways. There are a number of integer linear programmings [3], but the problem complexity is too high to use it in a real context, most of which are NP-Hard because it also generalizes the Traveling Salesman Problem with Time Windows (TSPTW). Therefore, the problem must be handled through heuristic techniques. [4] is an important work on the subject and uses the Tabu search to solve it. Other techniques work well like dynamic programming (e.g. [5] and [6]) or variable neighborhood

978-83-60810-53-8/$25.00

searches (VNS) (e.g. [7] and [8]). Moreover, a basic feature of DARP is that it usually derives from a dynamic context. So, algorithms for static DARP should be designed in order to take into account the fact that they will have to be adapted to dynamic and reactive contexts, which means synchronization mechanisms, interactions between the users and the vehicles, and uncertainty about for-coming demands. [9], and [10] later, developed the most used technique in dynamic context or in a real exploitation is heuristics based on insertion techniques. These techniques are a good solution when the people's requests have to be taken into account in a short period of time.

III. The Dial-a-Ride Problem: model and insertion greedy algorithm

A. The general notations

This section lets to set notations used throughout this document. For any sequence (or list) Γ_k we set:

- for any z in Γ_k :
  - Succ(Γ_k, z) = Successor of z in Γ_k ;
  - Pred(Γ_k, z) = Predecessor of z in Γ_k ;
- for any z, z' in Γ_k :
  - $z \ll_k z'$ if z is located before z' in Γ_k ;
  - $z \ll_k^= z'$ if $z \ll_k z'$ or z = z'.

B. The model

A Dial a Ride Problem instance is defined by a *Demand* set $D = (D_i, i \in I)$, a fleet of *K* vehicles with a common capacity CAP, and a transit network $G = (V, E)$. *V* contains some specific node *Depot* and demands' nodes (DepotD for the departure and DepotA for the arrival). Each arc $e \in E$ is endowed with riding times give by a distance function $DIST(e)$. Each demand includes o_i an *origin* node, d_i a destination node, $F(o_i)$ and $F(d_i)$ two time windows, Δ_i a maximum ride time and Q_i a description of the load such that $Q_i = q_{o_i} = -q_{d_i}$ with q the load related to a node. Finally, the total time of the *K* vehicles planning are limited by $\Delta^k, k \in K$.

Solving a DARP with such an instance means creating a scheduling for each vehicle handling demands of D. The routes are constructed while optimizing a performance, which could be a mix of costs (e.g. total distance) and QoS criteria (e.g. ride time).

C. A greedy insertion algorithm: the insertion mechanism

In [1], we present an insertion greedy algorithm based on constraint propagation in order to contract time windows according to the time constraints. An insertion which does not imply constraint violation is said *valid* if $\Gamma = \cup_{k \in K} \Gamma_k$, the resultant collection of routes, if *load-valid* and *time-valid*. A route is *load-valid* if the capacity is not exceed, so, the *load-validity* is obtained if $ChT_k(x) \leq CAP$ with $ChT_k(x) = \sum_{y \ll_k^= x} q_y$, x and y nodes in the route *k*. The *time-validity* is obtained if there is no violation of the time constraints modeling by, for each demand *i*, $i \in D$, Δ_i the maximum ride time, $\Delta^k, k \in K$ the maximum route time and the constraints modeled by each time window $F(o_i) = [F.min(o_i), F.max(o_i)]$ and $F(d_i) = [F.min(d_i), F.max(d_i)]$. Checking the *load-validity* on $\Gamma = \cup_{k \in K} \Gamma_k$ is easy, and we show the efficiency of the constraint propagation in order to prove to *time-validity* after each planned insertion once the *load-validity* is proved. According to a current time window set *FP* = {*FP(x) = [FP.min(x), FP.max(x)]*, $x \in \Gamma_k, k = 1..K$} the *time-validity* may be performed through propagation of the five following inference rules **Ri**, i = 1..5 in a given route Γ_k:

for each *(x,y)* pair of nodes such that y is the successor of x:

- **R1** : $FP.min(x) + DIST(x,y) > FP.min(y) \models (FP.min(y) \leftarrow FP.min(x) + DIST(x,y))$,
- **R2** : $FP.max(y) - DIST(x,y) < FP.max(x) \models (FP.max(x) \leftarrow FP.max(y) - DIST(x,y))$;

for each *(x,y)* pair of nodes such that both are related to the same demand, one is the *origin* so the other the *destination* :

- **R3** : $FP.min(x) < FP.min(y) - \Delta(x) \models (FP.min(x) \leftarrow FP.min(y) - \Delta(x))$,
- **R4** : $FP.max(y) > FP.max(x) + \Delta(x) \models (FP.max(y) \leftarrow FP.max(x) + \Delta(x))$;

and for each x, $x \in \Gamma_k, k = 1..K$:

- **R5** : $FP.min(x) > FP.max(x) \models REJET \leftarrow true.$

These 5 rules are propagated in a loop while there no time windows exists *FP* modified at the last iteration. The tour Γ_k, $k = 1..K$, is *time-valid* according to the input time window set *FP* if and only if the *REJET* Boolean value is equal to *false* as initialized at the beginning of the process. In such a case, any *valid- time* value set *t* related to Γ_k and *FP* is such that: for any *x* in Γ_k, *t(x)* is the appointment's date in *FP(x)*.

The greedy insertion algorithm includes this propagation constraint technique in order to evaluate each possible insertion. Each iteration of the algorithm selects one demand according to the number of vehicle able to integrate it. Once a demand is selected, the process chooses the insertion's parameters that are the vehicle and the location of the *origin* and *destination* nodes.

IV. *Insertability* optimization

A. State of the system

In the above algorithm, each iteration selects a demand, and then, it finds the way to insert while optimizing the performance. This greedy algorithm doesn't take in account the impact of this actual insertion on the future demands integration, but only the effect on the demands already inserted. In this section, we introduce a *Insertability* calculation

by integrating this impact of an insertion related to the time constraints (time windows, maximum ride time and maximum route time).

During the insertion process, the state of the system is given by:

- a set of demands $D-D1$ already integrated in the routes, and $D1$ is the set of demands not inserted,
- a collection $\Gamma = \cup_{k\in K}\Gamma_k$ of routes including a list of nodes related to the *Depot*, *origin* and *destination* nodes,
- a exhaustive list of insertion's parameters sets. Each set gathers 5 elements : k the vehicle, *i* the demand, (x, y) the pair of insertion nodes (locating respectively o_i between x and the successor of x, and d_i between y and the successor of y), and *v* the evolution of the collection $\Gamma = \cup_{k\in K}\Gamma_k$'s cost.

B. Insertion's parameters

Given that the difficulty of the instances' problem is linked to the time constraints, we introduce an *Insertability* calculation related to the times windows contractions. During an insertion's assessment, these reductions appear once the inference rules are propagated. Here, we try to find a good triple (k, x, y), the vehicle and the location of the *origin/destination* nodes, in order to give enough space to the future demands (which have to be integrated in $\Gamma = \cup_{k\in K}\Gamma_k$).

We set *INSER(i,* Γ *)* the *Insertability* measure of the demand *I*. The quantity $U_n^k(z)$ denotes the vehicle *k* time windows' amplitude of the node *n* once it has been inserted to the right of node z. *INSER* is calculated as follows:

- $INSER(i,\Gamma) = \sum_{k\in K} INSER1(i,\Gamma_k)$;
- $INSER1(i,\gamma) = Max_{(x,y)}INSER2(i,\gamma,x,y)$, γ a tour of Γ ;
- $INSER2(i,\gamma,x,y) = U^{\gamma}_{o_i}(x).U^{\gamma}_{d_i}(y)$.

$INSER1$ gives us the maximum of the product of the time's windows amplitude at the origin i and destination i over the possible insertion positions x and y in the route γ. When *INSER1* is equal to 0, the new route γ resulting to the new insertion isn't *time-valid*.

We set $Inserted(\Gamma, i_0, k, x, y)$ the updated collection of tours Γ with the insertion of the selected demand i_0 at the locations x and y in the vehicle k. The *INSER(i,* Γ*)* measure allows us to write the *Optimization Insertability Problem* which consists to find the best insertion parameters in order to keep the vehicles' scheduling more flexible:

Optimization Insertability Problem. Find the optimal parameters *(k,x,y)* inserting i_0 and maximizing the value $Min_{i\in D1-i_0}INSER(i, Inserted(\Gamma, i_0, k, x, y))$.

For instance, the value $Min_{i\in D1-i_0}INSER(i, Inserted(\Gamma, i_0, k, x, y))$ may be used if all the demands have to be inserted. Another optimization may be process as the maximization of the sum $\sum_{i\in D1-i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y))$. The choice is made according to the homogeneity of the demands and if the problem requires to insert all the set *D*.

This problem only optimizes the variation of the *Insertability* values and doesn't include other performance criteria like the minimization of the ride times, waiting times or distances. The *Insertability* criterion can be integrated in a mix of economical cost (point of view of the fleet manager) and of QoS criteria (point of view of the users). Then, the process maximizes the function $Perf = \mu.\sum_{i\in D1-i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y)) - v(Inserted(\Gamma, i_0, k, x, y))$ with μ a criterion coefficient and *v* the performance value function mixing the costs related to the both points of view.

C. Other uses of the Insertability measure

So far, we select the demand i_0 according to the number of vehicles available (taking in account all the time and load constraints). The *Insertability* measure *INSER(*i_0, Γ*)* may be also used in order to select the next request i_1 to insert. This application could be used in a context where all the demands of D have to be integrated. The selection is based on the smallest *Insertability* measure. Once a demand is selected, the problem may solve the *Optimization Insertability Problem.* Here, the two steps may be written in a non-deterministic way. The demand may be selected randomly through a set of N1 elements with the smallest *INSER* value. The same scheme may be applied on a set of a insertion parameters of N2 elements with a best (k, x, y) elements maximizing the quantity $Min_{i\in D1-i_0}INSER(i, Inserted(\Gamma, i_0, k, x, y))$.

Also, *INSER*(i_0, Γ*)* may be useful for a larger set D. If the instance doesn't have any solution integrating all the set D, it is preferable to identify requests to exclude as soon as possible. The exclusion of a demand i_0 may be set up if its insertion results in Γ not enough flexible to include the other elements of $D1$. In other words, the demands excluded will be those that will have the most impact of future insertions. The difference $\sum_{i\in D1-i_0}(INSER(i,\Gamma) - INSER(i, Inserted(\Gamma, i_0, k, x, y)))$ of the inequality (1) takes in account the *Inserabilty* measure of $D1-i_0$ before and after the insertion of i_0 in the routes of Γ. If this difference is larger than the threshold ξ, the demand is excluded. In the experimentation' section, we will discuss the fact this threshold should be dynamic and decreases over the execution.

$$\sum_{i\in D1-i_0} (INSER(i,\Gamma) - INSER(i, Inserted(\Gamma, i_0, k, x, y))) > \xi \quad (1)$$

D. The Insertability optimization suited to the greedy insertion algorithm

The calculation of *INSER(i, Γ)*, $i \in D$, begins to be time consuming starting from a medium size of *D* once the *INSER2* value is based on the time windows' amplitude obtained after the propagation of the time constraints. So, this is important to spot each step of the process where the *Insertability* measure doesn't have to be updated. When i_0 is selected, $INSER2(i, \Gamma_k, x, y)$, $INSER1(i, \Gamma_k)$ and $INSER(i, \Gamma)$ are known for all demand in $D1 - i_0$ and all $k = 1..K$. Once i_0 is about to be inserted, the process computed the value $H(i), i \in D1 - i_0$ (cf. formulation (2)). Then, the algorithm tries the insertion of each i from $D1 - i_0$ in $Inserted(\Gamma, i_0, k, x, y)$ and deduce the value *K(i)* given in formula (3) for all $i \in D1 - i_0$ and ultimately the quantity $Val(k, x, y) = Min_{i \in D1 - i_0}(K(i) + H(i))$.

$$H(i) = INSER(i, \Gamma) - INSER1(i, \Gamma_k) \tag{2}$$

$$\begin{aligned} K(i) &= INSER(i, Inserted(\Gamma, i_0, k, x, y)) \\ &= H(i) + INSER1(i, Inserted(\Gamma, i_0, k, x, y)_k) \end{aligned} \tag{3}$$

Other calculations may be avoided. We set W_1 such that $W_1 = Min_{i \in D1 i_0} INSER\ (i, \Gamma)$. If the quantity $INSER(i, \Gamma) - INSER1(i, \Gamma_k)$ is larger than W_1, there is no need to test the impact of the insertion of i_0 on *i*.

Finally, we're able to use *INSER(i, Γ)* once we integrate the future demands presented in the next section. In a dynamic context, the *Insertability* measure helps the routes to be enough flexible for the next insertion process. Moreover, the appointments have to be set with the same purpose and *INSER(i, Γ)* is able to help to do it.

V. Introduction to the robustness in the DARP: Anticipation of the future demands and *Insertability* measure integration

The problem may have to be handled according to a *dynamic* context and the greedy insertion algorithm is easily adaptable to this context. Once the *Insertability* measure is included in the performance criteria, the system may increase its robustness. In order to accomplish this, we need to exploit knowledge about future demands. In our case, this knowledge is related to the type of on demand transportation service. In this paper, we will use a simple extrapolation of this probable demand based on the demand already broadcasted.

We won't take into account the way the system supervises its various communication components with the users. In reality, there are eventual divergences between the data which were used during the planning phases and the situation of the system.

We set $D-V$ the virtual demands, $D-R$ the real demands, and $D-Rejet$ the set of the ones excluded from the insertion algorithm such that $D-Rejet = DV-Rejet \cup DR-Rejet$. The $D-V$ formulation is given in (4). p_i gives us the number of times the demand D_i, $i \in D$, will appear for each period of each discrete planning horizon.

$$D - V = \sum_{i \in D} D_i.p_i \tag{4}$$

Then, we're able to update the formula (5) giving the performance function *Perf*.

$$\begin{aligned} Perf = \alpha. \sum_i p_i INSER(i, Inserted(\Gamma, i_0, k, x, y)) \\ +\mu. \sum_{i \in D1 - i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y)) \\ -v(Inserted(\Gamma, i_0, k, x, y)) \end{aligned} \tag{5}$$

As in the previous sections, the process may exclude some demands taking in account the future requests. We updated the inequality (1) by the (6). α is a coefficient based on the importance of the future demands.

$$\begin{aligned} \alpha. \sum_i p_i.(INSER(i, \Gamma) \\ -INSER(i, Inserted(\Gamma, i_0, k, x, y))) \\ + \sum_{i \in D1 - i_0} (INSER(i, \Gamma) \\ -INSER(i, Inserted(\Gamma, i_0, k, x, y))) > \xi \end{aligned} \tag{6}$$

VI. Discussion about the appointments and the dynamic context

Most works on vehicle scheduling problems including time window studies how to integrate a set of demands in the vehicle planning. Making an appointment anticipating the future is especially rare. Previous sections explained how to select and integrate user's request while keeping enough space for the next set of demands.

Once routes are built and integrated a first set *D*, the users expect the date when the vehicle selected will pick them up. In the lists forming the *K* routes, each node has a time window. After the appointment's date is set, each time window becomes tight with zero amplitude or equals a very small delay. How the appointments' dates are made is very important for the next insertion's process. For instance, we consider a fleet of 2 vehicles with two plannings including 5 demands while the distances are minimized (cf. Figure VI). The time windows are relatively wide so, while the distance traveled is minimized, the difference of each appointment's time between two nodes is the exact time to join them. The vehicle *k=2* from the Figure VI may integrated the node o_7 between its depot node and o_5 even if its time windows have a zero amplitude (the vehicle will only have to leave the depot earlier). On the other hand, if the difference on the appointment' times given to the users related to the nodes d_5 and o_3 equals to $DIST(d_5, o_3)$, the insertion of d_7 will be forbidden. In the same way, there will be a violation of some constraint once nodes o_6 and d_6 will be inserted in the vehicle $k = 1$.

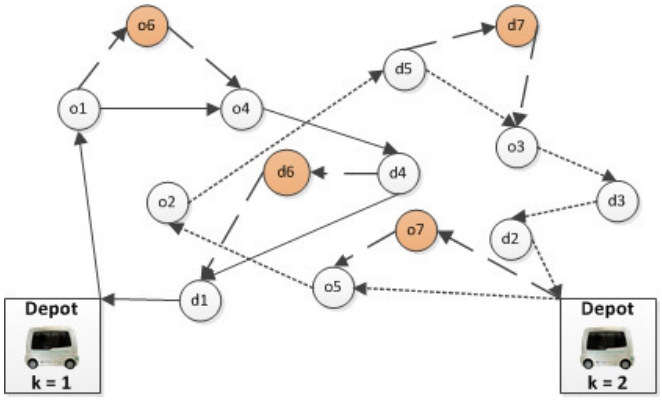


Fig. 2. New insertions after the first set of appointments

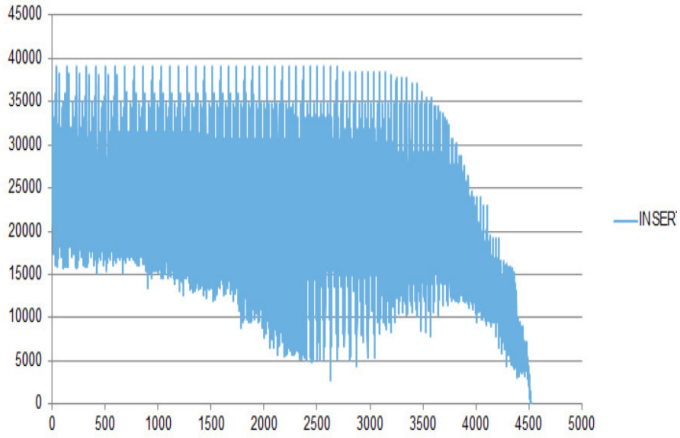


Fig. 3. INSER values on the not inserted demands

One more time, the *INSER(i, Γ)* values may be used in order to set the appointment dates without to have the problem above. The appointment's dates may be calculated once the process have inserted the virtual demands $D-V$ and the real demands $D-R$.

The previous section shows the way to anticipate the future demands $D-V$. These demands are related to a dynamic context. Note again that our greedy algorithm is easily adaptable to this context. More specifically, the technique doesn't change unlike the state of each route. The first node isn't a depot node anymore but a dynamic node related to the vehicle's location. The entire constraint propagation process is applied on these new routes. A simulation will be necessary to evaluate the anticipation of the future demands including in the dynamic context.

VII. Computational Experiments

In this section, we study the behavior of our *Insertability* measure used in the resolution of Dial-a-Ride instances. The algorithms were implemented in C++ and compiled with GCC 4.2. In [1], we solve the [4]'s instances by our greedy insertion algorithm based on constraint propagation. We obtained good results in the majority of instances, but, only 1% of the replications gave us a feasible solution on the tenth instance (*R10a*). The CPU time was smallest or equal to the best times in the literature; we don't work on this feature for this experiment.

A. First experimentation: the optimization of the selection of the demand to insert

1) INSER used in the selection of a demand: We note by R^{DARP} the rate of 100 replications which give us a feasible solution obtained by using the solution of [1]. Here, the selection of the demand is based on the lowest number of cars which are able to accept it. R^{DARP}_{Rob} is the rate obtained with the same process except that each demand is selected at each iteration by the lowest *Insertability* value *INSER*.

The *Insertability* measure is already efficient once it's used in the selection of the demands to insert. The rate obtained for the pr08, pr09, pr10 and pr19 are clearly more interesting as shown in table I (e.g. for the instance pr08, the rate increases by 56% to 91% of success).

Inst.	R^{DARP}	R^{DARP}_{Rob}
pr01	99	100
pr02	100	100
pr03	97	100
pr04	100	100
pr05	100	100
pr06	100	100
pr07	90	96
pr08	56	**91**
pr09	18	21
pr10	1	**7**
pr11	100	100
pr12	100	100
pr13	99	100
pr14	100	100
pr15	100	100
pr16	100	100
pr17	98	100
pr18	99	100
pr19	64	**99**
pr20	43	56
Av.	83,2	88.5

TABLE I
R^{DARP} VS $R^{DARPRob}$

2) INSER behaviour: Each time a replication can't integrate all the request, the *INSER* value of the demands not inserted has to be null. In Figure VII-A2, while the resolution process applied to the R10a instance, we note the evolution of more than 4500 *INSER*'s demands not inserted. The technique used is the second approach selecting the demand by the smallest *Insertability*. The values noted are from a failed replication.

One can observe big gaps between the different *INSER* until the 4000 first values. After that, for the remaining requests, the *Insertability* values decrease strongly because the routes begin to be not flexible. Between the 2500^{th} and the 3500^{th}, for some demands, the values are very low at the beginning just before increasing strongly. This is explained by the fact the process inserts the demand with the lowest

INSER but their insertion don't make a big impact on the other demands not inserted. This impact is related to the *Optimization Insertability Problem* studied below.

B. Second experimentation: the optimization of the insertion parameters

In a second experimentation, we compare the [1]'s approach and another algorithm based on the optimization of the parameters *(x,y,k)*. The selection of the request to insert is the same for both solutions. For the second one, once a demand i_0 is selected, we maximize the sum $\sum_{i \in D1-i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y))$ in order the find the best parameter *(x,y,k)* which will integrate i_0 in the route k. We don't optimize $Min_{i \in D1-i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y))$ because we create instances especially with a set *D* too large for inserting all the requests. So, the demand with the smallest value *INSER* for a given parameters *(x,y,k)* could never be integrated into the routes.

The two algorithms were applied to five sets of 5 randomly generated instances. All the instances have their time constraints related to the interval [0;400] and all the load was unit. We set by $e_{F(o)}$ and $e_{F(d)}$ the amplitude of the time windows at the *origin* and the *destination* given by the users, respectively. The other parameters are given in table II.

K	$e_{F(o)}$	$e_{F(d)}$	Δ	CAP
10	35	10	∞	10

TABLE II
PARAMETERS' INSTANCES

We generate 5 different sets of 5 instances with a variation of the number of demands $|D|$. We set by T_{Insert} and by $T_{Insert_{Rob}}$ the demand inserted's rate the first resolution and the second technique, respectively. Finally, Gap_{Insert} is the gap in percentage between each rate. Its calculation is given by $Gap_{Insert} = 100.(T_{Insert_{Rob}} - T_{Insert})/T_{Insert}$. We launched 100 replications of each technique on the 5 sets. The results are provided by the table III.

$\|D\|$	50	75	100	150	200
T_{Insert}	100	93.2	78.9	64.2	52.6
$T_{Insert_{Rob}}$	100	96.8	85.3	66.4	54.1
Gap_{Insert}	0	3.86	8.11	3.43	2.81

TABLE III
GAP BETWEEN THE *INSERT* RATES

In future experiments, we need to optimize the value $Perf = \mu.\sum_{i \in D1-i_0} INSER\ (i, Inserted(\Gamma, i_0, k, x, y)) - v(Inserted(\Gamma, i_0, k, x, y))$ to calculate each best insertion parameters. Here, we're just taken into account the *INSER* values in order to integrate the most requests possible. The results show us that the larger of $|D|$ defines if the system needs to optimize the *Insertability* measure. For $|D| = 50$, all the requests are able to be inserted easily, so, the *INSER* values doesn't have any interest. When the set is composed of 100 demands, we obtained a Gap_{Insert} of 8,11% meaning there are more than 8% more requests inserted by the second approach.

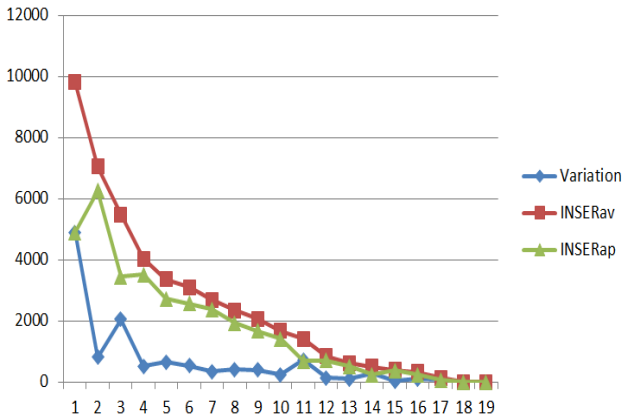


Fig. 4. Variation of the *Insertability* values between each insertion

For this set of instance, we also tried to integrate a new feature in our algorithm: we've added the ability to exclude a request if the impact of one insertion involving a significant drop of the general *Insertability's* demands from $D1 - i_0$. Before that, we study the threshold which limits the variation of *Insertability*.

We exclude a demand selected i_0 if $\sum_{i \in D1-i_0}(INSER(i, \Gamma) - INSER(i, Inserted(\Gamma, i_0, k, x, y))) > \xi$ is true with ξ a threshold. The calculation of the threshold is a difficult problem. In the figure VII-B, we report the $\sum_{i \in D1-i_0}(INSER(i, \Gamma) - INSER(i, Inserted(\Gamma, i_0, k, x, y)))$ *Variation* with *INSERav* and *INSERap* the values $\sum_{i \in D1-i_0} INSER(i, \Gamma)$ and $\sum_{i \in D1-i_0} INSER(i, Inserted(\Gamma, i_0, k, x, y))$, respectively. This figure shows us that the threshold ξ have to be calculated dynamically according to the average of *INSER*.

We used this type of dynamic threshold for the third set of instances with 100 demands. We exclude an request if the current ξ is exceeded, and only this feature is added in the second approach. We obtained a gain of 1,3% in average (from 85,3% to 86,6%) meaning approximately one more demand is able to be inserted.

VIII. CONCLUSION

The Dial-a-Ride Problem is one of the transport problems with the highest number of hard constraints like time windows. The insertion techniques are able to obtain a good solution in a reasonable time. Their adaptability to a dynamic context is easy but a lack of robustness could appear once the goal is to integrate requests as much as possible.

We have introduced a way to measure the impact of each insertion on the other demands not inserted. This *Insertability*

measure could be used in order to exclude a demand, to select a demand to insert and also to calculate the best insertion parameters. This value, named *INSER*, leads to a large amount of work opportunities. We have introduced a simple way to make the model of the future demands, and how to adapt our greedy insertion algorithm based on the constraint propagation to the dynamic context. In future work, we will develop a simulation which is necessary to show the efficiency of the demands anticipation. The final goal will be to develop the most robust algorithm possible in order to adapt it to a real context.

ACKNOWLEDGMENT

This work was founded by the French National Research Agency, the European Commission (Feder funds) and the Region Auvergne in the Framework of the LabEx IMobS3.

REFERENCES

[1] S. Deleplanque, A. Quilliot, Constraint Propagation for the Dial-a-Ride Problem with Split Loads, 2013, Recent Advances in Computational Optimization. Studies in Computational Intelligence, Vol. 470. ISBN 978-3-319-00409-9, Volume 470, 2013, Springer, 31-50.

[2] K. Menger, Das botenproblem, 1932, Ergebnisse eines mathematischenkolloquiums 2, 11-12.

[3] J.F. Cordeau, G. Laporte, The dial-a-ride problem: models and algorithms, 2007, Annals of Operations Research,153(1):29-46.

[4] J.-F. Cordeau, G. Laporte, A tabu search heuristic algorithm for the static multi-vehicle dial-a-ride problem, 2003, Transportation Research B 37, 579-594.

[5] H. Psaraftis, An exact algorithm for the single vehicle many-to-many dial-a-ride problem with time windows, 1983, Transportation Science 17, 351-357.

[6] R. Chevrier, P. Canalda, P. Chatonnay, D. Josselin, Comparison of three algorithms for solving the convergent demand responsive transportation problem, 2006, Intelligent Transportation Systems, Toronto, Canada, 1096-1101.

[7] S.N. Parragh, K.F. Doerner, R.F. Hartl, Variable neighborhood search for the dial-a-ride problem, 2010, Computers & Operations Research, 37, 1129-1138.

[8] P. Healy, R. Moll, A new extension of local search applied to the dial-a-ride problem, 1995, European Journal of Operational Research 83, 83-104.

[9] H. Psaraftis, N. Wilson, J. Jaw, A. Odoni, A heuristic algorithm for the multi-vehicle many-to-many advance request dial-a-ride problem, 1986, Transportation Research B 20B, 243-257.

[10] O. Madsen, H. Ravn, J. Rygaard, A heuristic algorithm for the a dial-a-ride problem with time windows, multiple capacities, and multiple objectives, 1995, Annals of Operations Research 60, 193-208.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 307–313

Multiple shooting SQP-line search algorithm for optimal control of pressure-constrained batch reactor

Paweł Drąg
Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
Janiszewskiego 11-17, 50-372 Wrocław, Poland
Email: pawel.drag@pwr.wroc.pl

Krystyn Styczeń
Institute of Computer Engineering, Control and Robotics,
Wrocław University of Technology
Janiszewskiego 11-17, 50-372 Wrocław, Poland
Email: krystyn.styczen@pwr.wroc.pl

***Abstract*—In the article a new approach for control of a pressure-constrained batch reactor and a new multi-step optimization algorithm were presented. The considered batch reactor was described by both differential and algebraic equations. State constraints incorporate always difficulties into a mathematical model of the reactor, so a new algorithm based on a multiple shooting SQP-line search method was proposed and tested. The multiple shooting method was used not only to ensure a stability of the solution, but to divide a system into smaller subsystems, so a large-scale problem is considered. The considerations were made for a simultaneous approach, which allows to apply this algorithm to a wide class of differential-algebraic systems. The simulations were executed in Matlab environment using Wroclaw Centre for Networking and Supercomputing.**

***Index Terms*—optimal control, DAE systems, multiple shooting method, state constraints.**

I. Introduction

SEARCHING for controls that will result in a desired behavior of a system plays a key role in a process design [1], [2], [11]. To describe the system often only algebraic equations are enough. Especially, when changes in the state variables are slow and algebraic equations accurately reflects the behavior of the system. More complex processes are described by differential equations. Suitable numerical methods and optimization algorithms were proposed and implemented, so it is a group of well-known problems . It seems, that the most important feature of the differential systems is a existence of a solution for all initial conditions. Difficulties may, however, be caused by the instability of the equations and selection of the appropriate numerical methods for the equations [5].

Often, however, it happens that there are in the system simultaneously both algebraic and differential relations. Description with the system of equations, which can be easily divided in part consisting solely of differential equations and a group of algebraic equations is desirable for several reasons. (1) During the construction of the mathematical model one does not need to perform additional transformations to obtain allowed equations. (2) The variables in a model are known to have physical interpretation. When the equations are well scaled, then no other transformations are needed. Additionally, (3) one can explore the impact of different variables on the behavior of the model. But the searching for the solution of the initial value problem for differential-algebraic equations, which does not exist for all possible values of parameters, was always a challenge [3], [8], [9].

The main motivation of this paper is to present an algorithm, which can treat the large-scale optimal control problem. Every system with path constraints on state trajectory can be considered as an optimization problem with arbitrarily large number of variables [2]. Even systems with simple path constraints, but with large number of decision variables, may require a huge computational effort. The aim of this paper is to present a feasible-type algorithm that improves a feasible initial solution of a large-scale problem in a reasonable time. These features enable the use of this approach in the task of design and control of chemical processes.

The pressure-constrained batch reactor is usually described by the nonlinear differential-algebraic equations [6]. The control problem of the chemical reactor belongs to the group of tasks, the size of which is not clearly defined. Especially if the model takes into account constraints on the state variables. The statement that the size of the task is infinitely large does not help much in solving the problem. The practical approach leads to the use of the existing finite-dimensional methods.

To solve the control problem of the chemical reactor with the constraints on state variables, the new multi-step algorithm was designed. Its particular advantage is the possibility to take into account the large number of variables and to preserve the feasibility of all iterates, which start from the feasible initial conditions. The study was carried out on the large-scale task of about 4 000 variables and 3 000 differential equations [4]. The algorithm combines the multiple shooting method and the simultaneous approach.

The article is structured as follows. At the beginning the optimal control problem of DAE system was formulated. Then the simultaneous approach for the optimal control of differential algebraic systems and its relationship with the multiple shooting method is discussed. The Multi-step SQP-line search algorithm using the multiple shooting method is presented. The differential-algebraic model of the pressure constrained batch reactor is described and solved by the designed algorithm. Finally, the results of the large-scale simulations, which were performed using Wroclaw Centre for Networking and Supercomputing, were discussed.

978-83-60810-53-8/$25.00

II. The simultaneous approach for multiple shooting optimal control of differential-algebraic systems

In the paper the following multiple shooting optimal control problem of differential-algebraic systems is considered

$$\min_{p} \phi(p) = \sum_{l=1}^{N_T} \Phi(z^l(t_l), y^l(t_l), p^l), \tag{1}$$

subject to

$$z^{l-1}(t_{l-1}) = z_0^l = 0; \qquad l = 2, \ldots, N_T, \tag{2}$$

$$z^{N_T}(t_{N_T}) - z_f = 0; \qquad z^l(0) = z_0^l, \tag{3}$$

$$p_L^l \leq p^l \leq p_U^l, \tag{4}$$

$$y_L^l \leq y^l(t_l) \leq y_U^l, \tag{5}$$

$$z_L^l \leq z^l(t_l) \leq z_U^l; \quad l = 1, \ldots, N_T, \tag{6}$$

with the DAE system

$$\frac{dz^l(t)}{dt} = f^l(z^l(t), y^l(t), p^l); \qquad z^l(t_{l-1}) = z_0^l, \tag{7}$$

$$g^l(z^l(t), y^l(t), p^l) = 0; t \in [t_{l-1}, t_l]; l = 1, \ldots, N_T. \tag{8}$$

In equations (1)-(8) $z(t)$ denotes the differential state trajectory and $y(t)$ denotes the algebraic state trajectory. The control profile is represented as a parametrized function with coefficients that determine the optimal profil [12], [13]. The decision variables on DAE equations appear only in the time independent vector p. The assumption on the invertibility of $g(-, y(t), -)$ permits an implicit elimination of the algebraic variables $y(t) = y[z(t), p]$ [3]. While there are N_T periods in DAE equations, the time dependent bounds and other path on the state variables are no longer considered. The algebraic constraints and terms in the objective function are applied only at the beginning of each period.

The mentioned optimal control problem of the reactor is an example of wide range control problems of systems described by differential-algebraic equations (eg. [11]). The instability of this type of equation resulted in development of shooting methods. The shooting method was adjusted for solving more difficult systems and is usually known as the multiple shooting method or the parallel shooting method. As a result of the application of the shooting methods, tested systems exhibit new properties, which could not be expected considering the general formulation of the optimal control problem of DAE systems. When the multiple shooting approach is used, the time domain is partitioned into smaller time periods and the DAE models are integrated separately in each element. To provide the continuity of the states across elements, the equality constraints are added to the nonlinear program. The inequality constraints for states and controls are then imposed directly at the grid points t_l [1].

The aim of the simultaneous approach is searching for the optimal control trajectory, the differential and algebraic state trajectories in a special manner. A sketch of the sequential

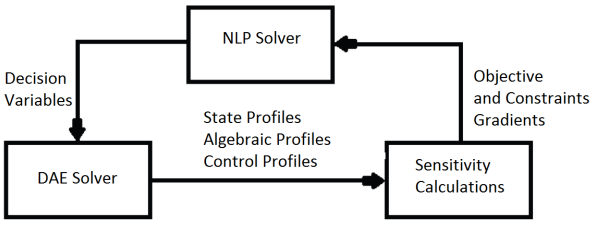


Fig. 1. Sequential dynamic optimization strategy.

dynamic optimization strategy for the problem (1)-(8) is presented on Fig. 1. At l-iteration, the variables p^l are specified by NLP solver. In this situation, when the values of p^l are known, one can treat DAE system as an initial value problem and integrate (2)-(4) forward in time for periods $l = 1, \ldots, N_T$. For these purposes Backward Differentiation Formula was used, which can solve index-1 DAEs. The Differential state profile, the algebraic state profile and the control function profile were obtained as results of this step. Next component evaluates the gradient of the objective and constraint functions with respect to p^l. Because function and gradient information are passed to the NLP solver, then the decision variables can be updated [2].

III. Dynamic optimization of the pressure-constrained batch reactor

The reactions taking place in the reactor are

$$A \xrightarrow{k_1} 2B, \tag{9}$$

$$2B \xrightarrow{k_2} A, \tag{10}$$

$$A + B \xrightarrow{k_3} D. \tag{11}$$

The dynamic optimization problem is described as follows

$$\min_{F} J = C_D(t_f), \tag{12}$$

subject to

$$\dot{C}_A = -k_1 C_A + k_2 C_B C_B + \frac{F}{V} - k_3 C_A C_B, \tag{13}$$

$$\dot{C}_B = k_1 C_A - k_2 C_B C_B - k_3 C_A C_B, \tag{14}$$

$$\dot{C}_D = k_3 C_A C_B, \tag{15}$$

$$N = V(C_A + C_B + C_D), \tag{16}$$

$$PV = NRT, \tag{17}$$

$$P \leq 340000, \tag{18}$$

$$0 \leq F \leq 8.5, \tag{19}$$

$$[C_A(0), C_B(0), C_D(0)] = [100, 0, 0]. \tag{20}$$

The rate constants are $k_1 = 0.8$ per h, $k_2 = 0.02\ m^3/(mol \cdot h)$, $k_3 = 0.003\ m^3/(mol \cdot h)$, the volume $V = 1.0\ m^3$, the temperature $T = 400\ K$. There is one path constraint on the state variable P. The process duration is 2 hours.

The task is to find the optimal flow rate profile, which is treated as a control variable, to minimize the objective function (12). There are some possibilities of parametrization of the control variable. The most popular are piecewise constant, piecewise linear with continuity, piecewise linear without continuity, piecewise quadratic with continuity [12]. In [6] the simultaneous approach with piecewise linear parametrization with continuity was considered. It means, that for 11 time intervals the size of NLP was originally 42, and 35 by model decomposition method presented in this article. Because "the optimal control is highly nonlinear which makes this problem difficult for general control parametrization methods" [6], the new control algorithm for the simultaneous approach with constant parametrization of the control variables was proposed and tested. So, there are decision variables connected only with the control function and the state variables.

IV. The Multi-step SQP line search algorithm

The designed algorithm belongs to a group of the Sequential Quadratic Programming methods [10]. Its main part is as follows.

The equality constrained problem is considered

$$\min_p f(p), \tag{21}$$

subject to

$$c(p) = 0, \tag{22}$$

where the objective function $f : \mathcal{R}^n \to \mathcal{R}$ and the vector of equality constraints $c : \mathcal{R}^n \to \mathcal{R}^m$ are smooth functions. The idea behind the SQP approach is to model (21)-(22) at the current iterate p_k by a quadratic programming subproblem. Then the subproblem is minimized and the new iterate p_{k+1} is defined.

The Lagrangian function for this problem is

$$\mathcal{L}(p, \lambda) = f(p) - \lambda^T c(p). \tag{23}$$

The matrix $A(p)$ were used to denote the Jacobian matrix of the constraints

$$A(p) = [\nabla c_1(p), \nabla c_2(p), \ldots, \nabla c_m(p)]^T, \tag{24}$$

where $c_i(p)$ is the ith component of the vector $c(p)$.

The first order KKT conditions of the equality constrained problem (21)-(22) can be written as the system on $n + m$ equations and the $n + m$ unknowns p and λ,

$$\begin{bmatrix} \nabla f(p) - A(p)^T \lambda \\ c(p) \end{bmatrix} = 0. \tag{25}$$

Any solution (p^*, λ^*) of the equality constrained problem (21)-(22) for which $A(p^*)$ has full row rank satisfies (25). The nonlinear system (25) can be solved by the Newton method.

The Jacobian of (25) with respect to p and λ is given by

$$\begin{bmatrix} \nabla^2_{pp}\mathcal{L}(p, \lambda) & -A(p)^T \\ A(p) & 0 \end{bmatrix} = 0. \tag{26}$$

The Newton step from the iterate (p_k, λ_k) is given by

$$\begin{bmatrix} p_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} d_k \\ d_\lambda \end{bmatrix}, \tag{27}$$

where d_k and d_λ solve the Newton-KKT system

$$\begin{bmatrix} \nabla^2_{pp}\mathcal{L}(p, \lambda) & -A(p)^T \\ A(p) & 0 \end{bmatrix} = \begin{bmatrix} d_k \\ d_\lambda \end{bmatrix} + \tag{28}$$

$$+ \begin{bmatrix} -\nabla f(p) + A(p)^T \lambda \\ -c(p) \end{bmatrix},$$

The Newton step is well defined when KKT matrix in (26) is nonsingular. This is satisfied, when the following assumptions hold [10]

Assumption 1: The Jacobian of the constraints $A(p)$ has full row rank.

Assumption 2: The matrix $\nabla^2_{pp}\mathcal{L}(p, \lambda)$ is positive definite on the tangent space of the constraints, that is, $d^T \nabla^2_{pp}\mathcal{L}(p, \lambda)d > 0$ for all $d \neq 0$ such that $A(p)d = 0$.

Suppose that at the iterate (p_k, λ_k) the problem (21)-(22) is modeled by the quadratic program

$$\min_p f_k + \nabla f_k^T p + \frac{1}{2}\nabla^2_{pp}\mathcal{L}_k p, \tag{29}$$

subject to

$$A_k(p) + c_k = 0. \tag{30}$$

If Assumptions 1 and 2 hold, then this problem has the unique solution (d_k, l_k) that satisfies

$$\nabla^2_{pp}\mathcal{L}_k d_k + \nabla f_k - A_k^T l_k = 0, \tag{31}$$

$$A_k d_k + c_k = 0. \tag{32}$$

The vectors d_k and l_k can be identified with the solution of the Newton equation (28).

ALGORITHM 1. Local SQP Algorithm for solving the equality constrained problem

Choose an initial par (p_0, λ_0);
  (if p_0 is given, then λ_0 is given by eq. (25))
Set $k \leftarrow 0$;
REPEAT UNTIL convergence test is satisfied
  evaluate f_k, ∇f_k, $\nabla^2_{pp}\mathcal{L}_k$, c_k, A_k;
  solve (29)-(30) to obtain d_k and l_k;
  set $p_{k+1} \leftarrow p_k + d_k$; $\lambda_{k+1} \leftarrow l_k$;
END (REPEAT)

On this basis, the new algorithm was designed.

ALGORITHM 2. The line search SQP algorithm

choose parameters $\eta \in (0, 0.5)$, $\tau \in (0, 1)$
and an initial pair (p_0, λ_0);
evaluate $f(p_0)$, $\nabla f(p_0)$, $c_i(p_0)$,
$A_0 = [\nabla c_1(p_0), \nabla c_2(p_0), \ldots, \nabla c_m(p_0)]^T$;
if a quasi-Newton approximation is used, choose
an initial $n \times n$ symmetric positive definite Hessian
approximation B_0, otherwise compute $\nabla^2_{pp}\mathcal{L}_0$;
WHILE convergence test is not satisfied DO
  compute d_k by solving (28), let λ be
  the corresponding multiplier;
  $d_\lambda \leftarrow \hat{\lambda} - \lambda_k$;
  choose μ_k to satisfy eq. (33) with $\sigma = 1$;
  set $\alpha_k \leftarrow 1$;
  WHILE $\Phi_1(p + \alpha_k d_k; \mu_k) >$
    $\Phi_1(p_k; \mu_k) + \eta\alpha_k D_1(f(p_k; \mu_k); d_k)$ DO
    reset $\alpha_k \leftarrow \tau_\alpha \alpha_k$ for some $\tau_\alpha \in (0, \tau]$;
  END (WHILE)
  set $p_{k+1} \leftarrow p_k + \alpha_k d_k$ and $\lambda_{k+1} \leftarrow \lambda_k + \alpha_\lambda d_\lambda$;
  IF a quasi-Newton approximation is used THEN
    set $s_k \leftarrow \alpha_k d_k$;
    set $\hat{y}_k \leftarrow \nabla_p \mathcal{L}(p_{k+1}, \lambda_{k+1}) - \nabla_p \mathcal{L}(p_k, \lambda_{k+1})$;
    obtain B_{k+1} by updating B_k using
    a quasi-Newton formula
    $B_{k+1} = B_k + \frac{(\hat{y}_k - B_k s_k)(\hat{y}_k - B_k s_k)^T}{(\hat{y}_k - B_k s_k)^T s_k}$
  END (IF)
END (WHILE)

The strategy for choosing μ in the Algorithm 2 considers the effect of the step on a model of the merit function, so μ has to satisfy the inequality

$$\mu \geq \frac{\nabla f_k^T d_k + \frac{\sigma}{2} d_k^T \nabla^2_{pp}\mathcal{L}_k d_k}{(1-\rho)\|c_k\|_1}. \tag{33}$$

If the value of μ from the previous iteration of the SQP method satisfies eq. (33), it is left unchanged. Otherwise, μ is increased, so that satisfies this inequality with some margin. The constant σ is used to handle the case in which Hessian $\nabla^2_{pp}\mathcal{L}_k$ is not positive definite. We define $\sigma = 1$ if $d_k^T \nabla^2_{pp}\mathcal{L}_k d_k > 0$, and $\sigma = 1$ otherwise.

The l_1 merit function for the problem (21)-(22) takes the form

$$\Phi_1(p; \mu) = f(p) + \mu\|c_k\|_1. \tag{34}$$

The directional derivative of Φ_1 in the direction d_k satisfies

$$D(\Phi_1(p_k; \mu); d_k) = \nabla f_k^T d_k - \mu\|c_k\|_1. \tag{35}$$

ALGORITHM 3. The SQP-line search algorithm for solving the equality constrained problem

BEGIN
  define a vector of decision variables $\tilde{p}$
    and its initial conditions;
  choose from vector $\tilde{p}$ a subvector p,
    which describes a subsystem
    $S = f(p)$
  solve problem (36) using Algorithm 2;
  update values of vector $\tilde{p}$ using results
    from the previous step;
END

As one can see, the Algorithm 2 can be thought as an inner loop in Algorithm 3. The last question is, what is the rate of convergence of the considered algorithm.

Assumption 3: The point p^* is a local solution of the problem (21)-(22) at which the following conditions hold.

a) The functions f and c are twice differentiable in a neighborhood of p^* with Lipschitz continuous second derivatives.

b) The linear independence constraint qualification holds at p^*.

c)The second order sufficient conditions hold at (p^*, λ^*).

Now one can call the theorem, which justifies the correctness of the presented algorithm.

Theorem ([10]): Suppose, that Assumption 3 holds and that the iterates p_k generated by Algorithm 1 with guasi-Newton approximate Hessian B_k, converge to p^*. Then p_k converges superlinearly if and only if the Hessian approximation satisfies

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2_{pp}\mathcal{L}_*)(p_{k+1} - p_k)\|}{\|p_{k+1} - p_k\|} = 0. \tag{36}$$

Lemma 4: Algorithm 3 generates a sequence of the feasible solutions with decreasing values of the goal function. In this bounded sequence one can distinguish a subsequence, which is superlinearly convergent to the locally optimal solution p^*.

V. Numerical results

Simulations were executed on the large-scale model of the pressure-constrained batch reactor.

When the model was divided into 1 000 submodels, then for solving the KKT system more than 24 hours was needed. So, the reactor was divided into 100 parts and the solution was obtained in 12 hours. At this step the vector of decision variables was stated as follows

$$p = [u_1 \cdots u_{100}, C_{A0,2} \cdots C_{A0,100}, \tag{37}$$
$$C_{B0,2} \cdots C_{B0,100}, C_{D0,2} \cdots C_{D0,100}].$$

Solution of this model was used as the initial conditions in the further work.

The question is, how to choose the vector of decision variables, to obtain in a reasonable time a possibly greatest improvement of the solution.

Then the reactor was divide into 1 000 parts. There are 3 997 decision variables in the system (1 000 piecewise

TABLE I
RESULTS OF THE SIMULATIONS IN CASE 1.

Size of subvector p	Number of iterations	$\bar{d}$
10	37	$8.6448e-004$
20	17	$8.4317e-004$
50	6	$9.3396e-004$
100	3	$9.3396e-004$

TABLE II
RESULTS OF THE SIMULATIONS IN CASE 2.

Size of subvector p	Number of iterations	$\bar{d}$
10	130	$7.5896e-004$
20	82	$7.6063e-004$
50	40	$7.6116e-004$
100	19	$7.6136e-004$

constant control functions and 2 997 variables treated as initial conditions for differential state trajectories).

$$\widetilde{p} = [u_1 \cdots u_{1000}, C_{A0,2} \cdots C_{A0,1000}, \quad (38)$$
$$C_{B0,2} \cdots C_{B0,1000}, C_{D0,2} \cdots C_{D0,1000}].$$

The simulations were executed for 4 different possible number of variables in the subvector $\widetilde{p}$: 10, 20, 50 and 100. As decision variables only control function, especially in the initial phases of the process, were considered. This enables increase the accuracy of the calculation, when the reactions proceed quickly. The initial average discontinuity in the state variables was $\bar{d} = 1.1e-3$ and $C_D(t_f) = 10.7240\ mol/m^3$. The final value of $C_D(t_f)$ is about 8.7% better then result presented in [6].

In the simulations two different stop criteria in Algorithm 2 were used. In the implementation convergence is declared when $TolFun < \epsilon_1$ and $TolCon < \epsilon_2$. $TolFun$ denotes termination tolerance on the function value, and $TolCon$ denotes tolerance on the constraint violation.

1) Case 1: In the first case the local optimization processes were performed more precisely. So, $TolFun < 1e-6$ and $TolCon < 1e-6$.

2) Case 2: In the second case stop criterion in local optimization were no so rigorous: $TolFun < 1e-3$ and $TolCon < 1e-3$.

The main stop criterion was the performance time. When computing time exceeded 12 hours, then optimization process was stopped.

In both cases the augmented objective function was considered

$$f(p) = C_D(t_f) + \rho \sum_{l=1}^{N_T} (z_0^{l+1} - \hat{z}_l)^2, \quad (39)$$

where penalty parameter $\rho = 10^6$.

Equation (39) shows the balance in the quest to minimize the concentration of component D and to meet the continuity constraints in differential-algebraic equations.

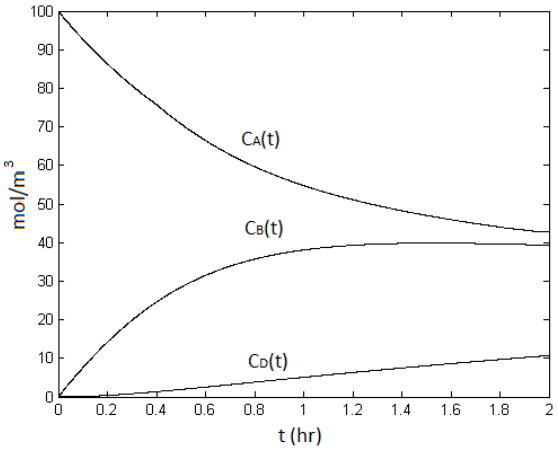


Fig. 2. Differential state trajectories. Results for size of subvector p=10. Stop criteria like in case 2.

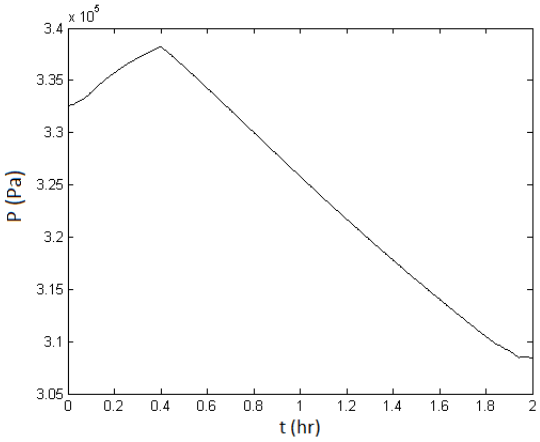


Fig. 3. The optimal pressure trajectory.

Results presented in the Table 1 and Table 2 show, that the inexact algorithm with the weak stop criteria, can obtain a better improvement of the initial solution.

As it was mentioned, in the considered problem two opposite tasks were considered. The minimization of the component D stands in opposition to fulfilment the constraints. As a result, the final concentration was improved and the obtained solution meets the constraints with high accuracy, so this method can be applied in real-life chemical processes.

The solutions obtained for size of the subvector $p = 10$ and stop criteria like in case 2 were presented on the figures 2-4. There are the differential state trajectories in the figure 2, the optimal pressure trajectory in the fig. 3 and the optimal flowrate profile in the fig. 4

VI. CONCLUSION

In the article the task of control of the pressure-constrained batch reactor was considered. The complex model of the reactor was designed using the simultaneous approach. The

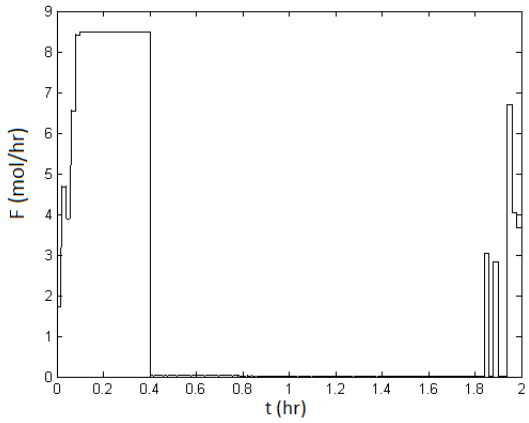


Fig. 4. Control variable - the optimal flowrate profile.

new SQP-line search algorithm was designed and tested. The algorithm, which takes in each iteration only a few number of decision variables into account, can do new iterations and improve the initial solution. But in both approaches a large number of variables were considered.

Because in the pure form, SQP algorithm is convergent to the locally solution, line search was used as a globalization approach to construct a sequence of feasible solutions with decreasing values of the objective function.

This type of algorithms can be successfully applied to the large systems, when Jacobian and Hessian matrices are dense and structure of these matrices can not be effectively used.

As the conclusion we want to pay attention to need for solver for the large-scale optimization and optimal control problems. Second order information, which can be approximated using BFGS method, can be unavailable when Jacobian matrix is difficult to calculate. This situation one can be met very often, when simultaneous approach is used. Multistep algorithms, which need feasible initial conditions, can improve the solution in considerable short time. At the end we want to emphasize the need for Jacobian-free optimization algorithm, which could solve the large-scale optimization tasks [7].

VII. Nomenclature

C	concentration (mol/m^3)
$\bar{d}$	average discontinuity in the state variables
F	flowrate (mol/hr)
f	objective function in optimization problem
g	function of algebraic constraints
k_1, k_2, k_3	rate constants
N_T	number of shots
n, m	dimensions of the space
p	vector of decision variables
S	function describing subproblem
t	time (hr)
z	state variables
y	algebraic variables
$\mathcal{L}$	Lagrangian function
$\mathcal{R}$	real numbers

Greek symbols

Φ	function describing optimization problem
ϕ	function describing system
λ	Lagrangian multipliers
ρ	penalty parameter
ϵ	tolerance

Superscripts

T	transposition of the matrix

Subscripts

A, B, D	components of the reaction
L	lower bound
U	upper bound
l	number of a shot

References

[1] J.T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming. 2nd edition*, SIAM, Philadelphia 2010.
[2] L.T. Biegler, *Nonlinear Programming. Concepts, Algorithms, and Applications to Chemical Processes*, SIAM, Philadelphia 2010.
[3] K.E. Brenan, S.L. Campbell, L.R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM, Philadelphia 1996.
[4] W.E. Feehery, J.E. Tolsma, P.I. Barton, "Efficient sensitivity analysis of large-scale differential-algebraic systems", *Applied Numerical Mathematics*, vol. 25, 1997, pp. 41-54.
[5] E. Hairer , C. Lubich, M. Roche, "The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Method", *Lecture Notes in Mathematics*, vol. 1409, Springer-Verlag, Berlin, 1989.
[6] Y.J. Huang, G.V. Reklaitis, V. Venkatasubramanian, "Model decomposition based method for solving general dynamic optimization problems", *Computers and Chemical Engineering*, vol. 26, 2002, pp. 863-873.
[7] D.A. Knoll, D.E. Keyes, "Jacobian-free Newton-Krylov methods: a survey of approaches and applications", *Journal of Computational Physics*, vol. 193, 2004, pp. 357-397.
[8] D.B. Leineweber, A. Schaefer, H.G. Bock, J.P. Schloeder J. P., "An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part II: Software aspects and appliactions", *Computers and Chemical Engineering*, vol. 27, 2003, pp. 167-174.
[9] R. Maerz, "Numerical methods for differential algebraic equations", *Acta Numerica*, vol. 1, 1992, pp. 141-198.
[10] J. Nocedal, S.J. Wright, *Numerical Optimization. 2nd edition*, Springer, New York, 2006.

[11] M. von Schwerin, O. Deutschmann ,V. Schulz, "Process optimization of reactives systems by partially reduced SQP methods", *Computers and Chemical Engineering*, vol. 24, 2000, pp. 89-97.
[12] V.S Vassiliadis, R.W.H. Sargent, C.C. Pantelides, "Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems without Path Constraints", *Ind. Eng. Chem. Res*, vol. 33, 1994, pp. 2111-2122.
[13] V.S Vassiliadis, R.W.H. Sargent, C.C. Pantelides, "Solution of a Class of Multistage Dynamic Optimization Problems. 2. Problems with Path Constraints", *Ind. Eng. Chem. Res*, vol. 33, 1994, pp. 2123-2133.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 315–319

Bicriteria Fuzzy Optimization Location-Allocation Approach

Santiago García-Carbajal
Escuela Politécnica Superior de Ingeniería, Universidad de Oviedo, 33204 Gijón, Spain
Email: carbajal@lsi.uniovi.es

Belarmino Adenso-Díaz
Escuela Politécnica Superior de Ingeniería, Universidad de Oviedo, 33204 Gijón, Spain
Email: adenso@epsig.uniovi.es

Sebastián Lozano
Escuela Superior de Ingenieros, University of Seville, 41092 Seville, Spain
Email: slozano@us.es

***Abstract*—Distribution network design deals with defining which elements will be part of the supply chain and how they will be interrelated. Many authors have studied this problem from a cost minimization point of view. Nowadays the sustainability factor is increasing its importance in the logistics operations and must be considered in the design process. We deal here with the problem of determining the location of the links in a supply chain and the assignment of the final customers considering at the same time cost and environmental objectives. We use a fuzzy bicriteria model for solving the problem, embedded in a genetic algorithm that looks for the best trade-off solution. A set of experiments have been carried out to check the performance of the procedure, using some instances for which we know a priori a good reference solution.**

I. Introduction

THE fierce competition between the different supply chains makes it necessary that efficiency be continuously pursued. One of the most important strategic decisions, and one that has a long-term impact in the economic results of the logistics operations, is the design of the distribution network.

Distribution network design is the process of determining the structure of a supply chain, defining which elements will be part of it (i.e., where locate the facilities), and what will be the interrelationships between them (i.e., the allocation of customers to facilities and how the material and products will flow in the network between the nodes in the network). For that reason the problem is often called location-allocation (e.g. [1]).

In [2], Akkerman *et al.* consider, from a hierarchical point of view, a second level called distribution network planning that includes the decisions related to fulfilling the aggregate demand (i.e., aggregate product flows and delivery frequencies)

Many authors have studied these problems, most of them (around two thirds according to [3]) by considering as the objective the minimization of the costs involved in the process. However, for different reasons (legal pressure, customers demand, ethical consciousness, etc) nowadays the sustainability factor is increasing its importance in the business management and specifically in the logistics operations, where transportation of goods is a high pollutant activity. It is therefore necessary to consider the operations impact when defining the distribution network.

The aim of this study is the formulation of a model and a solution procedure for the location-allocation problem when two criteria (cost and environmental impact of transportation) are considered at the same time.

II. Problem Setting

Let us suppose that there is an uncapacitated central plant that must distribute a single product among many customers. Those customers have uncertain (i.e. fuzzy) demands. We need to define the distribution network, choosing the capacitated intermediate warehouses to set up, and allocating each customer to one of warehouses (or to the central facility). There are two types of vehicles: large trucks (used for high demand customers and for serving the warehouses from the central plant) and smaller trucks.

We are going to consider two objective functions. One is the minimization of the logistics costs (transportation and warehouses set-up). The transportation costs will be proportional to distances and depend on type of truck used. The second objective function is the minimization of the environmental impact of the Greenhouse Gases (GHG) emissions (e.g. CO_2) due to transportation.

Note that, in principle, every customer could be served from the central facility, but if the demand is small, the cost and environmental impact of such direct shipments would be very high, likely bigger that delivering the goods from a near warehouse.

Our problem consists in deciding which of the potential warehouse locations will be opened and from which warehouse should each customer be served warehouse each customer will be allocated (considering the limited capacities of the warehouses) in such a way that total cost and GHG emissions are minimized.

III. Model Formulation

Table I shows the notation used for modeling the problem. Note that the set of potential warehouse locations are given together with the distance, unit transport cost and unit GHG emissions factor from the central plant to each potential warehouse location j. From each potential warehouse loca-

This research was funded by the Spanish Ministry of Science and FEDER, grant DPI2010-16201.

TABLE I.
NOTATION

i	Index on customers (i=1..N)
j	Index on potential warehouse locations (j=1..A)
I(j), $\hat{I}$	Subsets of customers that can be served from warehouse j and from the central plant, respectively
$\widetilde{D}_i$	Fuzzy demand of customer i. A Triangular Fuzzy Number membership function is assumed. $\mu_{D_i}(x) = \begin{cases} 0 & \text{if } x \le D_i^- \\ \frac{x - D_i^-}{D_i^0 - D_i^-} & \text{if } D_i^- \le x \le D_i^0 \\ \frac{D_i^+ - x}{D_i^+ - D_i^0} & \text{if } D_i^0 \le x \le D_i^+ \\ 0 & \text{if } x \ge D_i^+ \end{cases}$
$\widetilde{U}_j$	Fuzzy capacity of warehouse j. A decreasing linear membership function is assumed $\mu_{U_j}(x) = \begin{cases} 1 & \text{if } x \le U_j^- \\ \frac{U_j^+ - x}{U_j^+ - U_j^-} & \text{if } U_i^- \le x \le U_j^+ \\ 0 & \text{if } x \ge U_j^+ \end{cases}$
$\widetilde{L}_j$	Fuzzy minimum flow of warehouse j. An increasing linear membership function is assumed (with parameter $L_i^+ << U_j^-$). $\mu_{L_j}(x) = \begin{cases} 0 & \text{if } x \le L_j^- \\ \frac{x - L_j^-}{L_j^+ - L_j^-} & \text{if } L_i^- \le x \le L_j^+ \\ 1 & \text{if } x \ge L_j^+ \end{cases}$
f_j	Fixed cost of warehouse j
c_{ji}	Unit transport cost between warehouse j and customer i
$\hat{c}_i$	Unit transport cost between central plant and customer i
$\hat{\hat{c}}_j$	Unit transport cost between central plant and warehouse j
e_{ji}	Unit GHG emissions factor for transport between warehouse j and customer i
$\hat{e}_i$	Unit GHG emissions factor for transport between central plant and customer i
$\hat{\hat{e}}_j$	Unit GHG emissions factor for transport between central plant and warehouse j
t_{ji}	Distance between warehouse j and customer i
$\hat{t}_i$	Distance between central plant and customer i
$\hat{\hat{t}}_j$	Distance between central plant and warehouse j
x_{ji}	Amount of product shipped from warehouse j to customer i∈ I(j)
$\hat{x}_i$	Amount of product shipped from central plant to customer
y_j	Amount of product shipped from central plant to warehouse j

tion j only a subset of customers I(j) can be served. The distance, unit transport cost and unit GHG emissions factors from each warehouse location to each customer $i \in I(j)$ are given.

The membership function of the demand of each customer i is given by a Triangular Fuzzy Number (TFN) with parameters $\left(D_i^-, D_i^0, D_i^+\right)$. Each warehouse has a capacity, i.e. an upper bound on the flow of goods that it can convey from the central plant to its allocated customers. The membership function of the capacity of warehouse j is given by a linear decreasing function with parameters $\left(U_j^-, U_j^+\right)$. Each warehouse also has a lower bound on the flow that it should handle in case it is selected. This minimum flow is imposed to guarantee an economic operation of the warehouse. The membership function of the minimum flow of warehouse j is given by a linear increasing function with parameters $\left(L_j^-, L_j^+\right)$.

The proposed bicriteria optimization model consist in the minimization of both cost and GHG emissions:

$$\text{Min} \quad \sum_j f_j\, y_j + \sum_{i \in \hat{I}} \hat{c}_i\, \hat{t}_i\, \hat{x}_i + \sum_j \sum_{i \in I(j)} \left(\hat{\hat{c}}_j\, \hat{\hat{t}}_j + c_{ji}\, t_{ji}\right) x_{ji} \tag{1}$$

$$\text{Min} \quad \sum_{i \in \hat{I}} \hat{e}_i\, \hat{t}_i\, \hat{x}_i + \sum_j \sum_{i \in I(j)} \left(\hat{\hat{e}}_j\, \hat{\hat{t}}_j + e_{ji}\, t_{ji}\right) x_{ji} \tag{2}$$

subject to

$$\hat{x}_i + \sum_{\{j: i \in I(j)\}} x_{ji} = \tilde{D}_i \qquad \forall i \in \hat{I} \tag{3}$$

$$\sum_{\{j: i \in I(j)\}} x_{ji} = \tilde{D}_i \qquad \forall i \notin \hat{I} \tag{3'}$$

$$\tilde{L}_j\, y_j \le \sum_{i \in I(j)} x_{ji} \le \tilde{U}_j\, y_j \tag{4}$$

$$\hat{x}_i \ge 0 \;\; \forall i \in \hat{I} \quad x_{ji} \ge 0 \;\; \forall j \forall i \in I(j) \quad y_j \in \{0,1\} \;\; \forall j \tag{5}$$

In order to solve this model, a Fuzzy Multiobjective Optimization approach based on the additive model of Tiwari [4] is proposed. Thus, the new objective function, to be maximized, will be the sum of the membership functions of the fuzzy constraints and of the two objective functions. The latter are fuzzified using decreasing linear membership functions, between the thresholds (C^-,C^+) and (E^-,E^+), respectively. These total cost and total emissions thresholds are evaluated in the following way. For C^+, model (1),(3)-(5) is solved maximizing transportation costs and assuming all the potential warehouses are closed. Let Ψ be the resulting maximum transportation cost, then $C^+=\Psi^+ + \Sigma f_i$.

For the calculation of C^-, model (1),(3)-(5) is solved minimizing transportation costs and assuming that all the warehouses are open. Let Ψ^- be the resulting minimum transportation cost, then $C^-=\Psi^- - \Sigma f_i$. For the calculation of E^+, model (2)-(5) is solved maximizing total emissions and assuming that all the warehouses are closed. Finally, for calculating E^-, model (2)-(5) is solved minimizing total emissions and assuming that all the warehouses open.

We assume that both objectives (minimizing total costs and total emissions) are equally important. As regards the constraints we shall request that their membership function values should be higher than a lower bound μ_{min} (see [5]). The model to solve is, thus, the following

$$\text{Max} \quad \lambda_1 + \lambda_2 \tag{6}$$

subject to

$$C = \sum_j f_j\, y_j + \sum_{i \in \hat{I}} \hat{c}_i\, \hat{t}_i\, \hat{x}_i + \sum_j \sum_{i \in I(j)} \left(\hat{\hat{c}}_j\, \hat{\hat{t}}_j + c_{ji}\, t_{ji}\right) x_{ji} \tag{7}$$

$$E = \sum_{i \in \hat{I}} \hat{e}_i\, \hat{t}_i\, \hat{x}_i + \sum_j \sum_{i \in I(j)} \left(\hat{\hat{e}}_j\, \hat{\hat{t}}_j + e_{ji}\, t_{ji}\right) x_{ji} \tag{8}$$

$$\lambda_1 \le \frac{C^+ - C}{C^+ - C^-} \tag{9}$$

$$\lambda_2 \le \frac{E^+ - E}{E^+ - E^-} \tag{10}$$

$$D_i^- + \mu\left(D_i^0 - D_i^-\right) \le \hat{x}_i + \sum_{\{j: i \in I(j)\}} x_{ji} \le D_i^+ - \mu\left(D_i^+ - D_i^0\right) \;\; \forall i \in \hat{I} \tag{11}$$

$$D_i^- + \mu\left(D_i^0 - D_i^-\right) \le \sum_{\{j: i \in I(j)\}} x_{ji} \le D_i^+ - \mu\left(D_i^+ - D_i^0\right) \;\; \forall i \notin \hat{I} \tag{11'}$$

$$\mu\, y_j \le \frac{U_j^+ - \sum_{i \in I(j)} x_{ji}}{U_j^+ - U_j^-}\, y_j \qquad \forall j \tag{12}$$

$$\mu\, y_j \le \frac{\sum_{i \in I(j)} x_{ji} - L_j^-}{L_j^+ - L_j^-}\, y_j \qquad \forall j \tag{12'}$$

$$\sum_{i \in I(j)} x_{ji} \le U_j^+\, y_j \qquad \forall j \tag{12''}$$

$$0 \le \lambda_1 \le 1; \quad 0 \le \lambda_2 \le 1; \quad \mu_{min} \le \mu \le 1 \tag{13}$$

$$x_{ji}, \hat{x}_i \ge 0 \quad \forall i \in \hat{I} \, \forall j \in I(j) \quad y_j \in \{0,1\} \quad \forall j \tag{14}$$

IV. Solution procedure

In order to solve the above model a Genetic Algorithm (GA) will be used. The GA explores which warehouses are to be opened (binary variables y_j) and, for each individual, a Linear Programming (LP) solver is used to compute the corresponding fitness function selecting is the best customer allocation, using model (6)-(14) with variables y_j fixed (see Fig. 2). Note that, in principle, not every subset of warehouses is feasible, i.e., there is not always enough demand in the area of influence of the warehouses I(j) to cover the minimum flow required to open the facilities as per constraints (12'). Therefore, a check needs to be done previous to calling the optimization software that solves the LP model. In case the candidate warehouses to be opened are seen to lead to an infeasible solution, changes in the

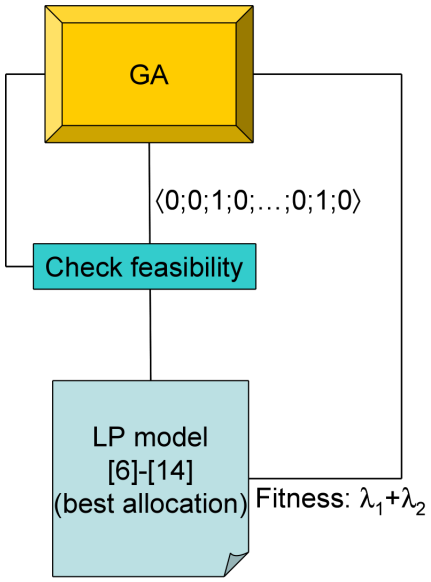


Fig. 1 GA solution procedure solves LP model for fitness evaluation

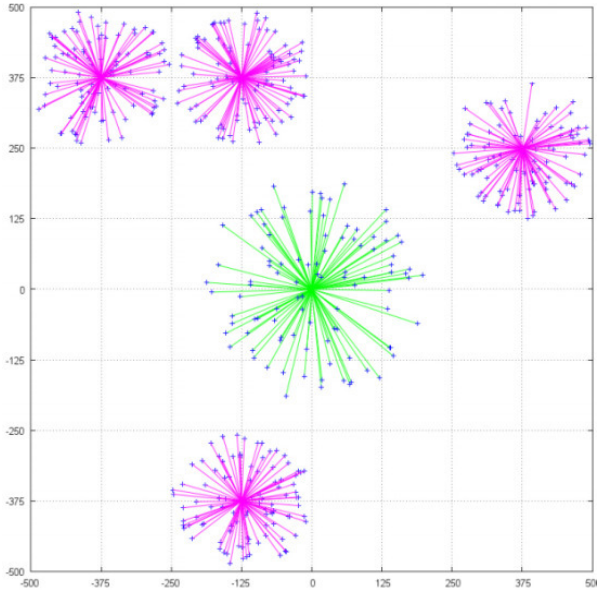


Fig. 2 Example of a priori solution with 4 warehouses

warehouses subset are made until it can be assured there that the LP optimization software will return a feasible solution. This can be seen as a repair operator, which is one of the possible ways of handling constraints in GA.

Since the solution space explored by the GA corresponds to binary variables (y_j) a binary codification of the solution is used, i.e. each chromosome is just a vector of as many components as potential warehouse locations. Each component encode whether a warehouse is open or not. In order to assign a fitness value to an individual a linear solver is used to solve model (6)-(14) also obtaining the complete specification of the solution, including the flows between the central plant and the open warehouses and from these to their allocated customers.

About the crossover and mutation operators, standard binary coding operators have been used, namely the 1-point crossover (1X crossover) and the bitwise mutation. Fitness-proportional selection (i.e. roulette wheel) is used to choose the individuals to cross over. A generational GA is used with a maximum number of generations. An additional stopping criterion consists in a limit on the number of generations without improving the best solution found.

As regards the implementation of the GA, an efficient parallel Python code has been programmed. Although the details of the parallelization strategy is out of the scope of this paper, let us just say that parallel python allows for calculating in parallel of the fitness of all the individuals in initial population as well as of the new individuals created in each generation.

V. Computational Experiments and Results

For testing the good performance of the proposed approach, we have created a testbed of instances, each one with a 7x7 square grid of potential warehouses locations and with the central plant in the middle of the grid. The size of each of the grid cells is 100 km×100 km. The data were created in such a way that we have a clue about which could be the best possible solution, and them we shall check if our procedure is able to find a solution at least as good as that. With that purpose, customers were created locating them around a specific warehouse, forming a kind of cluster. Thus, for example, Fig. 2 shows an instance with four clusters of customers generated around four chosen warehouses. An additional cluster of customers, not in the vicinity of the four chosen warehouses, is also generated, with the expectation that these customers will likely be allocated to the central plant.

Two sets of 20 instances each were created. In the first set 2 warehouses are opened and 4 in the other 4. Therefore, 40 instances were solved and compared with the corresponding a priori "cluster" solution.

For each of the N selected warehouses, a set of (500/N)·4/5 clients in a radius distance of 125 km, all with the same demand, are randomly generated. The other fifth of the warehouse customers were generated out of that neighborhood. Note that there is always a feasible solution since we assume that the central facility can always deliver goods to any client (although at a higher cost). Capacity is assigned to each warehouse in such a way that the defined solution is feasible.

For the two types of vehicles (trucks and vans) cost and emission factors are shown in Table II and include the corresponding corrections to deal with non-full truckloads. The emission factors used correspond to those computed by the LIPASTO model developed by the Technical Research Centre of Finland (VTT) ([6]).

For the GA a population size of 100 was used, mutation probability was set to 0.001, maximum number of generations was 100 but stopping before reaching that limit if 10 generations pass without improving the best solution found.

Comparing the results obtained with the clustered solution from which the instance customer data were generated, it can be seen in Fig. 3 that the GA procedure has been suc-

TABLE III.
COSTS AND EMISSIONS OF TRUCKS AND VANS DEPENDING ON DISTANCE AND LOAD

	Truck	Van	Non-full truck, from central depot, >125km	From warehouse,>125km
c_{ji}	0.00004 €/kg/km	0.00030 €/kg/km		0.00045 €/kg/km
$\hat{c}_i$	0.00004 €/kg/km	0.00030 €/kg/km	0.00006 €/kg/km	
$\hat{\hat{c}}_j$	0.00004 €/kg/km	0.00030 €/kg/km		
e_{ji}	0.0621 gr Eq-CO_2/kg/km	0.0950 gr Eq-CO_2/kg/km		0.1425 gr Eq-CO_2/kg/km
$\hat{e}_i$	0.0621 gr Eq-CO_2/kg/km	0.0950 gr Eq-CO_2/kg/km	0.1425 gr Eq-CO_2/kg/km	
$\hat{\hat{e}}_j$	0.0621 gr Eq-CO_2/kg/km	0.0950 gr Eq-CO_2/kg/km		

cessful in 27 out of the 40 instances (two thirds of the cases) location the warehouses according to the corresponding a priori clustered solution considered. Overall, the fitness of the GA solution (measured by $\lambda_1+\lambda_2$) is 2.2% below that of the a priori clustered solution. Note that as the problem complexity increases (as the number of clusters in the instance increases), it occurs more often (0% in the case of two clusters, 45% in 4 clusters case) that the GA does not find the a priori clustered solution. Different ways to compensate this effect are being studied to make the GA more robust.

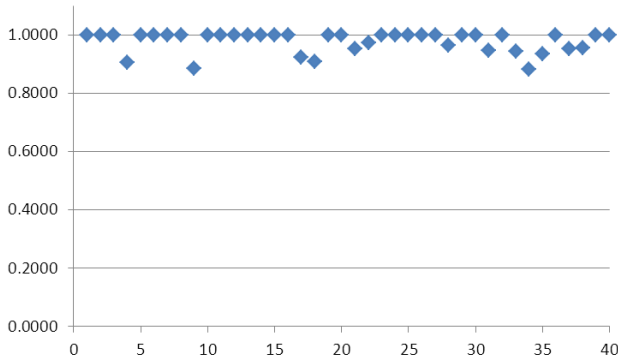


Fig. 3 Ratio between the fitness of the final GA solution and the fitness of the original clustered solution

VI. CONCLUSIONS

This research has proposed a new network design approach that aims not only at cost minimization but also at minimizing GHG emissions from goods transportation. This second objective function will contribute to the sustainability of logistics operations. The decision variables are the selection of the warehouse location (from a set of discrete potential locations) and the allocation of customers to the selected warehouses.

A fuzzy bicriteria optimization model for solving the problem has been formulated and a GA solution procedure has been implemented. The GA explores the space of solutions corresponding to the selection of warehouses to open. A binary codification has been used so that if a potential location is opened the corresponding gene is one and zero otherwise. Standard crossover and mutation operator are used. Since there are both lower and upper bounds on the capacity of the open warehouses, a repair mechanism is needed to guarantee that these constraints hold and that the individual whose fitness is to be evaluated leads to a feasible solution.

A set of experiments have been carried out to check the performance of the procedure, using some instances for which a good reference solution is known a priori. The results indicate that the proposed approach generally finds (or gets close to) this reference solution.

REFERENCES

[1] M.J. Meixell and V.B. Gargeya, "Global supply chain design: a literature review and critique", *Transport Research Part E*, vol. 41, pp. 531-550, 2005.

[2] R. Akkerman, P. Farahani, and M. Grunow, "Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges", *OR Spectrum*, vol. 32, pp. 863-904, 2010.

[3] M.T. Melo, S. Nickel and F. Saldanha-da-Gama, "Facility location and supply chain management: a review", *European Journal of Operational Research*, vol. 196, pp. 401-412, 2009

[4] R.N. Tiwari, S. Dharmar and J.R. Rao, "Fuzzy goal programming - an additive model", *Fuzzy Sets and Systems*, vol. 24, pp. 27-34, 1987

[5] L.H. Chen and F.C. Tsai, "Fuzzy goal programming with different importance and priorities", *European Journal of Operational Research*, vol. 133, pp.548-556, 2001

[6] LIPASTO Unit emissions of vehicles, Freight transport - road traffic, http://lipasto.vtt.fi/yksikkopaastot/tavaraliikennee/tieliikennee/tavara_tiee.htm

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 321–328

Branch and Price for Preemptive Resource Constrained Project Scheduling Problem Based on Interval Orders in Precedence Graphs

Aziz Moukrim
HEUDYASIC Laboratory
Technological University
COMPIEGNE
Email: aziz.moukrim@hds.utc.fr

Alain Quilliot
LIMOS CNRS UMR 6158
LABEX IMOBS3, Université
Blaise Pascal
Bat. ISIMA, BP 10125
Campus des Cézaux,
63173 Aubière, France
Email: quilliot@isima.fr

Hélène Toussaint
LIMOS CNRS UMR 6158
LABEX IMOBS3, CNRS
Bat. ISIMA, BP 10125
Campus des Cézaux,
63173 Aubière, France
Email: toussain@isima.fr

***Abstract*—This paper describes an efficient exact algorithm to solve Preemptive Resource Constrained Project Scheduling Problem (Preemptive RCPSP). We propose a very original and efficient branch and bound procedure based upon minimal interval order enumeration, which involves column generation as well as constraint propagation and which is implemented with the help of the generic SCIP software. We perform tests on the famous PSPLIB instances which provide very satisfactory results. To the best of our knowledge it is the first algorithm able to solve at optimality all the set of j30 instances of PSPLIB in a preemptive way. Moreover, this algorithm allows us to update several best known lower bounds for the j60, j90 and j120 instances of PSPLIB.**

I. Introduction

THIS paper deals with the *Preemptive Resource Constrained Project Scheduling Problem* (RCPSP: see [1], [2]). RCPSP aims at scheduling a set of activities, submitted to precedence and resource constraints, while minimizing the induced *makespan* (total duration of the project) value. The precedence constraints mean that some activities must be completed before others can start. The resource constraints specify that each activity requires constant amounts of renewable resources during all the time it is processed, these resources having limited capacities. This problem has been extensively studied in its non preemptive version ([3], [4]), which means that every activity has to be run as a whole, without any kind of interruption. There exist several variants of RCPSP (see [5], [6] for recent surveys). We talk about Preemptive RCPSP when an activity may be run in several steps: one may launch such an activity, interrupt it, keep on with this activity a little further, and so on. There exists few works on Preemptive RCPSP: [7] developed a branch and bound algorithm, [8] proposed a tree search procedure augmented with pruning rules (best-first tree search), [9] proposed an integer linear program which add preemption penalties, [10] and [11] dealt with preemption in an heuristic way and [12] designed a genetic algorithm for multi mode Preemptive RCPSP.

For the sake of simplicity, authors often assume that all processing times are integral and that preemption only occurs at integer valued dates. Still, one easily checks that such a hypothesis is very restrictive, and only allows to get an approximation of the optimal value of the problem. In this paper, we consider the problem under its most general form and suppose that preemption is allowed for all activities and may occur at arbitrary rational dates, and that no penalties are related to preemption.

Our approach is a Branch and Bound one which involves constraint propagation, as well as the management of specific rational *Antichain* linear program whose variables are associated with subsets of activities which may be simultaneously processed during the schedule. This LP, which was was first introduced by [13], provides us with a lower bound of both Preemptive and Non Preemptive RCPSP. But dealing with it requires implementing a pricing or column generation scheme. It was proved in [14] that if the input RCPSP instance satisfies some ad hoc properties, then any optimal solution of the *Antichain* linear program may be turned into a feasible optimal schedule, without any increase of the *makespan* value. What we do here is to use this property in order to perform a tree search which may be viewed as being embedded into the enumeration process of all minimal extensions of the precedence relation which define *interval* orders. The resulting process happens to be very efficient, since it is able to solve in an exact way all 30 activity instances of the PSPLIB library, and to improve best existing lower bounds for several 60/120 activity instances of this library.

So the paper is organized as follows: we first recall what is Preemptive RCPSP (Section II), and next introduce the theoretical tools related to the *Antichain* LP and to interval orders (Section III), which will provide us with the basis of our algorithmic approach. Section IV describes the algorithm INT-ORD-ENUM and its implementation, and Section V is devoted to a presentation of experimental results.

II. Preemptive RCPSP

An instance $I = (X, K, \ll)$ of the *Resource Constrained Project Scheduling Problem* is defined by:

- A set $X = \{1,...,n\}$ of n *activities*: $\forall\ i \in X$, d_i denotes the *duration* of activity i
- A set $K = \{1,...,m\}$ of m *resources*: $\forall\ i \in X$, $\forall\ k \in K$, r_{ik} denotes the requirement of activity i for resource k; those resources are given back to the system once the activity is over

- $\forall (i,j) \in X^2$, $i \ll j$ means that *i precedes j*: activity *j* cannot start before *i* is over (*Precedence* constraints)

In the case of *Non Preemptive* RCPSP, scheduling only means computing the starting times t_i, $i \in X$, of the activities. A schedule $\sigma = (t_i, i \in X)$ is feasible if it satisfies:

- the *Precedence* constraints;
- the *Resource* constraints: at any time *t* during the process, and for any resource *k*, the sum $\sum_{i \in Act(\sigma,t)} r_{ik}$ does not exceed the global resource amount R_k, $Act(\sigma, t) = \{i$ such that $t_i < t < t_i + d_i\}$ denoting the set of the activities currently run at time *t* according to schedule σ.

So, solving *Non Preemptive* RCPSP means computing σ with a minimal *makespan* (total duration of the process).

In case preemption is allowed, scheduling an activity *i* means first decomposing *i* into a sequence of sub-activities $i_1, .., i_{h(i)}$, with durations $d_{i,1},.., d_{i,h(i)}$, such that: $\sum_{q = 1..h(i)} d_{i,q} = d_i$, and next scheduling all these sub-activities in the sense of standard RCPSP. Since there does not exist any "a priori" restriction either on the number of sub-activities or on their durations, which may be arbitrarily small, the existence of an optimal solution of Preemptive RCPSP has to be discussed.

III. Fundamental tools

A. The Antichain Linear Program: A Lower Bound

Let $I = (X, K, \ll)$ be some Preemptive RCPSP instance, defined according to notations of Section II. We suppose (we clearly may do it) that precedence relation $\ll$ is transitive. Then we define an *antichain* as being any subset *a* of X such that there does not exist $(i,j) \in a^2$ such that $i \ll j$. We say that such an antichain is *valid* if: $\forall k \in K, \sum_{i \in a} r_{ik} \leq R_k$.
It comes that a subset $a \subseteq X$ of activities is a *valid antichain* iff activities in *a* may be simultaneously run inside some feasible schedule. We denote by A the set of all *valid antichains*.

Then we become able to set the following linear program, which we call *Antichain Linear Program* associated with Preemptive RCPSP instance $I = (X, K, \ll)$, which was already introduced in [13, 14], and which we denote by $(\mathsf{P})_{\mathsf{Ant}}$:

$$(\mathsf{P})_{\mathsf{Ant}} \quad \begin{array}{l} \text{Minimize } \sum_{a \in \mathsf{A}} z_a \\ \text{Subject to} \\ \quad \forall i \in X, \quad \sum_{a \in \mathsf{A} \mid i \in a} z_a = d_i \\ \quad \forall a \in \mathsf{A}, \quad z_a \geq 0 \end{array} \qquad \text{(C1)}$$

Explanation: if σ is any feasible schedule related to instance *I*, we may associate with σ and with any valid antichain *a*, the total amount of time $z(\sigma)_a$ during which the activities which are simultaneously run according to σ are exactly the activities of *a*. Then we see that $z(\sigma) = (z(\sigma)_a, a \in \mathsf{A})$ is a feasible solution of $(\mathsf{P})_{\mathsf{Ant}}$ since constraints (C1) express the fact that any activity *i* has to be completely done, or, equivalently, that the duration of all antichains containing *i* must be equal to the duration of *i*. It comes that the optimal value of $(\mathsf{P})_{\mathsf{Ant}}$ provides us with a lower bound of the optimal value of *I*, which we denote by *LB*(*I*).

B. Dealing with $(P)_{Ant}$: Column generation

Since the set A may be very large, even when the activity set *X* is small, we need to handle the *Antichai*n LP $(\mathsf{P})_{\mathsf{Ant}}$ through *column generation*. Column generation is an usual technique to solve a LP which contains an exponential number of variables. It consists in initializing this LP with a few number of *active* variables (which may be obtained from application of some heuristic), and then in iteratively solving the induced *restricted problem* at optimality and using the dual variables to generate a new improving primal variable. The search for this improving primal variable is called the related *Pricing* Problem. The new variable is added to the restricted problem and the process goes on until no improving variable can be found: then the solution of the restricted problem is the optimal solution. When this technique is associated to a Branch and Bound process (usually for integer formulation) it gives rise to a *Branch and Price* process. In our case, let us consider some active antichain subset $B \subseteq \mathsf{A}$, together with some dual solution λ of the restricted LP $(\mathsf{P})^B_{\mathsf{Ant}}$ defined by (we suppose that *B* is such that this program admits a feasible solution):

$$(\mathsf{P})^B_{\mathsf{Ant}} \quad \begin{array}{l} \text{Minimize } \sum_{a \in \mathrm{B}} z_a \\ \text{Subject to} \\ \quad \forall i \in X, \quad \sum_{a \in B \mid i \in a} z_a = d_i \\ \quad \forall a \in \mathrm{B}, \quad z_a \geq 0, \end{array} \qquad \text{(C1)}$$

Then solving the related pricing problem PRICE(λ) means computing some valid antichain *a*, such that:

$$\sum_{i \in a} \lambda_i > 1.$$

Though this problem is NP-Complete, it may be efficiently handled through a combination of greedy search and Integer Linear Programming (LIP). A well-fitted LIP formulation of the PRICE(λ) problem comes as follows:

Pricing LIP Formulation L-PRICE(λ)

$$\begin{array}{l} \text{Maximize } \sum_{i \in X} \lambda_i y_i \\ \text{Subject to} \\ \quad \forall (i,j) \in X^2 \mid i \ll j, \quad y_i + y_j \leq 1 \qquad \text{(C2)} \\ \quad \forall k \in K, \ \sum_{i \in X} r_{ik} y_i \leq R_k \qquad \text{(C3)} \\ \quad \forall i \in X, \ y_i \in \{0,1\} \end{array}$$

C. Turning a Solution of $(P)_{Ant}$ into a Feasible Schedule ?

Unfortunately, Linear Program $(\mathsf{P})_{\mathsf{Ant}}$ only provides us with a lower bound of Preemptive RCPSP instance *I*: if vector $z = (z_a, a \in \mathsf{A})$ is a feasible solution of $(\mathsf{P})_{\mathsf{Ant}}$, it may not be possible to turn it into a feasible solution of *I*. As a matter of fact, we may provide the valid antichain set A with an oriented graph structure $(\mathsf{A}, E_{<<})$ by setting that there exists an arc $(a, b) \in E_{<<}$ from antichain *a* to antichain b, if there exist activities $i \in a$ and $j \in b$, such that $i \ll j$.

Then we easily check that:

Theorem 1: *Let z be some feasible solution of* $(\mathsf{P})_{\mathsf{Ant}}$, *and* $\mathsf{A}(z) \subseteq \mathsf{A}$ *be the set* $\mathsf{A}(z) = \{a \in \mathsf{A}$ *such that* $z_a \neq 0\}$ *of active*

antichains according to z. Then there exists a feasible schedule σ such that z = z(σ) if and only if the subgraph (A(z), $E_{<<}$) does not contain any circuit.

Proof: Left to the reader.

Still, we may notice that program $(P)_{Ant}$ provides us with additional understanding of Preemptive RCPSP: if σ is any feasible Preemptive RCPSP schedule, if $z(\sigma) = (z(\sigma)_a, a \in A)$ is the related solution of $(P)_{Ant}$, and if $A(z(\sigma))$ is the related active antichain set, then one sees that solving the restricted linear program $(P)^{A(z(\sigma))}_{Ant}$ through Primal Simplex Algorithm provides us with another feasible schedule σ* with makespan no larger than the makespan of σ. Moreover, Linear Programming Theory tells us that the number of active antichains related to σ*, that means the cardinality of $A(z(\sigma^*))$ does not exceed the number of constraints of $(P)^{A(z(\sigma))}_{Ant}$, which is equal to the cardinality of the activity set *X*. This makes appear Preemptive RCPSP as a combinatorial problem related to the search of some acyclic subgraph (*B*, $E_{<<}$) of the antichain graph (*A*, $E_{<<}$), such that:

- Card(*B*) ≤ Card(*X*);
- The optimal value of the program $(P)^{B}_{Ant}$ is minimal.

This confirms the existence of an optimal solution.

Also, we may notice that no activity which is not in the set *Min*(*X*) of the activities which are minimal in the sense of the precedence relation ≪, may start before the time when at least one activity in *Min*(*X*) is completed. We deduce that the lower bound which derives from the resolution of the $(P)^{A}_{Ant}$ program may be improved by adding the following constraint to $(P)^{A}_{Ant}$:

$\sum_{a \in A\text{-}Min} z_a \geq \text{Inf}(d_i,\ i \in Min(X))$, with *A-Min* = {$a \in A$, such that $a \subseteq Min(X)$}.

We denote by *LB**(*I*) this improved lower bound.

D. Interval Orders

A partially ordered set (*Z*, <) is an *interval order* if the elements *z* of *Z* may be represented as closed intervals [*o*(*z*), *d*(*z*)] of the real line, in such way that, for any pair *z*, *z*' in *Z*:

- *z* < *z*' if and only if *d*(*z*) < *o*(*z*').

It is known (see [20]), that the partially ordered set (Z, <) is an interval order if and only if there does not exist *x*, *y*, *z*, *t* ∈ Z such that:

- *x* < *y* and *t* < *z*; (C4)
- there does not exist any other pair *u* < *v* with *u*, *v* ∈ {*x*, *y*, *z*, *t*} than the pairs in (C4) above.

Figure 1 below shows the forbidden pattern associated with interval orders:

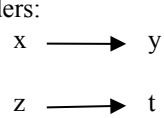


Fig. 1: Interval order forbidden pattern

If we consider now our Preemptive RCPSP instance *I* = (*X*, *K*, ≪), we see that:

Theorem 2: *If the partial order (X, ≪) is an interval order, then the oriented antichain graph (A, $E_{<<}$) is acyclic (does not contain any circuit).*

Proof: We suppose the converse, and consider some circuit Γ in (*A*, $E_{<<}$) with minimal length. Then we must distinguish two cases:

- *first case*: Length(Γ) = 2, which means that Γ contains two antichains *a* and *b*. Then we see that there must exist $i_1, j_2 \in a$, $i_2, j_1 \in b$ such that: $i_1 \ll j_1$ and $i_2 \ll j_2$. Then it becomes easy to check that i_1, j_1, i_2, j_2 define a forbidden pattern in the above sense, which induces a contradiction.
- *second case*: Length(Γ) ≥ 3, which means that Γ contains 3 consecutive antichains *a*, *b*, *c*, and that there must exist $x \in a$, y, $z \in b$, $t \in c$, such that $x \ll y$ and $z \ll t$. But we also deduce from the minimality of Length(Γ) and from the fact that *a*, *b*, *c* are antichains that *x*, *y*, *z*, *t* must define a forbidden pattern in the above sense, which induces again a contradiction. **End-Proof**.

This result will impact in a very significant way the design of the algorithm which will be presented in the next section. Clearly, if σ is a feasible schedule for the Preemptive RCPSP instance *I* = (*X*, *K*, ≪), it is possible to extend the precedence relation ≪ into an interval order $\ll_\sigma$, in such a way σ remains consistent with $\ll_\sigma$. In order to do it, we only need to set, for any activity pair *i*, *j* in *X*:

- *i* ≪* *j* iff End-Time(*i*) ≤ Start-Time(*j*).

Putting this last remark and Theorem 2 together makes appear that we only need, in order to deal with the Preemptive RCPSP instance *I*, to enumerate the extensions ≪* of the order relation ≪ which are interval orders. As a matter of fact, we may restrict ourselves to those extensions ≪* which are minimal for inclusion, that means which are such that there does not exist any extension ≪' of ≪ which is an interval order and which is such that: ≪' ⊂ ≪*, ≪' ≠ ≪*. So, next section is devoted to an accurate description of the way this enumeration process is performed.

IV. THE BRANCH/BOUND ALGORITHM INT-ORD-ENUM

A. A Reformulation of Preemptive RCPSP Instance I

Sections II and III lead us to reformulate Preemptive RCPSP instance *I* = (*X*, *K*, ≪) as follows:

> **Preemptive RCPSP Reformulation**: Compute an extension ≪* of the precedence relation ≪ which is an interval order and which is such that, if *z** is an optimal solution of the related LP $(P)_{Ant}$, obtained through Primal Simplex Algorithm and column generation, the optimal value 1.*z** is the smallest possible.

Remark: Clearly, program $(P)_{Ant}$ must be understood here with respect to ≪*and to the related *Antichain* set $A^* \subseteq A$.

So, our algorithm INT-ORD-ENUM is a Branch/Bound algorithm, which performs some enumeration of the exten-

sions $\ll^*$ of $\ll$. We must now specify the main components of such a tree search process, which are about:

- the extensions of Preemptive RCPSP instance $I = (X, K, \ll)$ which define the nodes of the related search tree;
- the way branching is performed;
- the way bounding and related filtering are performed;
- the way constraint propagation is performed;
- the branching strategy;
- the way the whole algorithm is implemented.

B. The Nodes of the INT-ORD-ENUM Search Tree

A node of the search tree induced by a branch/bound algorithm is usually defined by a set of additional constraints imposed to the initial problem. In the case of the Preemptive RCPSP instance $I = (X, K, \ll)$, those constraints are:

- additional precedence constraints $i \ll j$;
- *anti-precedence* constraints $i \dashrightarrow j$: $i \dashrightarrow j$ means that $i \ll j$ is forbidden.

So, we may identify any node of the search tree with a pair $(Add_{\ll}, Add_{\dashrightarrow})$, where $Add_{\ll}$ and $Add_{\dashrightarrow}$ are respectively the sets of additional precedence constraints and anti-precedence constraints which constrain $\ll^*$ as follows:

- $(\ll \cup Add_{\ll}) \subseteq \ll^*$;
- $(Add_{\dashrightarrow} \cap \ll^*) = \text{Nil}$.

Explanation of the *anti-precedence* constraints: it must be understood that those constraints have sense only with respect to the reformulation of subsection IV.*A*. That means that they are not going to play any role either with respect to an eventual feasible schedule or with respect to the program $(\mathsf{P})_{\mathsf{Ant}}$, but that they only will impact the way additional precedence constraints may be added to the initial ones.

Clearly, if current order relation happens to define an interval order, the related node is a terminal node (a leaf).

C. The Branching Mechanism

If current precedence relation, which is managed in such a way it always remains transitive, is not an interval order, then it must contain some forbidden pattern i_1, j_1, i_2, j_2:

- $i_1 \ll j_1$ and $i_2 \ll j_2$; (C5)
- no other pair $u \ll v$ exists with $u, v \in \{ i_1, i_2, j_2, j_1\}$ than the pairs in (C5) above.

This forbidden pattern allows us to perform a binary branching process by successively considering the 2 following alternatives:

- 1 th alternative (1 th son): insert $i_1 \ll j_2$ into $Add_{\ll}$;
- 2 th alternative (2 th son): insert $i_2 \ll j_1$ into $Add_{\ll}$ and insert $i_1 \dashrightarrow j_2$ into $Add_{\dashrightarrow}$;

D. Lower Bound, Upper Bound and Related Filtering

Lower Bound: The lower bound which derives from a current node defined by a pair $(Add_{\ll}, Add_{\dashrightarrow})$, is provided by the optimal value of the program $(\mathsf{P})_{\mathsf{Ant}}$, where valid antichains are considered as deriving from $(\ll \cup Add_{\ll})$. This problem is handled through column generation, as explained in Section III.*C*, and the Pricing problem PRICE(λ) is handled while using the ILP model of Section III.*C*. Every column which has been generated at some time during the process is kept into memory.

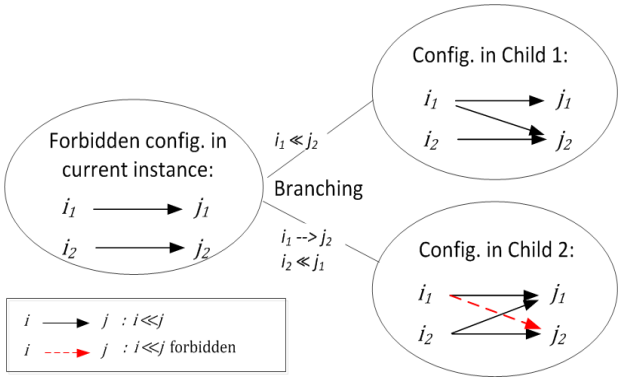


Fig. 2: the Branching Mechanism

Upper Bound: Also, we make in such a way that we are provided, as part of a pretreatment, with an initial upper bound *UB*: in order to get this initial upper bound, we apply to instance *I*, a greedy randomized algorithm designed for the Non Preemptive RCPSP (see [19]) and which, in case of 30 activity PSPLIB instances, approximates the optimal Non Preemptive RCPSP optimal value by less than 2% in average. Of course, *UB* is updated as soon as some feasible solution is computed by the INT-ORD-ENUM search process.

Related Filtering: Of course, if the optimal solution z^* of LP $(\mathsf{P})_{\mathsf{Ant}}$, is such that the subgraph $(A(z^*), E_{\ll})$ does not admit any circuit, we consider that we have been reaching some terminal node of the search tree. In case related value $1.z^*$ is smaller than the value of the current solution (current upper bound *UB* of the forthcoming section IV.*E*, we update this current solution as a feasible schedule σ such that $z^* = z^*(\sigma)$. We notice that it is sufficient to consider the subgraph $(A(z^*), E_{\ll})$ as defined with respect to initial precedence relation $\ll$, since our goal is to make possible turning a solution of $(\mathsf{P})_{\mathsf{Ant}}$, into a feasible schedule in the sense of initial precedence relation $\ll$.

E. Constraint Propagation

We apply several kind of inference rules $\alpha \models \beta$, whose semantics come as follows:

- α (precondition part) denotes constraints which are already associated with current node *S* of the three search;
- β (consequent part) denotes the additional relations which have to inserted into sets $Add_{\ll}$ and $Add_{\dashrightarrow}$..

The first class of rules deals with transitivity, and makes in such a way that, at any time during the process, current relation $\ll^* = (\ll \cup Add_{\ll})$ remains transitive:

Rule 1: $i \ll^* j, z \ll^* i \models z \ll^* j$;
Rule 1': $i \ll^* j, j \ll^* z \models i \ll^* z$;

Of course, any relation $i \ll^* i$ induces a *Failure* signal.

The second one deals in a classical way with largest paths and current upper bound *UB*. We add two dummy activities: *s* (source) and *p* (sink) defined as usual and, at every time

during the process, we are provided, for every activity i, with:

- $\pi(i)$ = earliest finish time for i, which means the length of a largest path from s to i;
- $\Pi(i)$ = the length of the largest path between the beginning of i and p: $\Pi(i)$ = UB - LS(i) where LS(i) is the latest starting time f*or i*

Doing this allows us to implement the following classical inference rules, which tend to keep the current precedence relation ($\ll \cup Add_{<<}$) from inducing the existence of a largest path with length $\geq UB$:

Rule 2: $\pi(i) = \alpha$, $i \ll * y$ *and* $\alpha+d_y > \pi(y) \models \pi(y) = \alpha+d_y$;

Rule 2': $\Pi(i) = \alpha$, $y \ll * i$ *and* $\alpha+d_y > \Pi(y) \models \Pi(y) = \alpha+d_y$;

Rules 2 and 2' update values $\pi(y)$ and $\Pi(y)$, $y \in X$, as soon as necessary. Of course, the existence of any path with length $\geq UB$ induces a Failure signal.

Rule 3: $\pi(i) = \alpha$, $\alpha + \Pi(y) > \text{UB} \models i$ ---> y ;

Rule 3': $\Pi(i) = \alpha$, $\alpha + \pi(y) > \text{UB} \models y$ ---> i;

Rules 3 and 3' forbid any additional precedence relation which would induce the existence of a largest path with length $\geq UB$ to be inserted into $Add_{<<}$.

Rule 4: $\pi(i) = \alpha$, $UB - \Pi(y) + d_y \leq \alpha - d_i \models y \ll * i$;

Rule 4': $\Pi(i) = \alpha$, $UB - \alpha + d_i \leq \pi(y) - d_y \models i \ll * y$;

Rules 4 and 4' insert into $Add_{<<}$. additional precedence relations which should be satisfied in any schedule with makespan no more than UB.

The last class of rules deals with the forbidden patterns of Section III.*D*, and aims at keeping current relation ($\ll \cup Add_{<<}$) from containing any such a pattern:

Rule 5: $i \ll * j$, $z \ll * t$ *and* z ---> $j \models i \ll * t$;

Rule 5': $i \ll * j$, $t \ll * z$ *and* i ---> $z \models t \ll * j$;

Rule 6: i ---> j, $z \ll * j$ *and* $i \ll * t \models z \ll * t$;

Rule 6': i ---> j, $i \ll * z$ *and* t ---> $z \models t$ ---> j.

We see here the true role of constraints i ---> j, which help us in inserting additional precedence constraints into the $Add_{<<}$ set, with a strong impact on the antichain set $A*$ and on the optimal value of the related linear program $(P)_{Ant}$,. Of course, any time such a pattern appears, it induces a Failure signal.

F. Branching Strategy

We described in IV.*B* the Branching mechanism, which relies on the extraction of some forbidden pattern i_1, j_1, i_2, j_2:

- $i_1 \ll j_1$ and $i_2 \ll j_2$; (C5)
- no other pair $u \ll v$ exists with $u, v \in \{ i_1, i_2, j_2, j_1\}$ than the pairs in (C5) above.

Since it is known that the way branching parameters are chosen is a critical issue as soon as Branch/bound and constraint propagation are performed. So we must now specify the strategy which is used here in order to compute a well-fitted 4-uple i_1, j_1, i_2, j_2.

As a matter of fact, we apply here the well-known "most constraint variable" principle, and focus on the shortest circuits of the subgraph (A(z*), $E_{<<}$) and on the antichains in A(z*) which are the most involved in those circuits.

As told in Section IV.*D*, branching has to be performed only if there exists some circuit in the subgraph (A(z*), $E_{<<}$), where z* is the optimal solution of the LP $(P)_{Ant}$, solved after constraint propagation has been performed. Then we distinguish two cases:

- **First case**: there exists a circuit with length 2. In such a case, circuits with length 2 define in a natural way a non oriented graph (A(z*), F) on the set A(z*): two antichains a, a' in A(z*) define an edge of this graph if they also define a circuit of the oriented graph (A(z*), $E_{<<}$), We consider an antichain a_0 which is with maximal degree $D_F(a_0)$ in the graph (A(z*), F), together with some antichain a_1, with maximal degree $D_F(a_1)$ among the F-neighbours of a_0. Then we derive the forbidden pattern i_1, j_1, i_2, j_2, according to the proof of Theorem 2 in Section III.*D* and to Figure 3 (a).
- **Second Case**: there does not exist any circuit with length 2. Then we compute the largest strongly connected component A_0 of the oriented graph (A(z*), $E_{<<}$), together with the antichain a_0, which is such that:
  - There exists at least one pair a_1, a_2, such that (a_1, a_0) and (a_0, a_2) are in the arc set $E_{<<}$, while (a_1, a_2) $\notin E_{<<}$;
  - The sum $D^-_F(a_0) + D^+_F(a_0)$ of the inner and outer degrees of a_0 in the subgraph (A_0, $E_{<<}$) induced from by A_0 is maximal.

Finally we compute some circuit Γ which contains a_0 as well as a_1, a_2 above and which is with minimal length, and we derive the forbidden pattern i_1, j_1, i_2, j_2, according to the proof of Theorem 2 in Section III.*D* and to Figure 3 (b).

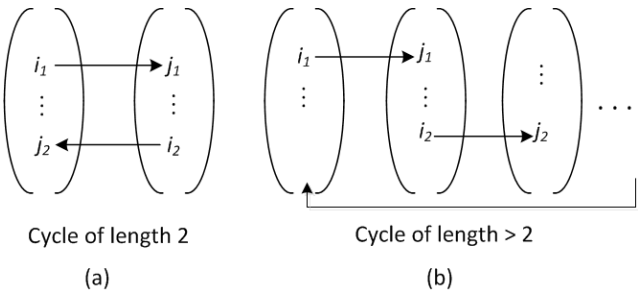


Fig. 3: Extracting a forbidden pattern

G. Implementation

The global Branch/bound algorithm INT-ORD-ENUM Branch/Bound is implemented according to a Breadth First Search strategy which may be summarized as follows:

***INT-ORD-ENUM** Algorithmic Scheme*.

1. **Pretreatment**: Compute a feasible Non Preemptive RCPSP schedule σ, while using a greedy randomized insertion flow heuristic as in [19]. Derive an upper bound UB, together with an initial antichain subset $B \subset A$, such that the linear program $(P)^B_{Ant}$ admits a feasible solution; Initialize the breadth search node list L as the list {($Add_{<<}$ = Nil, $Add_{--->}$ = Nil)};

2. **Main Process: Breadth First Tree Search**.
   Let L be the current node list, ordered according to LP $(\mathsf{P})^{B}_{\mathsf{Ant}}$ related values, and S be the first node in L; S is defined by two additional constraint sets $Add_{<<}$ and $Add_{--->}$; Delete S from L; Perform Constraint Propagation and extend $Add_{<<}$ and $Add_{--->}$; If *Failure* then go back to 2. Else go to 3.;
3. Solve the LP $(\mathsf{P})^{B}_{\mathsf{Ant}}$ related to S through column generation and test the oriented graph $(A(z^*), E_{<<})$ deriving from the obtained optimal solution z^*; If $1.z^* \geq$ UB then go to 2. Else go to 4.;
4. If the graph $(A(z^*), E_{<<})$ is acyclic then derive from z^* a feasible schedule σ, update the upper global bound UB and go back to 2. Else go to 5.;
5. Compute branching parameters i_1, j_1, i_2, j_2, according to Section IV.*F* and create both related children:
   - 1 th son: insert $i_1 << j_2$ into the set $Add_{<<}$;
   - 2 th son: insert $i_2 << j_1$ into the set $Add_{<<}$; and i_1 ---> j_2 into the set $Add_{--->}$;

   Insert those two children nodes in L, according to their related LP $(\mathsf{P})^{B}_{\mathsf{Ant}}$ value; Go back to 2.;

Process ends as soon as the LP value related to the first element of S is no smaller than UB. Then current value UB provides the optimal makespan value;

This algorithm is implemented in C++, and linear programs $(\mathsf{P})^{B}_{\mathsf{Ant}}$ and *L-PRICE(λ)* are handled by CPLEX.12 linear solver. But the global INT-ORD-ENUM process is embedded into the SCIP framework for branch cut and price algorithms [16]. The SCIP framework consists in a template library which implements through breadth first search generic branch and bound schemes involving Linear Programming together with pricing scheme. In the present case, what we mainly had to do was providing the C++ procedures which performed, for every node S, the construction of the $(\mathsf{P})^{B}_{\mathsf{Ant}}$ and *L-PRICE(λ)* programs, the constraint propagation process and the branching strategy, and assembly them inside SCIP.

H. An Example

Let us consider an instance of 6 activities and 1 resource. Each activity has a duration equal to 1 and a resource requirement equal to 1. The precedence constraints are given by the following precedence graph:

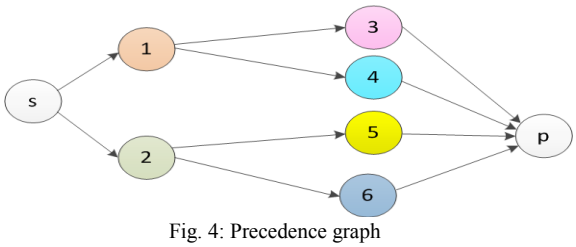


Fig. 4: Precedence graph

We initialize the set of antichains with the 6 singleton antichains. The tree constructed by our method and the branching decisions are given as below:

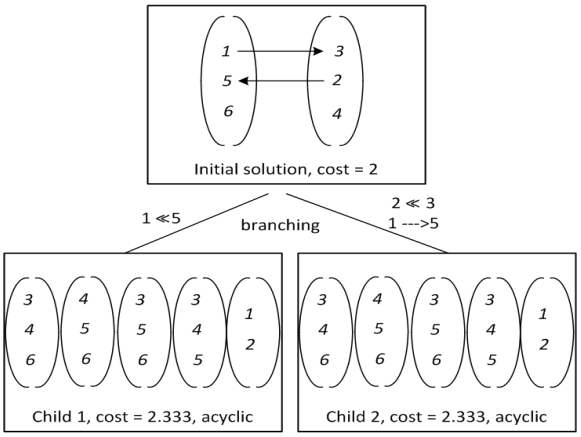


Fig. 5: Solving instance of figure 4

The resulting optimal solution is given according to the following Gantt chart.

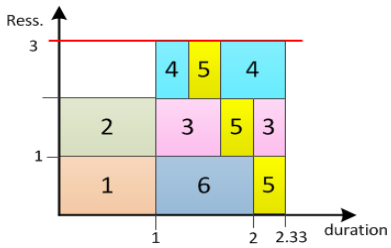


Fig. 6: Gantt chart of optimal preemptive solution

V. Numerical Experiments

Experiments were carried on in C++, on linux CentOS proc. Intel(R) Xeon(R) 2.40GHz. The instances which we used were PSPLIB instances ([17]).

Our main achievement here was to solve in an exact way and in a rather short time (never more that 95 CPU seconds) Preemptive RCPSP on all 30 activity instances of the PSPLIB library, which had been, until now, never done. This first experiment is described in coming section V.*A*.

Also, we could get an evaluation of the bounding process related to linear program $(\mathsf{P})^{B}_{\mathsf{Ant}}$, and check that in average, $LB(I)$ approximates the optimal Non Preemptive RCPSP optimal value by less than 6%. By the same way, we checked that the augmented lower bound $LB^*(I)$ hardly improve $LB(I)$ by less than 0.5%.

Finally, though we were not able to handle in an exact way all 60/120 activity instances of the PSLIB library, we could derive from the instances which we were able to handle new lower bounds for several Non Preemptive RCPSP instances of the PSPLIB library. This second part of the experiment will be described in Section V.*B*.

A. Exact results on j30

The columns of table I have the following meaning:

- *No Preemp. opt*.: optimal value for non preemptive RCPSP (available in PSPLIB website)
- *Preemp. opt*.: optimal for preemptive RCPSP (our results)
- *#nodes*: number of nodes created (0 means optimal value was found by heuristic in preprocessing and prooved to be optimal by the first constraint propagation)
- *cpu* (*s*): cpu time in seconds

TABLE I
RESULTS ON J30 INSTANCES OF PSPLIB

	No Preemp. opt.	*Preemp. opt.*	*#nodes*	*cpu*(s)
Mean	58.99	58.07	72.73	2.04
Min	34.00	34.00	0.00	< 0.1
Max	129	129	2130	94.11
std dev.	14.09	13.80	214.87	8.03

B. Comparative Analysis of LB(I), MB(I)*

The following table II provides us with average values for the 480 instances of PSPLIB with 30 activities:

TABLE II: EVALUATION OF THE BOUND *LB*(*I*)

Mean LB(*I*)	*Mean LB**(*I*)	*Mean Premp. opt.*	*Mean No Premp. opt.*
56.73	56.79	58.07	58.99

Remark: in almost 50% of the cases (exactly 236 instances among 480), the values *LB*(*I*), *Premp. Op*. and *No Premp*. *Opt*. coincide.

C. New best lower bounds

Our method gives new best lower bound for j60, j90 and j120 instances (in a limit of time of 3 hours).

The columns of table III have the following meaning:

- *best No preemp. UB*: best known upper bound for no preemptive RCPSP (available in PSPLIB website)
- *Preemp. LB*: lower bound for preemptive RCPSP (our method)
- *deduced no preemp. LB:* lower bound for no preemptive RCPSP which we deduce from *Preemp. LB*
- *Best known LB:* the best known lower bound currently available in PSPLIB website and updated with the recent results of [18].

TABLE III
NEW BEST LOWER BOUNDS

instance	best no preemp. UB	Preemp. LB	deduced no preemp. LB	Best known LB
j6013_1.sm	112	106.41	107	105
j6029_2.sm	133	126.20	127	123
j6029_3.sm	121	117.29	118	115
j6029_4.sm	134	129.29	130	126
j6029_5.sm	110	104.04	105	102
j6029_6.sm	154	145.30	146	144
j6029_7.sm	123	116.00	116	115
j6029_9.sm	112	106.83	107	105
j6045_1.sm	96	91.00	91	90
j6045_2.sm	144	137.32	138	134
j6045_3.sm	143	137.50	138	133
j6045_4.sm	108	102.49	103	101
j6045_5.sm	106	100.41	101	100
j6045_6.sm	144	136.42	137	132
j6045_7.sm	122	116.04	117	113
j6045_8.sm	129	122.17	123	119
j6045_9.sm	123	118.20	119	114
j6045_10.sm	114	106.48	107	104
j9045_6.sm	175	163.26	164	163
j12036_4.sm	236	217.35	218	217
j12056_3.sm	241	222.12	223	220
j12056_4.sm	222	206.62	207	205
j12056_5.sm	280	261.80	262	261
j12056_7.sm	283	263.29	264	260
j12056_8.sm	289	268.04	269	265
j12056_9.sm	288	266.34	267	264

VI. CONCLUSION

Our method is very efficient. Besides exactly solving small size Preemptive RCPSP instances, it is also able to provide us with very good lower bounds for larger scale Non Preemptive RCPSP. We are looking for adapting this method to the non preemptive RCPSP.

REFERENCES

[1] R. Kolisch, R. Padman. Deterministic project scheduling, *Omega*, 48, pp. 249-272 (1999)

[2] P. Brucker, A. Drexl, R. Mohring, K. Neumann, E. Pesch. Resource-constrained project scheduling: notation, classification, models and methods, *EJOR* 112, pp. 3-41 (1999)

[3] W. Herroelen. Project Scheduling-Th./Pract., *Prod./Op. Management*, 14, 4, pp. 413-432 (2006)

[4] S.S. Liu, C.J. Wang. RCPSP profit max with cash flow, *Aut. Const.* 17, pp. 966-74 (2008)

[5] S. Hartmann, D. Briskorn. A survey of variants of RCPSP. *EJOR* 207, pp.1-14 (2010)

[6] M.J. Orji, S. Wei. Project Scheduling Under Resource Constraints: A Recent Survey. *International Journal of Engineering Research & Technology (IJERT)* Vol. 2 Issue 2, (2013)

[7] E. Demeulemeester and W. Herroelen. An efficient optimal solution procedure for the preemptive resource-constrained project scheduling problem. *European Journal of Operational Research*, 90, pp. 334–348 (1996).

[8] S. Verma. Exact methods for the preemptive resource-constrained project scheduling problem, research and publication, *Indian institute of management* 2006, ahmedabad, india,w.p.no. 2006-03-08.

[9] B.A. Nadjafi, S. Shadrokh. The preemptive resource-constrained project scheduling problem subject to due dates and preemption penalties: an integer programming approach, *Journal of Industrial Engineering*, 1 pp. 35-39 (2008)

[10] F. Ballestin, V. Valls, S. Quintanilla, Preemption in resource- constrained project scheduling, *European Journal of Operational Research*, 189 pp.1136-1152 (2008)

[11] M. Vanhoucke, D. Debels, The impact of various activity assumptions on the lead time and resource utilization of resource- constrained projects, *Computers and Industrial Engineering*, 54 pp.140-154J (2008)

[12] M. Vanhoucke, A genetic algorithm for the net present value maximization for resource constrained projects, EVOComp; *LNCS* 5482 pp. 13-24 (2009)

[13] A. Mingozzi, V. Maniezzo, S. Ricciardelli, and L. Bianco. An exact algorithm for the resource-constrained project scheduling based on a new mathematical formulation. *Management Science*, 44 pp. 714–729, (1998).

[14] A. Damay, A. Quilliot, E. Sanlaville. Linear programming based algorithms for preemptive and non preemptive RCPSP, *EJOR*, 182, 3, pp. 1012-1022 (2007)

[15] A. Mehrotra, and M. A. Trick. A column generation approach for exact graph coloring, INFORMS Journal on Computing, 8:4, pp. 133-151 (1996)

[16] http://scip.zib.de/

[17] http://www.om-db.wi.tum.de/psplib/

[18] Andreas Schutt, Thibaut Feydy, Peter J. Stuckey Explaining Time-Table-Edge-Finding Propagation for the Cumulative Resource Constraint. *Lecture Notes in Computer Science* Volume 7874, pp. 234-250 (2013) (last results on http://ww2.cs.mu.oz.au/~pjs/rcpsp/)

[19] A.Quilliot, H.Toussaint: Flow Polyedra and Resource Constrained Scheduling Problem, *RAIRO-RO*, 46-04, p 379-409, (2012)

[20] F.S.Roberts: *Discrete Maths Models*; Prentice Hall, Englewood Cliffs, N.Y, (1976).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 329–336

A Beam Search Based Algorithm for the Capacitated Vehicle Routing Problem with Time Windows

Hakim Akeb
ISC Paris Business School
22 Bd du Fort de Vaux
75017 Paris, France
Email: hakeb@iscparis.com

Adel Bouchakhchoukha[(1)(2)]
[(1)]MSE, Université Paris 1 Panthéon Sorbonne
106–112 Bd de l'Hôpital
75013 Paris, France
Email: adel.bouchakhchoukha@
malix.univ-paris1.fr

Mhand Hifi[(1)(2)]
[(2)]Université de Picardie Jules Verne
UR EPROAD, Équipe ROAD
7 rue du Moulin Neuf
80039 Amiens, France
Email: mhand.hifi@u-picardie.fr

***Abstract*—In this paper the capacitated vehicle routing problem with time windows is tackled with a beam-search based approximate algorithm. An instance of this problem is defined by a set of customers and a fleet of identical vehicles. A time window is associated with each customer and a maximum capacity characterizes a vehicle. The aim is then to serve all the customers by minimizing the number of vehicles used as well as the total distance and by respecting the time windows.**

The proposed method follows three complementary phases: (i) dividing the set of customers into disjunctive clusters, (ii) determining a feasible solution in each cluster by using beam search, and (iii) applying a local search in order to improve the quality of the solutions. The proposed method is analyzed computationally on a set of benchmarks due to Solomon. Encouraging results have been obtained.

I. Introduction

THE *Vehicle Routing Problem* (VRP) is well known and well studied in the literature. It consists, in its simplest version, to visit or deliver a set $N = \{1, ..., n\}$ of customers by using a fleet of m vehicles $V = \{v_1, ..., v_m\}$. The objective is often to minimize the number of vehicles m as well as the sum of distances traveled by these ones. Note that if $m = 1$ then the problem becomes the *Traveling Salesman Problem* (TSP). Is some cases, a time window $W_i = [e_i, l_i]$ is associated to customer i, where e_i is the earliest time to begin the service of this customer and l_i the latest time. Parameters e_i and l_i are also known as *ready time* and *due date* respectively. The aforementioned problem is known as the *Vehicle Routing Problem with Time Windows* (VRPTW). A service time s_i can be associated with customer i, this means that the arrival time at this customer must be at most $l_i - s_i$. Furthermore, each customer $i, i \in N$ has a demand d_i and a capacity may be associated with each vehicle denoting the maximum of the sum of quantities that can be put inside the corresponding vehicle. Such a problem is called *Capacitated VRPTW* (CVRPTW).

VRP was addressed by many authors and several methods and strategies were proposed to solve its different versions. These methods can be categorized into two categories: exact methods and approximate ones. In the first category, Azi et al. [3] proposed an exact algorithm, based on column generation and branch-and-price, to solve VRPTW including multiple use of vehicle, i.e., a given vehicle may be associated with several routes. The same authors [2] proposed, several years before, another exact algorithm for a single vehicle routing problem with time windows and multiple routes. Baldacci et al. [4] employed branch-and-price in order to solve a capacitated vehicle routing problem (CVRP) by using an integer programming formulation. New lower bounds were presented and an algorithm to find the optimal solution for CVRP was given. Baldacci and Maniezzo [5] proposed exact methods based on node-routing formulations to tackle the undirected arc-routing problems. Feillet et al. [11] developed an exact algorithm for the elementary shortest path problem with resource constraints where the authors indicated an application to some vehicle routing problems.

The second category of methods consists to search for approximate solutions by using essentially heuristics and metaheuristics. Solomon [19] proposed different algorithms in order to solve the vehicle routing and scheduling problems with time window constraints. A two-stage heuristic including ejection pools was for example proposed by Lim and Zhang [14] in order to tackle VRPTW. Chen et al. [9] proposed a heuristic that combines mixed integer programming and a record-to-record travel algorithm in order to solve approximately the *split delivery vehicle routing problem*, a variant of CVRP where a customer may be served by more than one vehicle. Tan et al. [22] proposed several heuristic methods, including simulated annealing, to solve VRPTW . Insertion heuristics were proposed by Campbell and Savelsbergh [7] for vehicle routing and scheduling problems. Chao et al. [8] proposed a fast heuristic for the orienteering problem, i.e., a vehicle routing problem where a profit is associated with each customer and the objective is to visit a subset of customers in order to maximize the total benefit and by respecting some constraints. Pisinger and Ropke [16] developed a general heuristic for vehicle routing problems able to solve five different variants of VRP, including CVRP and VRPTW. A genetic algorithm

978-83-60810-53-8/$25.00

was proposed in [12], [21], and [17] for the VRPTW. Ant colony optimization is very effective and was adapted for the various variants of VRP (see for example [13] where the *open vehicle routing problem* was considered). Finally, tabu search was considered by Cordeau et al. [10] for solving VRPTW and by Brandão and Mercer [6] for solving the multi-trip vehicle routing and scheduling problem, i.e., the case where a vehicle may perform several trips.

In this work, we propose a three-phase algorithm for solving CVRPTW. The first phase consists to divide the set of customers into m clusters. After that, at phase 2, the shortest path that visits each customer once in each cluster is computed by using beam search. Each path must verify the problem constraints, i.e., must not violate the time window. In the third and last phase, a local search is applied on the solution in order to try to decrease the total distance traveled by the m vehicles.

II. Problem Statement and Mathematical Formulation

The problem to solve (CVRPTW) consists to visit n customers $i \in N = \{1, ..., n\}$, where each customer (or *vertex*) has coordinates (x_i, y_i) in the Euclidean plan. The float used contains m identical vehicles $V = \{v_1, ..., v_m\}$, each with the same maximal capacity $C_{\max}$. All the vehicles start their travel at the *depot* D, whose coordinates are (x_D, y_D) in the Euclidean plan, visit a set of distinct customers, and returns to the depot.

In addition, a demand d_i is associated with each customer $i \in N$, d_i has the same unit of measurement as the vehicle capacity, this may be a volume or a weight for example. Customer i must be served at time t_i that may be the earliest time after its ready time (e_i) and no later than its due date l_i, this corresponds to the time window $W_i = [e_i, l_i]$ associated with customer i ($l_i - e_i$ is called the *width* of the time window). A service time s_i is also defined for customer i and is equal to the time spend to serve the customer.

Each vehicle $v_j, 1 \leq j \leq m$ performs then a route R_j by visiting a number of customers or vertices. These ones correspond to a cluster denoted by $\mathcal{C}_j$. Let also denote by $\mathcal{D}_j$ the sum of distances covered by vehicle v_j. Note that the distance between to customers i and j, denoted by $dist_{ij}$ corresponds to the euclidean distance between these two points, i.e., $dist_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

CVRPTW can then be formulated as follows:

$$\min\ m \tag{1}$$

$$\min \sum \mathcal{D}_j,\ 1 \leq j \leq m \tag{2}$$

subject to

$$\sum_{i \in C_j} d_i \leq C_{\max},\ \ \forall i \in C_j \tag{3}$$

$$t_i \in W_i = [e_i, l_i]\ \ \forall i \in N \cup \{D\} \tag{4}$$

Equation 1 represents the first objective to minimize, that is the number of vehicles to use. Equation 2 is the second objective to minimize and corresponds to the total distance traveled by all vehicles. The first constraint is indicated in (3) and ensures that the sum of demands for each cluster C_j is at most equal to the vehicle capacity $C_{\max}$. The second constraint (4) means that each customers i must be served inside its time window $W_i = [e_i, l_i]$, i.e., $e_i \leq t_i \leq l_i,\ \forall i \in N \cup \{D\}$. Then the depot can be considered as a customer for which the demand is null $(d_D = 0)$ and is the only point that is visited twice. The time window $W_D = [e_D, l_D]$ associated to the depot defines the *scheduling horizon* and means that each vehicle cannot leave the depot before its opening (at time e_D) and must return to the depot before its closure (at time l_D).

III. A three-phase method for solving CVRPTW

It is well known that many methods for solving vehicle routing problems often contain three phases (steps):

P1. Consists to divide the set of customers into m disjoint clusters, i.e., $C_1, ..., C_m$ such that $C_i \cap C_j = \emptyset$ for $1 \leq i < j \leq m$.

P2. Apply a given method in order to compute the shortest path, or a path of maximum benefit inside each cluster.

P3. Try to improve the solution obtained after P2, by applying another method such as local search.

A. Clustering

The first phase in solving our problem (Capacitated Vehicle Routing Problems with Time Windows) consists to divide the n customers into m disjoint sets or clusters. Then each vehicle will visit all the customers in the clusters that is assigned to it. Fig. 1 shows an example where a set containing 16 points and a depot (D) is divided into three disjoint clusters $\{C_1, C_2, C_3\}$.

There exists several methods for clustering and many of them are based on the dispersion (distribution) of the point around a central point called *centroid*. In our case we choose the well-known *k-means* method in order to compute the clusters. Actually, this is an adaptation of k-means in order to compute m clusters each of total capacity smaller than of equal to the capacity $C_{\max}$ of the vehicle.

Algorithm 1 explain how procedure k-means works. It receives the set N of customers as input parameter. The procedure's output corresponds to m disjoint clusters respecting

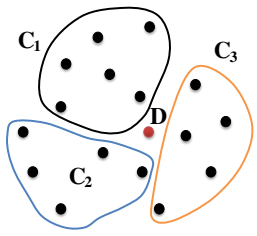


Fig. 1. An example of clustering a set of points into three disjoint clusters.

the vehicle capacity constraint. k-means begins by choosing m random points $c_1, ..., c_m$ from N (line 4), each point $c_j, (1 \leq j \leq m)$ corresponds to a centroid (belonging to cluster C_j). After that (at line 5), each not-assigned yet point $i \in N$ is assigned to the nearest centroid c_j of coordinates (x_{c_j}, y_{c_j}) in the sense euclidean distance. This means that point i of coordinates (x_i, y_i) is assigned to cluster C_j that minimizes the euclidean distance $\sqrt{(x_i - x_{c_j})^2 + (y_i - y_{c_j})^2}$, for $1 \leq j \leq m$.

After that, in the **while** loop that begins at line 6, the coordinates of the m centroids are recomputed (line 7). This is done by assigning to each $c_j, (1 \leq j \leq m)$ the center of the points belonging to cluster j. More precisely: $x_{c_j} = \frac{1}{|C_j|} \sum x_{c_k}$ and $y_{c_j} = \frac{1}{|C_j|} \sum y_{c_k}$ for all points $c_k \in C_j$. Each point $i \in N$ is after that assigned to the nearest centroid (among the new computed centroids), this is done in line 8. At line 9, if no point has moved from a centroid to another one, then a stable clustering is obtained: variable $move$ is set to the value $false$ (line 10) in order to stop recomputing of the centroids. Otherwise, this means that at least one point has moved, then instructions in lines 7–10 are repeated until a stable configuration is obtained. After that, the procedure verifies that the total capacity of each cluster does not exceed the capacity $C_{\max}$ of the vehicles. If so, then variable $found$ is set to "true" (line 14) in order to stop the procedure and return m clusters respecting the capacity constraint. If at least one cluster violates the capacity constraint, then the procedure restarts with m other random points by using the first **while** loop (line 2).

Note that the optimal number of clusters (m) is not known in the general case. So a dichotomous search can for example be used in order to test several values and determine the best one. Of course, increasing the value of m increases the probability to find clusters respecting the capacity constraint. For well-studied benchmarks, e.g. those proposed by Solomon [19], the same best value of m was found by many authors, so this value can be fixed in advance in order to save computation time.

B. Beam search for computing the shortest paths (routes)

The second phase takes place after the m clusters were generated by the k-means procedure (algorithm 1). The objective of the second phase is to compute the shortest path in each cluster. A path correspond to a route beginning at the depot D, visiting exactly once each point (customer) are returning after that to the depot. In addition, one vehicle is associated to each cluster. Fig. 2 shows an example of solution for the example indicated in figure 1.

It is to note that the time windows associated to each customer as well as the capacity of the vehicle make the problem hard to solve, harder than the traveling salesman problem (TSP) in which there are no time windows and no limit to the vehicle capacity.

Remember that the objective in CVRPTW is to minimize the sum of distances traveled by the m vehicles. In order to compute the shortest paths, we propose to use beam search on each cluster. Beam search is a tree search and is a modified version of the well known branch-and-bound method.

Algorithm 1 Procedure k-means for CVRPTW

Require: Set N containing n points (customers);
Ensure: m disjoint clusters each of total capacity $\leq C_{\max}$;
1: $found \leftarrow$ false;
2: **while** ($found =$ false) **do**
3: $move \leftarrow true$;
4: Choose randomly m distinct points $\{c_1, ..., c_m\}$ from N. Let these points be the m centroids (clusters);
5: Assign each point $i \in N$ to the nearest centroid c_j, $1 \leq j \leq n$;
6: **while** ($move$ = true) **do**
7: Recompute the coordinates (x_{c_j}, y_{c_j}) of each centroid, i.e., each c_j becomes the center of the points assigned to that centroid;
8: Assign each point $i \in N$ to the nearest centroid c_j, $1 \leq j \leq n$;
9: **if** no point has moved from a cluster to another one **then**
10: $move \leftarrow$ false;
11: **end if**
12: **end while**
13: **if** the capacity of each a cluster $\leq C_{\max}$ **then**
14: $found \leftarrow$ true;
15: **end if**
16: **end while**

Beam search was used to solve different combinatorial problems, such as Scheduling [15] and Cutting–and–Packing problems [1]. In its width–first implementation, the method starts by creating the root node which may contains an initial (starting) partial solution. After that, each node at level ℓ generates a set of descendants, these correspond to level $\ell+1$. Each node of the new level is then evaluated by using an evaluation criterion and only a subset containing the ω best nodes are retained, the other nodes are discarded. Parameter ω is known as the *beam width*. If a node contains a final solution, then this one is evaluated and stored. The corresponding node is after that deleted because no branching is possible from it (leaf). The beam search stops when no branching becomes possible from any node of the current level. The best solution, among the different solutions obtained, is then retained as the final result.

1) Content of a node in the search tree: Let $V = \{v_1, ..., v_{|C_j|}\}$ be the set of vertices (customers) in cluster C_j $(1 \leq j \leq m)$.

It is important to define clearly the content of a node in the beam search tree. Each node η_ℓ at level ℓ contains the following elements:

- The set of vertices (customers) already visited $V^+ = \{v_1^+, ..., v_\ell^+\}$.
- The set of vertices that have not yet been visited $V^- = V \backslash V^+$.
- The distance $dist$ corresponding to the length of the path

$D \rightarrow v_1^+ \rightarrow v_2^+ \rightarrow ... \rightarrow v_\ell^+$.

Note then that if a node corresponds to a complete solution, then the path obtained is $D \rightarrow v_1^+ \rightarrow ... \rightarrow v_{|C_j|}^+ \rightarrow D$ and the total distance is the sum of euclidean distances of the corresponding arcs in the path.

As a result, a node η_ℓ at level ℓ in the search tree can be designated by the elements described above, i.e, $\eta_\ell = \{V^+, V^-, dist\}$, where $|V^+| = \ell$ and $|V^-| = |C_j| - \ell$.

2) Selection criterion for the next customer to visit: As explained above, branching from a node η_ℓ (or more exactly from the last node v_ℓ^+ in the path under construction) consists to choose the successors of the vertex v_ℓ^+ among the vertices in V^-. The next vertex $v_i \in V^-$ may be for example the closest one to v_ℓ^+ in the sense of euclidean distance or the time window interval $[e_i, l_i]$. For example, for the two sets of instances examined in this work (see Section IV below), the next vertex to visit is the closest one in the sense of parameter e_i (the earliest time) in the time window. For beam search, all the successors v_i^- are ranked in increasing value of parameter e_i and then the ω first ones are chosen to create ω distinct branches.

Of course, others criteria were tested, including the latest time l_i and/or the distance between the current customer and the remaining customers to visit, but the experimentations showed that the criterion based on parameter e_i is the best one for the instances tested.

3) Algorithm Beam Search: Algorithm 2 explains how beam search works in order to compute a route. Note that the capacity constraint is not taken into account in the algorithm since the sum of the capacities of the vertices in each cluster is less or equal to the vehicle capacity. The capacity constraint is always respected after the clustering phase (see Section III-A), then only the time window constraint is checked.

Algorithm 2 receives three input parameters: the cluster C_j, i.e., the set of vertices or customers to visit (to serve), the value of the beam width ω, and the selection criterion ρ that will serve to sort the nodes at each level of the tree and then to determine the *best* ones according to this criterion. As output, the algorithm computes the best route (path) R_j of minimal distance beginning at the depot D, visiting each customer once, and then returning to the depot.

The root node η_0 of the search tree is created in line 1. This node contains the set of vertices already visited ,i.e. the depot D (so $V^+ = \{D\}$), and the set of vertices not already served $V^- = V$. Since no customer has been already visited, then the distance is equal to 0.

Set B (line 2) corresponds to the nodes at the current level of the search tree. Each node $\eta \in B$ contains a partial route (path) from the depot to a given customer (set V^+) as well as the set of remaining customers to visit V^-. The total distance to the current customer is also known since this distance is updated each time a new customer is visited and then added to the path (route). B is initialized to η_0 (line 4). Set B_{off} (line 3) contains the offspring nodes after branching from each node in B.

Algorithm 2 Beam Search for computing the shortest path in a cluster

Require: Cluster C_j, the beam width ω, and the selection criterion ρ;

Ensure: The best shortest route R_j starting from the depot D, visiting all the vertices of the cluster, and returning to the depot;

1: Let $\eta_0 \leftarrow \{\{D\}, V, 0\}$ be the root node;
2: Let B be the set containing the nodes at a given level of the tree;
3: Let B_{off} the offspring nodes (descendants of nodes in B);
4: $B \leftarrow \{\eta_0\}$;
5: $\ell \leftarrow 0$;
6: $\eta^* \leftarrow \eta_0$; (the best solution found)
7: $\eta^*.dist \leftarrow +\infty$; (best distance)
8: **while** $(B \neq \emptyset)$ **do**
9: Branch out of each node $\eta_{\ell_i} = \{V_i^+, V_i^-, dist_i\} \in B$ and create the offspring nodes B_{off} (each node in B_{off} must respect the time windows);
10: $\ell \leftarrow \ell + 1$;
11: **if** $(V_i^- = \emptyset$ for a node $\eta_{\ell_i} \in B_{\text{off}})$ **then**
12: Add vertex D (depot) to that node and compute the total distance;
13: **if** $(\eta_{\ell_i}.dist < \eta^*.dist)$ **then**
14: $\eta^* \leftarrow \eta_{\ell_i}$;
15: Remove η_{ℓ_i} from B_{off};
16: **end if**
17: **end if**
18: Sort the nodes in B_{off} according to parameter ρ and then keep only the $\min(\omega, |B_{\text{off}}|)$ first nodes, remove the other nodes from B_{off};
19: $B \leftarrow B_{\text{off}}$;
20: $B_{\text{off}} \leftarrow \emptyset$;
21: **if** there is a node $\eta_{\ell_i} \in B$ for which V_i^- contains a vertex with a violated time window **then**
22: Remove η_{ℓ_i} from B;
23: **end if**
24: **end while**

Since the current level ℓ is 0 (root node), then this is indicated in line 5, while the best solution η^* is initialized to the root node η_0 at line 6. The best distance $\eta^*.dist$ is set equal to $+\infty$ (line 7) because this value is to be minimized.

At line 8 the **while** loop starts. So at a given level ℓ of the tree, B contains at most ω distinct partial paths (routes) computed in parallel from the depot (root node). Then branching from a node η_{ℓ_i} (line 9) consists to explore the successors of the last visited vertex (customer) and to create as many nodes as there are successors with nonviolated time windows. So each node in B may have several descendants. Each descendant is then inserted into the set of offspring node B_{off}, that corresponds to level $\ell + 1$. This is why the level is incremented at the next line (10).

After that, at line 11, if there is a node in B_{off} in which all the customers were served $(V^- = \emptyset)$, then the complete

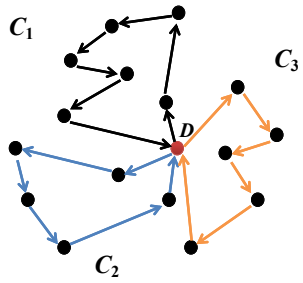


Fig. 2. An example of solution obtained after the second phase (beam search).

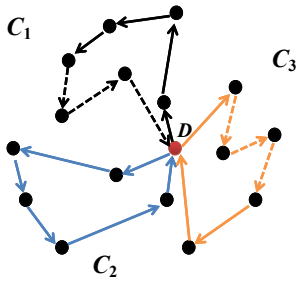


Fig. 3. A solution obtained after the third phase (local search).

solution is computed by adding the returning arc to the depot (line 12). The total distance for the obtained complete solution is then computed and compared to the best known one (line 13). If a lower distance is obtained then the best solution is updated (line 14) and the corresponding node is removed from B_{off} (line 15).

The most important instruction in beam search is certainly that given in line 18. Indeed, this step consists to sort the nodes according to the selection criterion ρ from the most important node to the least important one. Then the ω first nodes are kept and the other ones are removed from B_{off}. Note that if there are less than ω nodes in B_{off} then all the nodes are kept. After that set B_{off} is assigned to B and B_{off} reset to the empty set (lines 19–20). The last instruction in algorithm 2 consists to remove from B all the nodes that cannot lead to feasible solutions, i.e, that containing violated time windows.

The algorithm stops when set B becomes empty meaning that there is no node to explore or more precisely no customer to serve. Two cases can be distinguished: the algorithm has computed a feasible solution and this one is indicated in node η^* as well as the best corresponding distance, or there is no solution (if the distance in node η^* is equal to $+\infty$).

Fig. 2 shows an example of a solution that may be obtained after the second phase (beam search) on the example (clusters) shown in Fig. 1.

C. Local search for improving solution quality

In order to try to improve the result obtained after the second phase (beam search), a local search is performed on each cluster. This consists to execute the well-known 2-opt algorithm on each cluster (route).

2-opt is an iterative method that consists, at each iteration, to *break* two nonconsecutive arcs in the route and to link the four extremities in order to form another path and by respecting the time windows of course. The replacement is kept if the obtained solution is better.

The 2-opt method is given in algorithm 3. In each iteration, the algorithm examines each two distinct arcs $v_i \rightarrow v_{i+1}$ and $v_j \rightarrow v_{j+1}$ in the route R. These two arcs are replaced by the arcs $v_i \rightarrow v_j$ and $v_{i+1} \rightarrow v_{j+1}$ if and only if the distance decreases and the time windows are not violated. This process is repeated as long as there is improvement. Fig. 3 shows an example of improvement obtained by the 2-opt procedure on the solution of fig. 2. The arcs that had changed are indicated in dotted lines.

D. The three-phase algorithm (3PA) for solving CVRPTW

The three-phase algorithm based on clustering, beam search, and 2-opt local search is given in algorithm 4. It receives as input parameters the set of customers N, the depot D, and the number of vehicles. The algorithm's output corresponds to a set containing m feasible routes (respecting the constraints) of minimum distance, each one starts and ends at the depot D.

At line 1, the clustering phase (algorithm 1) is called in order to create m distinct clusters. The selection criterion (ρ), that serves to choose the next customer to serve is set at line 2. Then, algorithm 2 (beam search) is executed on each cluster (line 5), and this for several values of the beam width, i.e., for all values $\omega \in [1, ..., \omega_{\max}]$. The local search (algorithm 3) is then executed (line 6) on each solution computed by beam search.

Algorithm 3 2-opt algorithm

Require: A route $R = v_0 \rightarrow v_1 \rightarrow \ldots \rightarrow v_{|C_j|} \rightarrow v_{|C_j|+1}$;
Ensure: A route R' with a length at most equal to that of V;
1: $improvement \leftarrow$ true;
2: **while** ($improvement$ = true) **do**
3: $improvement \leftarrow$ false;
4: **for each** vertex $v_i \in R$ **do**
5: **for each** vertex $v_j \in R$ $(j \neq i-1, j \neq i+1)$ **do**
6: **if** $(\text{dist}(v_i, v_{i+1}) + \text{dist}(v_j, v_{j+1}) > \text{dist}\ (v_i, v_j) + \text{dist}(v_{i+1}, v_{j+1})$ **AND** the time windows will not be violated) **then**
7: Replace arcs $(v_i \rightarrow v_{i+1})$ and $(v_j \rightarrow v_{j+1})$ by arcs $(v_i \rightarrow v_j)$ and $(v_{i+1} \rightarrow v_{j+1})$;
8: $improvement \leftarrow$ true;
9: **end if**
10: **end for**
11: **end for**
12: **end while**

Algorithm 4 The three-phase algorithm 3PA for solving CVRPTW

Require: A set $N = \{1, ..., n\}$ of customers, the depot D, and m the number of vehicles (clusters).

Ensure: A set of routes minimizing the total distance and respecting the capacity and the time windows constraints.

1: Call the clustering phase (algorithm 1) and create m clusters $\{C_1, ..., C_m\}$ respecting the vehicle capacity constraint;
2: Define the selection criterion ρ;
3: **for each** cluster $C_j, (1 \leq j \leq m)$ **do**
4: **for** $\omega = 1$ to $\omega_{\max}$ **do**
5: Call algorithm 2: Beam-Search(C_j, ω, ρ);
6: Apply algorithm 3 (2-opt) on the solution returned by Beam-Search;
7: **end for**
8: **end for**

IV. Computational results

The proposed method is coded in C++ and the program run under Microsoft Windows environment on a computer with 2 GB of RAM and a 2.26 GHz Intel processor.

The algorithm was tested on two sets of instances, namely C1 and C2, proposed by Solomon. The characteristics of the two sets are summarized in table I. Each instance of each set contains 100 customers (column 2), they have also all the same service time (time needed to serve a customer) which is equal to 90 (column 4). The depot D has also the same coordinates for all the instances.

The first common characteristic between two distinct instances of the same set is the customer *demand*, i.e., the quantity to deliver to each customer. This value is fixed in each set for a given customer. More precisely, for two distinct instances in the same set (C1 or C2) each customer i has the same demand d_i. The second common characteristic is that a given customer i has the same coordinates (x_i, y_i) in two distinct instances of the same set (C1 or C2).

The third common characteristic appears in the scheduling horizon (column 3) of Table I, which is *short* for instances of set C1 (1236) and *large* for the instances of set C2 (3390). This means for example that the tours in instances C1 will all finish at most after 1236 units of time and at most after 3390 units of time for the instances of set C2. Finally, the fourth common characteristic concerns the vehicle capacity (column 5) of table I. In set C1, vehicles of capacity $C_{\max} = 200$ are used while this capacity is equal to 700 for instances of set C2.

TABLE I
Characteristics of the C1 and C2 instances

Set	Number of customers	Scheduling horizon	Service time	Vehicle Capacity
C1	100	1236	90	200
C2	100	3390	90	700

From these characteristics Solomon designed several instances in the same set by changing the time windows from an instance to another one. More precisely, there are nine instances C101–C109 in the first set C1 and eight instances C201–C208 in the second set C2. For two distinct instances in the same set (C1 or C2) we have:

- the same coordinates for a given customer i as well as for the depot D
- the same demand for a given customer i
- the same vehicle capacity
- different time windows

For more details, the reader can refer to Solomon's web site [20].

Then, one can surmise that less vehicles (clusters) will be needed for instances of set C2 comparing to set C1 because of the larger value of the capacity and the greater length of the time windows. This is in fact the case as proven in different works published in the literature.

Finally, note that the value of m (number of clusters or vehicles) is fixed to the best value found by several authors in the literature. More precisely, $m = 10$ for instances C1 and $m = 3$ for set C2.

Table II shows the results obtained on the 17 instances where column 1 indicates the name of each instance. Columns 2–4 contain the best known results in the literature (to our knowledge). Column 2 shows the best value for m (the number of vehicles) and column 3 the best distance. Column 4 (Ref.) indicates the reference to the paper where the best values for m and Dist were obtained. Columns 5–7 contain the results obtained by a method based on goal programming and genetic algorithm (GP-GA) proposed by Ghoseiri and Ghannadpour [12]. So columns 5 and 6 indicate the best value for m and the best distance respectively. Column 7 corresponds to the gap between the distance obtained by GP-GA (column 6) and the best known distance in the literature (column 3). This gap is computed as follows: $\text{gap} = 100\% \times (\text{Dist}_{\text{bestknown}} - \text{Dist}_{\text{GP-GA}})/\text{Dist}_{\text{bestknown}}$. Columns 8–12 summarize the results obtained by the proposed algorithm 3PA. Column 8 indicates the minimum number of vehicles m while column 9 shows the best minimum distance obtained by algorithm 3PA. Column 10 ($\omega_{\max}$) corresponds to the maximum value of the beam width used for each instance, then $\omega_{\max} = 50$ means that beam search was executed for each value $1 \leq \omega \leq 50$. The next column (time) indicates the total computation time needed for the execution of algorithm 3PA (in seconds). The last column (gap) indicates the difference (in %) between the solution obtained by the proposed method 3PA and the best known solution in the literature (column 3). More precisely $\text{gap} = 100\% \times (\text{Dist}_{\text{Best known}} - \text{Dist}_{\text{3PA}})/\text{Dist}_{\text{best known}}$. Finally, the last row of table II indicates the average gap for the two compared methods (GP-GA and 3PA). As expected, the number of vehicles needed for C1 instances (10) is larger than that needed for instances of set C2 (only 3 vehicles).

TABLE II
RESULTS OBTAINED ON INSTANCES C1 AND C2

	Best known			**GP-GA [12]**			**The proposed method (3PA)**				
Inst.	m	**Dist.**	**Ref.**	m	**Dist.**	**gap(%)**	m	**Dist.**	$\omega_{\max}$	**time (s)**	**gap (%)**
C101	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	29	0.00
C102	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	66	0.00
C103	10	**828.06**	[12]	10	**828.06**	0.0	10	828.94	50	140	-0.11
C104	10	**824.78**	[12]	10	**824.78**	0.0	10	828.94	50	183	-0.50
C105	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	36	0.00
C106	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	41	0.00
C107	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	41	0.00
C108	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	1600	146 600	0.00
C109	10	**828.94**	[12]	10	**828.94**	0.0	10	**828.94**	50	80	0.00
C201	3	**591.56**	[12]	3	**591.56**	0.0	3	**591.56**	50	176	0.00
C202	3	**591.56**	[12]	3	**591.56**	0.0	3	**591.56**	50	240	0.00
C203	3	**591.17**	[12]	3	**591.17**	0.0	3	**591.17**	100	13 210	0.00
C204	3	**590.60**	[17]	3	599.96	-1.58	3	591.17	100	12 390	-0.10
C205	3	**588.16**	[21]	3	588.88	-0.12	3	588.49	50	528	-0.06
C206	3	**588.49**	[17]	3	588.88	-0.07	3	**588.49**	2000	286 700	0.00
C207	3	**588.29**	[18]	3	591.56	-0.56	3	588.32	2000	291 600	-0.01
C208	3	**588.32**	[18]	3	**588.32**	0.0	3	588.88	50	618	-0.09
Average						-0.14					**-0.05**

The results of table II indicate that the proposed method 3PA reached the best known results in 11 cases out of 17. In the six other cases, the result is very close to the best known value since the gap is often smaller of equal to -0.11%, except for instance C104 where the gap reaches -0.50%. Note that even if the GP-GA method reaches the best known results in 13 cases out of 17, its average gap (-0.14%) is worst then that obtained by algorithm 3PA (-0.05%).

Concerning the computation time of algorithm 3PA, it is at most 183 seconds for instances of set C1 (except for C108 which was hard to solve). Each cluster in instances of set C1 contains about 10 customers. The computation time is generally greater for the second set C2, this is due to larger number of customers in each cluster (which is about 30) and then the number of combinations in each cluster (route) becomes larger.

But how to determine the maximum value for the beam width $\omega_{\max}$, especially for new instances. One can for example fix the maximum value to 50 or 100 or use a limited computation time.

Table III indicates, for each of the 17 solutions of table II the best value ω^* that gave the best solution for each cluster $C_j,\ j = 1,..,m$. We can see for example that $w^* = 1$ for all $j = 1,..,10$ for instance C101, but the values of ω^* are heterogeneous for instances C103 and C104 for example, meaning that these two instances are harder to solve (due to the characteristics of the time windows).

Fig. 4 shows an example of solution obtained after the second step of the algorithm (beam search), i.e., the output of algorithm 2 on instance C206. The total distance obtained after this step is equal to 597.35. Fig. 5 indicates improvement of the solution of fig. 4 (instance C206) by the 2-opt procedure (algorithm 3). The new distance was decreased from 597.35 to 588.49, i.e., an improvement of 1.48%. The 2-opt phase has changed several arcs in clusters 1 and 2 while the third cluster

TABLE III
BEST VALUE OF THE BEAM WIDTH IN EACH CLUSTER FOR INSTANCES C1 AND C2

	Cluster									
Inst.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
C101	1	1	1	1	1	1	1	1	1	1
C102	7	7	1	1	1	6	7	6	8	1
C103	39	13	1	8	7	10	44	5	8	4
C104	24	5	1	5	7	8	14	5	8	7
C105	1	1	1	1	1	1	1	1	1	1
C106	8	1	1	1	1	1	1	1	1	2
C107	1	1	1	1	1	1	1	1	1	1
C108	1518	2	3	2	2	10	1	1	3	1
C109	1	1	1	2	1	1	1	1	1	2
C201	1	1	1	–	–	–	–	–	–	–
C202	43	19	26	–	–	–	–	–	–	–
C203	43	35	68	–	–	–	–	–	–	–
C204	43	22	68	–	–	–	–	–	–	–
C205	1	1	1	–	–	–	–	–	–	–
C206	6	1687	1	–	–	–	–	–	–	–
C207	43	1928	1	–	–	–	–	–	–	–
C208	1	1	1	–	–	–	–	–	–	–

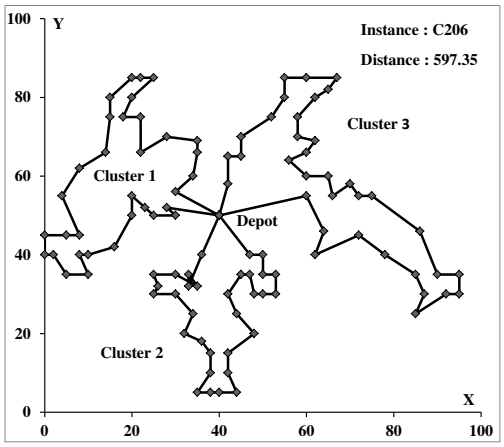


Fig. 4. Solution obtained by algorithm 3PA on instance C206 after the second step (Beam Search): $m = 3$, Distance=597.35

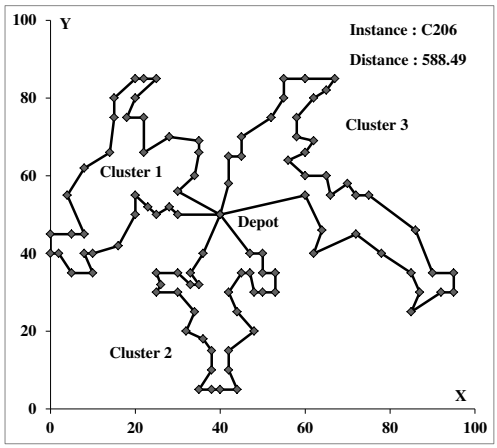


Fig. 5. Solution obtained by algorithm 3PA on instance C206 after the third step (Local Search): $m = 3$, Distance=588.49

has not changed.

V. Conclusion and future work

In this work a three-phase algorithm denoted by 3PA is proposed in order to solve the capacitated vehicle routing problem with time windows (CVRPTW). The central phase is that computing the shortest paths in a given cluster. To do so, a beam search is proposed. The most characteristic of beam search is that it explores several paths in parallel and increases then the probability to find *good* paths. The results obtained on the instances used show that the method is competitive since the computations revealed that algorithm 3PA was better than a method based on goal programming and genetic algorithm (GP-GA). Indeed, 3PA obtained a gap closer to the best known results in the literature than the gap obtained by GP-GA.

As a future work, it will be interesting to add a global evaluation criterion for beam search in order to provide solutions of better quality before calling the local search (third) phase and then to improve the overall solution.

References

[1] H. Akeb and M. Hifi. Algorithms for the circular two-dimensional open dimension problem. *International Transactions in Operational Research*, volume 15, pages 685–704, 2008.

[2] N. Azi, M. Gendreau, and J-Y. Potvin. An exact algorithm for a single vehicle routing problem with time windows and multiple routes. *European Journal of Operational Research*, volume 178, pages 755–766, 2007.

[3] N. Azi, M. Gendreau, and J-Y. Potvin. An exact algorithm for a vehicle routing problem with time windows and multiple use of vehicles. *European Journal of Operational Research*, volume 202, pages 756–763, 2010.

[4] R. Baldacci, E.A. Hadjiconstantinou, and A. Mingozzi. An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Operations Reserach*, volume 52, pages 723–738, 2004.

[5] R. Baldacci and V. Maniezzo. Exact methods based on node-routing formulations for undirected arc-routing problems. Networks, volume 47, pages 52–60, 2006.

[6] J.C.S. Brandão and A. Mercer. A tabu search algorithm for the multi-trip vehicle routing and scheduling problem. *European Journal of Operational Research*, volume 100, pages 180–191, 1997.

[7] A.M. Campbell and M. Savelsbergh. Efficient insertion heuristics for vehicle routing and scheduling problems. *Transportation Science*, volume 38, pages 369–378, 2004.

[8] I.M. Chao, B.L. Golden, and E.A. Wasil. A fast and effective heuristic for the orienteering problem. *European Journal of Operational Research*, volume 88, pages 101–111, 1996.

[9] S. Chen, B. Golden, and E. Wasil. The split delivery vehicle routing problem: Applications, algorithms, test problems, and computational results. *Networks*, volume 49, pages 661–673, 2007.

[10] J-F. Cordeau, G. Laporte, and A. Mercier. A unified tabu search heuristic for vehicle routing problems with time windows. *Journal of the Operational Research Society*, volume 52, pages 928–936, 2001.

[11] D. Feillet, P. Dejax, M. Gendreau, and C. Gueguen. An exact algorithm for the elementary shortest path problem with resource constraints: Application to some vehicle routing problems. *Networks*, volume 44, pages 216–229, 2004.

[12] K. Ghoseiri and S.F. Ghannadpour. Multi-objective vehicle routing problem with time windows using goal programming and genetic algorithm. *Applied Soft Computing*, volume 10(4), pages 1096–1107, 2010.

[13] X. Li and P. Tian. An ant colony system for the open vehicle routing problem. *Lecture Notes in Computer Science*, 4150, M. Dorigo et al. editions, Springer, pages 356–363, 2006.

[14] A. Lim and X. Zhang. A two-stage heuristic with ejection pools and generalized ejection chains for the vehicle routing problem with time windows. *INFORMS Journal on Computing*, volume 19(3), pages 443–457, 2007.

[15] P.S. Ow and T.E. Morton. Filtered beam search in scheduling. *International Journal of Production Research*, volume 26(1), pages 35–62, 1988.

[16] D. Pisinger and S. Ropke. A general heuristic for vehicle routing problems. *Computers & Computers Research*, volume 34, pages 2403–2435, 2007.

[17] J.Y. Potvin and S. Bengio. The vehicle routing problem with time windows. Part II. Genetic search. *INFORMS Journal of Computing*, volume 8, pages 165–172, 1996.

[18] Y. Rochat and E.D. Taillard. Probabilistic diversification and intensification in local search for vehicle routing. *Journal of Heuristics*, volume 1, pages 147–167, 1995.

[19] M.M. Solomon. Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research*, volume 35(2), pages 254–265, 1987.

[20] M.M. Solomon. VRPTW benchmark problems. http://w.cba.neu.edu/~msolomon.

[21] K.C. Tan, Y.H. Chew, and L.H. Lee. A hybrid multiobjective evolutionary algorithm for solving vehicle routing problem with time windows. *Computational Optimization and Applications*, volume 34, pages 115–151, 2006.

[22] K.C. Tan, L.H. Lee, Q.L. Zhu, and K. Ou. Heuristic methods for vehicle routing problem with time windows. *Artificial Intelligence in Engineering*, volume 15, pages 281–295, 2001.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 337–339

Real life cable constraints in designing Passive Optical Network architecture

Stanislas Francfort, Cédric hervet and Matthieu Chardy
Orange Labs
38-40 rue du Général Leclerc
92170 Issy mes Moulineaux
{*stanislas.francfort, cedric.hervet, matthieu.chardy*}*@orange.com*

Frédéric Moulis
SARL Moulis
204 rue Clémenceau
59139 Wattignies
frederic@moulis.eu

***Abstract*—Fiber To The Home (FTTH) deployment is crucial for telecommunication operators for both economical and quality of service reasons. This paper deals with a real-life Passive Optical Network (PON) design problem focusing on optical cabling constraints. This decision problem is formulated as an integer linear program (ILP) and several solving approaches are designed. Tests performed on real instances assess the efficiency of the proposed solution algorithms.**

***Index Terms*—Fiber optics, cables, integer programming.**

I. Introduction

INNOVATIVE bandwidth-requiring services lead telecommunication operators to the renewal of their fixed copper access networks, with the introduction of optical fibers. Among the optical fiber-based architectures, most telecommunication companies favor the PON architecture, appearing the best long-term technological solution [2]. In this paper, we thus focus on the design of FTTH-PON access network as experienced by field deployment teams from France Telecom-Orange. This decision problem arises as a joint optimization problem of optical splitters location, and cable routing and dimensioning.

Access network design problems have been intensively studied in the past decades. For a relevant survey, the reader can refer to [1]. To our knowledge, most papers related to PON design focus on fiber-oriented models i.e. skipping thus the cabling issues for later considerations (see. [3], [4]). Despite these problems being highly combinatorial, strengthened formulations of fiber-oriented PON design problems prove to be fairly tractable in practice (see. [4]). With regards to cabling issues, closest related work remains from Kim et al. [5] who propose ILP formulations for a global PON with cables design problem in a tree graph and a two-dimensional heuristic. The real-life PON deployment problem which is here dealt with, has at least two major differences: first no assumption is made on the underlying existing infrastructure except being with sufficient capacity (i.e. uncapacitated), and second, we exclude the possibility of gathering fibers with different types of end points (splitters or demand points) into the same cable, as well as fibers with different direction. Notice that this last possibility may appear as our underlying graph is not a tree. As far as the numerical experiment is concerned, the size of the instances and the set of available cables are as well very different in Kim et al. work compared to the present article. Those differences seem to us major requirements of an operational cabling policy. Formulated as an ILP, this problem proves intractable due to the specific cabling constraints. Therefore, we aim at proposing branch and bound-based algorithm based on its "fiber-oriented" relaxation, taking benefit from the practical tractability of the latter.

The remaining of this paper is organized as follows. Section II is dedicated to the modeling of the problem, proposing ILP formulations for both the real-life problem and its "fiber-oriented" relaxation. Section III reports for numerical tests performed on 4 real instances, before concluding in Section IV.

II. Problem modeling

In this section, we introduce a model for the PON deployment with cable constraints problem. Then, we show our decomposition approach to solve it more efficiently.

A. Main model and Real life constraints

Let $G = (V, E)$ be an undirected graph representing the infrastructure where the PON shall be deployed. Give orientation to each edge of G to obatin the directed graph $\overrightarrow{G} = (V, A)$ where for each edge $ij \in E$ define two reverse arcs ij and $ji \in A$. This architecture is modelled as in [4] under the form of an integer multiflow. Every edge has a length D_{ij} and every node has a demand a_i. We define 2 kinds of demand : a_i^h for "high" ones and a_i^l for "low" ones. High (respectively low) demands are defined as such if they exceed (respectively do not exceed) a given threshold t. For high demands, splitters of capacity m must be put directly on the demand node. Then, we have to transform the initial client demand into the last level splitter demand for the same node. For a given node $i \in V$, if $a_i < t$ (resp. $a_i \geq t$), then $a_i^h = \lceil \frac{a_i}{m} \rceil$ and $a_i^l = 0$ (resp. $a_i^h = 0$ and $a_i^l = a_i$. Low

978-83-60810-53-8/$25.00

demands are to be filled with fibers coming from an optical splitter. Splitters have a given capacity m, cost C_s and can be put on every nodes and are denoted by s_i^l. Fibers are coming out from a single node called the Optical Line Termination (OLT), of index 0. For the low demands architecture, there are 2 fiber levels. The level 1 fibers denoted by $f_{ij}^{l_1}$ (for all arcs $(i,j) \in A$), are used by splitters to produce level 2 fibers, denoted by $f_{ij}^{l_2}$ (for all arcs $(i,j) \in A$). For high demand nodes, technological concerns impose to put splitters on the demand node. It implies that there are only level 1 fibers to route for high demand nodes. We denote these fibers by f_{ij}^h (for all $(i,j) \in A$). Fibers are aggregated within cables of respective capacities $q \in Q$ (given in decreasing order, with $q_0 = \max_Q q$). We denote by $c_{ij}^{k,q}$ the number of cables of level $k \in \{1,2\}$ of capacity $q \in Q$ routed along the arc $(i,j) \in A$. Cables of capacity q cost C_c^q. Finally, we introduce the boolean variable b_{ij}^k which controls whether the arc $(i,j) \in A$ is used or not.

$$z_{cable} = \min_{c_{ij}^{k,q}, s_i} \sum_{ij \in E} (D_{ij} \cdot \sum_{\substack{k \in \{1;2\} \\ q}} C_c^q \cdot c_{ij}^{k,q}) + \sum_{i \in V} C_s \cdot s_i \quad (1)$$

$$c_{ij}^{k,q} \in \mathbb{N}, f_{ij}^{l_1} \in \mathbb{N}, f_{ij}^{l_2} \in \mathbb{N}, f_{ij}^h \in \mathbb{N}, s_i \in \mathbb{N}, b_{ij}^k \in \mathbb{N} \quad (2)$$

$$\forall i \in V \setminus \{0\} : s_i = \sum_{j \neq i} f_{ji}^{l_1} - \sum_{j \neq i} f_{ij}^{l_1} \quad (3)$$

$$\forall i \in V : a_i^l \leq \sum_{j \neq i} f_{ji}^{l_2} - \sum_{j \neq i} f_{ij}^{l_2} + m s_i \quad (4)$$

$$\forall i \in V : a_i^h = \sum_{j \neq i} f_{ji}^h - \sum_{j \neq i} f_{ij}^h \quad (5)$$

$$\forall k \in \{1;2\}, \forall i \in V : \sum_{j \neq i} b_{ji}^k \leq 1 \quad (6)$$

$$\forall k \in \{1;2\}, \forall ij \in A : b_{ij}^k \geq \frac{\sum_q c_{ij}^{k,q}}{N} \quad (7)$$

$$\forall ij \in A : \sum_{q \in Q} q \cdot c_{ij}^{1,q} \geq f_{ij}^{l_1} + f_{ij}^h \quad (8)$$

$$\forall ij \in A : \sum_{q \in Q} q \cdot c_{ij}^{2,q} \geq f_{ij}^{l_2} \quad (9)$$

$$\forall k \in \{1;2\}, \forall ij \in A : \sum_{q \in Q \setminus Q_0} c_{ij}^{k,q} \leq 1 \quad (10)$$

The objective function is denoted by z_{cable}. In the model presented above, constraints (3)-(5) ensure flow conservation for all level of fibers, according to the number of splitters and demand. Constraints (6) and (7) ensure that only one edge incident to a node will be used, in order for the deployed network to have tree properties. Constraints (8) and (9) allow aggregation of fibers within cables. Constraints (10) ensure that only one cable is routed through each edge (except for the biggest cables q_0).

B. Decomposing and using a warm start

The $\mathcal{P}_c$ model proves intractable on real-size instances (refer to section III), but it can be decomposed so that we obtain a model easier to solve. A fiber-based model, denoted by $\mathcal{P}_f$, can be derived from $\mathcal{P}_c$ as follows:

1) discard variable $c_{ij}^{k,q}$ describing the cables,
2) discard inequalities (8) to (10),
3) optimize along the objective function z_{fiber} instead of z_{cable}, that is replace equation (1) by equation (11) with C_f^k the cost[1] of fiber f^k for all $k \in \{h; l_1; l_2\}$.

$$z_{fiber} = \min_{f_{ij}^k, s_i} \sum_{ij \in E} (D_{ij} \cdot \sum_{k \in \{h;l_1;l_2\}} C_f^k \cdot f_{ij}^k) + \sum_{i \in V} C_s \cdot s_i \quad (11)$$

Solving $\mathcal{P}_f$ allows us to get values for fibers and splitters variables, so a solution of $\mathcal{P}_f$ is almost a feasible solution of $\mathcal{P}_c$ in a sense that only cable variables remain to be set. Therefore we design two solution algorithms based on feasible solution of $\mathcal{P}_f$.

- $\mathcal{P}_f + \mathcal{H}_c$: given any feasible solution of $\mathcal{P}_f$, we set cable constaints by use as much maximum capacity cables as necessary and cover the remaining fibers with the smallest cable whose capacity is greater than the number. $\mathcal{H}_c$ is delailed in Algorithm 1.
- $\mathcal{P}_f + \mathcal{P}_c$: given any feasible solution of $\mathcal{P}_f$, we set arbitrarily large values to c^k variables to satisfy the cable constraints (8) to (10) and be able to set a warm start to $\mathcal{P}_c$.

Algorithm 1 Heuristic H_c

Initialize $C_{ij}^{k,q} \leftarrow 0 \ \forall ij \in A; \forall k \in 1;2; \forall q \in Q$
for all arc $ij \in A$ **do**
  Set $f := f_{ij}^k$
  while $f \geq 288$ **do**
    $C_{ij}^{k,288} := C_{ij}^{k,288} + 1$
    $f := f - 288$
  end while
  choose $\min q$ s.t. $q \geq f$
  Set $C_{ij}^{k,q} := 1$
  Set $C_{ij}^{k,288} \leftarrow \lfloor f_{ij}^k 288 \rfloor$
  Set $C_{ij}^{k,\tilde{q}} \leftarrow 1$ where $\tilde{q} \leftarrow \min q$ s.t. $q \geq f_{ij}^k - \lfloor f_{ij}^k / 288 \rfloor$
end for

[1]Note that this cost is "virtual" in a sense as it should depend on the size of the cable which will be used for each level of fiber. In practice we will assume that capacity of cables are decreasing with the level of fiber and, given the concave cost structure of the cables, we will have decreasing fiber cost (with respect to their level)

III. Numerical results

Our goal is to be efficient on our very specific operational data. That is the reason why we have not conducted any experiments on any public Network or MIP library. In this context, we present numerical results from experiments on 4 data sets, named $Data1$ the smallest one, to $Data4$ the biggest one. The underlying infrastructure G and the constants (such as cables costs C_c^q or demands a_i^h and a_i^l) have been set to their actual values; splitters capacity set to $m = 8$, and the cables capacities are chosen in $Q = \{288 = q_0; 144; 96; 72; 48; 36; 24; 12\}$, while N is set to 1000. The Linear Programming and the branch and bound were performed by CPLEX 12.2 running on an AMD Athlon II X3 powered by a Linux 2.6 kernel. Computation times were set to 1800 seconds for $\mathcal{P}_f$ and $\mathcal{P}_c$, meanwhile $\mathcal{H}_c$ runs quite instantly. Table (I) and (II) summaries some numerical experiments.

TABLE I
Experimentation $\mathcal{P}_f$, 1800 seconds

Data	$\lvert V\rvert$	$\lvert E\rvert$	Demand	Cost $\mathcal{P}_f + \mathcal{H}_c$
$Data1$	583	838	8285	45606
$Data2$	808	2528	46294	110269
$Data3$	1232	3119	28080	105062
$Data4$	1624	2711	23774	110554

TABLE II
Experimentation $\mathcal{P}_c$, 1800 seconds

Data	Cost $\mathcal{P}_c$	Gap	Cost $\mathcal{P}_f + \mathcal{P}_c$	Gap	Δ
$Data1$	34624	23.95	34348	22.42	-24%
$Data2$	N/A	N/A	60202	27.94	-45%
$Data3$	N/A	N/A	71668	34.06	-32%
$Data4$	N/A	N/A	106372	52.46	-3%

Let us explain some results shown in table (I). $\mathcal{P}_c$ finds a feasible solution only on $Data1$, the smallest. Hence, by solving $\mathcal{P}_f$ first, we help to find a feasible solution even on biggest data sets. And then by solving $\mathcal{P}_c$ we lower the cost a lot more than by solving $\mathcal{H}_c$ for the three smallest data sets as shown in column Δ where Δ quantifies the difference between the cost of $\mathcal{P}_f + \mathcal{H}_c$ and $\mathcal{P}_f + \mathcal{P}_c$.

Moreover, since for $Data1$ the cost of $\mathcal{P}_c$ and $\mathcal{P}_f + \mathcal{P}_c$ are about the same, we think that the warm start helps to find a solution as good as the one that would have been found without it on other data sets.

Let us discuss about $Data4$. For this data set the gap is almost twice the gap on smaller data sets, and the value in column Δ is rather small. We think the number of vertices $|V|$ is too big for CPLEX to solve efficiently the branch and bound process. This thought is also supported by the fact that for model $\mathcal{P}_c$, if the computing time increases up to 10800 seconds, this does not achieve any improvement more than 1% gap. We expect to find a more efficient decomposition technique in order to lower the gap on big data sets. We should remark as well that for $\mathcal{P}_f$, gaps vary from 1% to 4%, which is very small.

IV. Conclusion

In the present article we have shown that by using a decomposition and a warm start, we make possible to solve a MIP model describing a PON access network design problem with cable constraints, even on very big data sets. We have shown that for some data sets, the entire model performs a good optimization once a warm start is given. The solutions found are admissible in an operational point of view, that means conform and checkable, fast to compute and easy deploy; moreover far cheaper than the best solution found by another mean.Notice that cable constraints (8) to (10) make the problem hard to solve on our large size of data sets. Notice that although it is easy to find a deployable, feasible solution without any computer assistance the price found by our model is around 50% of the best cost found by hand.

We think that the warm start could be improved by finding a good heuristic or by studying the link between objective function (1) and (11). To test wether this could help CPLEX to find a better solution on the entire problem is an open question. Some other constraints are post-processed after the MIP has been solved; we wish to encompass them inside the main model. Finally, since the gap for some data sets is quite large and that running CPLEX a long time doesn't lower it, we wish as well to better understand the decomposition used for the warm start in order to improve this step and achieve a better gap.

References

[1] B. Gavish, 1991. Design of telecommunication networks - Local access design methods. Annals of Operations Research 33, 17-71.
[2] D. Gutierrez, K.S. Kim, S. Rotolo, F. An, L.G. Kazovsky. 2005. FTTH standards, deployments and research issues. In: Proceeding of 8th Joint Conference on Information Sciences 2005, Salt Lake City, UT, USA
[3] H.D. Sherali, Y. Lee, T. Park. New modeling approaches for the design of local access transport area networks, European Journal of Operational Research, vol 127, 94-108, 2000
[4] Chardy, M., Costa, M-C., Faye, A. and Trampont, M. 2011. Optimizing the deployment of a multilevel optical FTTH network. Europeen Journal of Operations Research (under revision).
[5] Youngjin, K., Youngho, L. and Junghee, H. 2010. A splitter location-allocation problem in designing fiber optic access networks. European Journal of Operational Research Volume 210, Issue 2, 16 April 2011, 425-435.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 341–346

Energy-based Pruning Devices for the BP Algorithm applied to Distance Geometry

Douglas Gonçalves*, Antonio Mucherino*, Carlile Lavor†

*IRISA, University of Rennes 1, Rennes, France.
{douglas.goncalves,antonio.mucherino}@irisa.fr

†IMECC-UNICAMP, Campinas-SP, Brazil.
clavor@ime.unicamp.br

***Abstract*—The Molecular Distance Geometry Problem (MDGP) is the one of finding an embedding of a molecular graph in the three dimensional space, where graph vertices represent atoms and edges represent known distances between some pairs of atoms. The MDGP is a constraint satisfaction problem and it is generally cast as a continuous global optimization problem. Moreover, under some assumptions, this optimization problem can be discretized and so that it becomes combinatorial, and it can be solved by a Branch & Prune (BP) algorithm. The solution set found by BP, however, can be very large for some instances, while only the most energetically stable conformations are of interest. In this work, we propose and integrate the BP algorithm with two new energy-based pruning devices. Computational experiments show that the newly added pruning devices are able to improve the performance of the BP algorithm, as well as the quality (in terms of energy) of the conformations in the solution set.**

I. Introduction

THE Molecular Distance Geometry Problem (MDGP) is the one of finding the possible three-dimensional conformations of a molecule from the information about the relative distances between some pairs of its atoms [1], [2]. Let $G = (V, E, d)$ be a weighted undirected graph representing an instance of the MDGP. The vertex set describes the atoms forming the molecule, and there is an edge joining two vertices if and only if the distance between the two corresponding atoms is known. The MDGP can be seen therefore as the problem of finding a function $x : V \longrightarrow \Re^3$ such that

$$\forall (u, v) \in E, \qquad ||x(u) - x(v)|| = d_{uv},$$

where $||\cdot||$ represents the computed Euclidean distance between the coordinates $x(u)$ and $x(v)$, whereas d_{uv} is the weight of the edge (u, v). The MDGP is NP-hard [3].

By its nature, the MDGP is a constraint satisfaction problem that is generally formulated as a global optimization problem in a continuous space [4]. When some particular assumptions are satisfied, moreover, the search space of the optimization problem can be discretized, so that it becomes combinatorial. We refer to a subclass of MDGPs that can be discretized as the Discretizable MDGP (DMDGP) [5], [6], [7]. Instances of the DMDGP can be solved by applying an ad-hoc Branch & Prune (BP) algorithm [8].

The basic idea behind BP is as follows. Suppose that possible positions have already been computed for all the atoms of a given molecule that have rank smaller than i (we suppose that a total order relationship for the vertices of G exists). Because of the discretization assumptions, there are up to two possible positions for the current atom i, that can be obtained by intersecting three spheres centered in the already placed atoms $i-3$, $i-2$ and $i-1$, and having radii $d_{i-3,i}$, $d_{i-2,i}$ and $d_{i-1,i}$ respectively. In this way, a binary tree can be defined, which is the conformational search space of the discretized problem, where branches duplicate in number when passing on higher level layers. By using some additional information about the distances (that are not considered in the tree construction), the feasibility of the atomic positions can be verified, so that branches of the tree can be pruned in case they contain infeasible positions. Moreover, additional pruning devices can be conceived for improving the performance of the BP algorithm [9], [10].

In this paper, we propose two new pruning devices to be included in the BP algorithm in order to improve its performance. For the first time, we consider pruning devices that are based on the chemical nature of MDGP instances; in other words, in these new pruning devices, we exploit the chemical structure of the molecule, and not only the distance information. We will present two new pruning devices. The former basically considers the van der Waals (vdW) radii [11] and forbids any configuration where non-bonded atoms are too close to each other, by verifying the relative distances between spheres centered in the atoms and having as radii the corresponding vdW radii. The latter pruning device is instead based on the well-known Lennard Jones (LJ) potential [12], which is related to the internal energy of the molecule.

The rest of the paper is organized as follows. In Section II, we will briefly describe the BP algorithm and we will focus our attention on the symmetry properties of BP trees, that will be exploited later in the paper. Section III will introduce the two new pruning devices, while computational experiments will be presented in Section IV. Conclusions and future works will be discussed in Section V.

978-83-60810-53-8/$25.00

Algorithm 1 The BP algorithm.

```
1: BP(v, n, d)
2: compute x'_v;
3: if (x'_v is feasible) then
4:    if (v = n) then
5:       let nsols = nsols + 1;
6:    else
7:       BP(v + 1, n, d);
8:    end if
9: end if
10: compute x''_v;
11: if (x''_v is feasible) then
12:    if (v = n) then
13:       let nsols = nsols + 1;
14:    else
15:       BP(v + 1, n, d);
16:    end if
17: end if
```

II. The BP algorithm

The BP algorithm [8] is an exact algorithm for the solution of DMDGPs. It explores recursively the discrete search domain (a binary tree) of the DMDGP and it prunes infeasible branches of such a tree as soon as they are discovered. Alg. 1 gives a sketch of the algorithm. In the algorithm call, $v \in V$ is the current vertex for which we are looking for a position, n is the cardinality of V, and d represents the weights associated to the edges. During each call, the two possible positions for the vertex v, x'_v and x''_v, are computed, and their feasibility is verified. If, for example, x'_v is feasible, then this position could be part of a solution, and therefore the branch of the binary tree rooted at x'_v needs to be explored. In this case, the algorithm invokes itself for exploring the branch rooted at x'_v. Instead, if for example x''_v is not feasible, then the current branch does not contain any solution. It is therefore pruned: the algorithm does not invoke itself in this case.

The conditions in the two **if** control structures (see lines 3 and 11 of Alg. 1) can be verified by employing the so-called *pruning devices*. The easiest to conceive and to implement (and probably the most efficient) is the Direct Distance Feasibility (DDF) pruning device. DDF simply verifies whether additional distances (that are not employed in the computation of x'_v and x''_v) are satisfied by the obtained candidate positions. In the following, we will refer to such additional distances as *pruning distances*. The BP algorithm, together with the DDF pruning device, was shown to be very efficient for the solution of protein-like instances [5].

Fig. 1 shows a BP tree for a small instance and the set of solutions obtained when only the DDF pruning device is employed. In the following, we will say that this is the DDF solution set, in order to make a distinction between this solution set and the ones that we will obtain while employing the new pruning devices. In the tree representation, a solution is given by a path from the tree root to one of its leaf nodes.

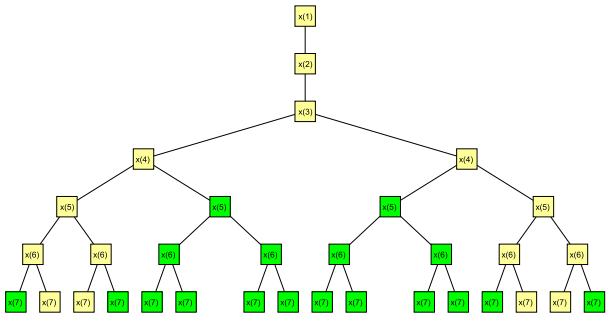


Fig. 1. The BP tree for a 7-atom instance, and the DDF solution set (in yellow).

As it is graphically shown in the picture, DDF solution sets are symmetric, i.e. one solution can be obtained from another by replacing a branch with its symmetric one [13], [14].

On the other hand, BP trees are highly symmetric: when no additional distances are available for applying the DDF pruning device, then every tree branch has a symmetric branch. A pair of symmetric branches shares the following property: given two vertices v and w, in the two branches, even if their coordinates are different, their relative distance is the same. Conversely, if we consider a vertex u having rank smaller than the common root of the two symmetric branches, its distance from any vertex v on the two branches is branch-dependent.

Let us consider a pair of symmetric branches of the BP tree that are rooted at the vertex v. If there is a vertex with rank $u < v$ and another vertex $w > v$ (which belongs to the two branches) such that the pruning distance d_{uw} is available, then, with probability 1, at most one of the two coordinates for w (one on the first branch, another on the second one) can satisfy this distance [15]. As a consequence, this kind of pruning distances are able to break the symmetries in BP trees, so that only a subset of symmetries is left in the DDF solution set. An easy test for verifying the presence of a symmetry in the solution set (before its explicit computation) is to check the existence of this kind of pruning distances. As reported in [13], there is a symmetry on the layer v of the solution set if and only if the vertex v belongs to the set

$$B = \{v \in V : \nexists (u, w) \text{ s.t. } u + 3 < v \leq w\}.$$

In this work, we will exploit these symmetry properties for generating instances having a predefined minimum number of solutions, and we will also discuss a possible improvement for the LJ pruning device which exploits the symmetries in BP trees.

III. Energy-based pruning devices

In this section, we propose two new pruning devices to be included in the BP algorithm. The first one is based on vdW radii (see Section III-A), while the second one is based on LJ energy (see Section III-B).

A. vdW radii pruning device

In this work, we consider the very common representation for an atom through the coordinates of its center. An atom, however, fills a certain portion of space: its nucleus, consisting of neutrons and protons, has a predetermined volume, while electrons orbit around this nucleus, at a given distance from it. Therefore, an atom can be seen as a sphere centered in its nucleus (which corresponds to the center of the atom) and having radius equal to the distance between the nucleus and the orbiting electrons. This distance can be estimated for each kind of atoms, and it is generally referred to as *atomic radius*. It is common to say that these electrons form a sort of *cloud*.

When two atoms are chemically bonded, their clouds of electrons tend to overlap. If they are not bonded, however, repulsion forces do not allow them to be too close to each other. The half of the distance (between atoms of the same kind) for which the attraction and repulsion forces are in equilibrium is called *van der Waals* (vdW) radius [11]. The vdW pruning device is therefore based on this simple idea: when two atoms are not bonded, their relative distance should be greater than the sum of the two corresponding vdW radii. This verification can be applied to all pairs of atoms for which no pruning distance is available.

Notice that, differently from the DDF pruning device, a precise distance is not available but rather only a lower bound for this distance. When the relative distance between two vertices u and v, with $u < v$ just positioned somewhere, goes below the predefined threshold, then the candidate position for v can be pruned. In the practice, we consider a relaxed condition, by setting the threshold to the 80% of the sum of the vdW radii. In the experiments presented in Section IV, we will consider instances containing Carbons (C) and Nitrogens (N) only. The vdW radius for C is set to 1.875; the vdW radius for N is set to 1.688. These values were extracted from the default parameters of the force field described in [16].

B. LJ pruning device

This pruning device is based on the overall internal energy of a molecule. As it is well-known, an accurate description of all interactions among the atoms in a molecule can be very complex, so that the overall energy can be only approximated by taking into consideration the most important interactions.

The vdW interactions between pairs of non-bonded atoms [17] are usually modeled by the Lennard Jones energy. In this case, we consider both repulsion and attraction forces, and for modeling the overall energy, we use the sum of pairwise LJ potentials 12–6 [12] :

$$E_{LJ} = \sum_{uv} 4\varepsilon_{uv} \left[\left(\frac{\sigma_{uv}}{d_{uv}} \right)^{12} - \left(\frac{\sigma_{uv}}{d_{uv}} \right)^{6} \right], \qquad (1)$$

where ε_{uv} and σ_{uv} are two parameters that can be defined by the relationships between the pairs of atoms u and v. The parameter σ_{uv} is the distance where the pair potential is zero, whereas ε_{uv} is the well depth. The minimum value for the LJ pair potential is $-\varepsilon_{uv}$ achieved in $r_{uv} = 2^{1/6}\,\sigma_{uv}$ (which corresponds to the sum of vdW radii).

During the execution of the algorithm, every time a leaf node is reached (on the layer n), a complete conformation is found, and its energy E_n can be computed. Let us suppose that $\hat{E}_n$ is the lowest energy found so far. The basic idea behind the LJ pruning device is to verify in advance whether new branches of the tree can actually contain conformations with an energy that can be potentially smaller than $\hat{E}_n$. This can be done by computing a lower bound on the energy concerning all the conformations belonging to a common branch.

Depending on the range in which the inter-atomic distances can vary, however, we can compute an accurate lower bound for the actual value. In case the BP algorithm is currently positioned on the layer v, then we have a partial energy value $E_{n(\leq v)}$ (computed by using the available coordinates) and a lower bound $L_{(>v)}$ on the energy $E_{n(>v)}$ (approximated by summing the minimum values given by the Lennard Jones terms for which no distance is available yet). Therefore, if $E_{n(\leq v)} + L_{(>v)} > \hat{E}_n$, there is no hope to identify a conformation with an energy smaller than $\hat{E}_n$ while exploring the current branch of the tree. This branch can be therefore pruned.

Notice that the LJ pruning device considers implicitly the vdW pruning device (see previous section). If a distance between a pair of atomic coordinates (for the atom v and a previous one) is small enough for the atomic position x_v to be pruned by the vdW pruning device, then the corresponding LJ potential is large, so that the LJ pruning device declares the atomic position infeasible as well. Therefore, when the vdW and LJ pruning devices work together, it is appropriate to apply LJ only after vdW.

IV. Computational experiments

This section presents some computational experiments where the BP algorithm is integrated with the new proposed pruning devices. All codes were written in C programming language and all the experiments were carried out on an Intel Core 2 Duo @ 2.4 GHz with 2GB RAM, running Mac OS X. The codes have been compiled by the GNU C compiler v.4.0.1 with the `-O3` flag.

In this paper, we suppose that all considered instances consist of a list of precise distances between some pairs of atoms of the molecule. This is an unrealistic assumption [18], because this information can be obtained through experiments of Nuclear Magnetic Resonance (NMR), where lower and upper bounds on the distances are actually provided. This assumption, however, allowed us to begin the investigation of new ideas that are potentially able to help in the solution of real instances of the problem (see Section V).

The instances considered in this paper are artificially generated by using the following procedure. Protein conformations are downloaded from the Protein Data Bank (PDB) [19] and the backbone atoms N–C_α–C of such proteins are extracted from such conformations. Distances are then computed between each possible pair of atoms, and the distances that are greater than 5Å are rejected (this is done because NMR data only consists of short range distances). As previously remarked

TABLE I
SOME COMPUTATIONAL EXPERIMENTS WITH BP AND THE NEW ENERGY-BASED PRUNING DEVICES.

Instance			only DDF			DDF + vdW			DDF + LJ			DDF + vdW + LJ		
name	n	m	BP calls	#L	time	BP calls	#L	time	BP calls	#L	time	BP calls	#L	time
1mbn-0	459	3200	1951	8	**0.03**	1951	8	0.05	1951	3	**0.03**	1951	3	0.05
1mbn-2	459	3169	11327	512	0.35	4071	96	0.16	9229	5	0.27	4071	4	**0.15**
1rgs-0	792	4936	10883	8	0.43	9059	8	0.51	9127	5	**0.35**	9059	5	0.51
1rgs-1	792	4857	165317	128	6.82	108033	96	6.27	123025	13	**4.90**	108033	11	6.29
1bpm-1	1443	9056	73147	128	6.88	14775	20	**1.89**	36599	7	3.26	14775	4	**1.89**
1bpm-2	1443	9027	1150409	2048	111	90143	108	**11.9**	272474	9	24.9	90143	6	**11.9**
1n4w-1	1610	10860	51521	32	3.39	21785	6	**1.52**	32037	7	1.79	21785	3	1.53
1n4w-2	1610	10675	182433	512	18.75	26855	24	**2.42**	52962	17	4.24	26855	4	2.43
1mqq-1	2032	12820	54175	128	8.62	17347	32	4.64	22067	4	**3.22**	17347	4	4.65
1mqq-2	2032	12807	812927	2048	144	161265	344	49.80	256059	6	**42.25**	161625	4	49.84
1rwh-1	2265	13908	38261	32	5.39	9727	2	**1.43**	17101	4	2.06	9227	2	**1.43**
1rwh-2	2265	13868	1152495	1024	169	35649	8	**5.43**	77613	9	9.32	35649	4	5.45
2e7z-1	2907	27509	342675	64	38.63	229483	8	30.91	256046	7	**25.44**	229483	3	30.89
2e7z-2	2907	27157	2041939	2048	391	271435	16	**42.61**	470990	21	68.11	271435	5	42.63
1epw-0	3861	35028	11975	2	3.12	11299	2	5.67	11489	2	**2.88**	11299	2	5.72
1epw-1	3861	34707	123561	32	43.27	22457	4	12.72	38580	4	**11.16**	22457	2	12.74
1epw-2	3861	34052	2815081	4096	1282	48297	32	**36.79**	192497	23	81.09	48297	3	36.81

in [5], the obtained set of distances forms a DMDGP instance in the majority of the cases (this is always the case for the proteins considered in this paper).

Successively, for all generated instances, a certain number of pruning distances is discarded. Let $\hat{V} \subset V$, containing K randomly selected vertices. For each $v \in \hat{V}$, all the pruning distances d_{uw} such that $u + 3 < v < w$ are removed from the instance. In this way, a symmetry in the DDF solution set is generated, and this makes the total number of solutions increase (see Section II). An instance generated by using this procedure has at least 2^K solutions [15], [13].

Table I shows some computational experiments. The name of each instance is composed by its label on the PDB, plus a number, representing the cardinality K of $\hat{V}$. For every instance, we also provide the number of atoms ($n = |V|$), and the total number of available distances ($m = |E|$).

The experiments are performed for different setups of the BP algorithm. We consider the four following setups:

- only the DDF pruning device is exploited;
- DDF is integrated with the vdW pruning device;
- DDF is integrated with the LJ pruning device;
- the three pruning devices are considered together.

For every setup and for every instance, we monitor the total number of BP calls necessary for enumerating the whole solution set, the number #L of times that BP reaches a leaf node of the search tree (#L corresponds to the cardinality of the solution set when only DDF is employed), and finally the CPU time, in seconds. CPU times in bold are used to mark the fastest executions.

The quality of the obtained solutions can be evaluated in two ways. In order to verify whether all available distances are satisfied, we use the Largest Distance Error (LDE) function:

$$LDE(x) = \sum_{(u,v)\in E} \frac{1}{m} \frac{|\,||x(u) - x(v)|| - d_{uv}\,|}{d_{uv}}.$$

In this work, moreover, it is of interest to verify the LJ energy of the obtained solutions, by employing Equation (1). In the presented experiments, the LJ energy is computed even when the LJ pruning device is disabled. The partial energy $E_{(>k)}$ is computed at each recursive call of BP, because we noticed empirically that this is more efficient than computing the LJ energy for the whole DDF solution set after the execution of the algorithm. For lack of space in the table, LDE and LJ values are omitted for all experiments, but some examples will be given in the text.

The experiments show the effectiveness of the two new pruning devices. Even if very simple, the vdW pruning device is able to reduce the total number of BP calls, as well as the computational time, because it can identify infeasible branches that were not selected by DDF. A similar observation can be done for the LJ pruning device. It is interesting to remark that, when the two new pruning devices are considered together with DDF, the performance of BP is similar to the case in which only the vdW pruning device is considered. As a consequence, the repulsive term in LJ, which dominates the potential for small distances (also considered in vdW), plays the most important role during the pruning process. However, notice that #L decreases when vdW and LJ work together: fewer leaf nodes are reached during the execution of BP.

Fig. 2 shows two solutions for the instance 1mbn-2. The leftmost conformation has minimum LDE value 1.79e-10 (energy -131.21), but it does not correspond to the solution having the smallest energy value. In fact, as it can be seen from the figure, one helix in this conformation slightly diverges from the rest of the molecule, so that some energetic terms increase in value. The rightmost conformation is the one with minimum LJ potential energy: -136.12, while the LDE value is 5.36e-08.

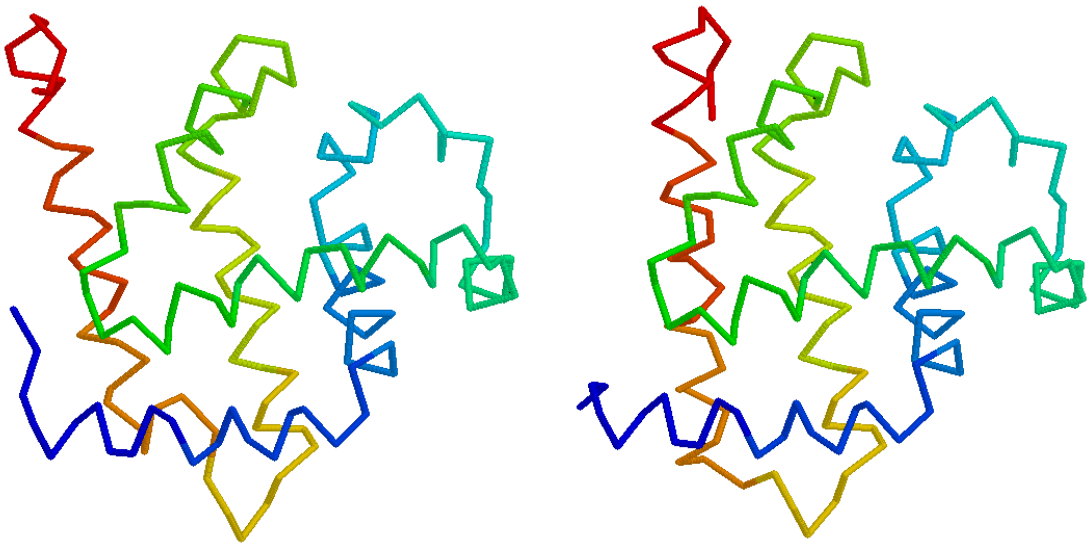


Fig. 2. Two solutions for the instance `1mbn-2`. The leftmost conformation has minimum LDE value, while the rightmost conformation has minimum LJ energy.

V. Conclusions and future works

We introduced two new pruning devices that are based on internal energy in molecules. We showed, through computational experiments, that they are able to improve the pruning capabilities of BP algorithm. The energetically most stable conformations, that previously could be selected from the entire DDF solution set, can be actually obtained by employing these two new pruning devices. The solution set is therefore reduced to energetically stable conformations only, while the performance of BP improves.

This work represents the first step for the integration of energy-based pruning devices inside BP. Next step will consist in testing such new pruning devices on instances of the problem containing interval data. As mentioned above, this kind of instances is more realistic, because, in biological applications, experimental data are generally imprecise. While the adaptation of these pruning devices to interval data is rather trivial, their impact on the performance of the algorithm is expected to be much more pronounced.

When working with interval data, the number of branches in BP trees increases because, instead of 2 possible positions for the current atom, we have $2D$ possible positions, where D corresponds to the number of samples in the discretized interval distance [20]. Thus, it is also expected an increase in the computational cost for the execution of these new pruning devices. For this reason, it will be worth implementing suitable strategies for making the new pruning devices more efficient. In the LJ pruning device, for example, the quality of the lower bound $L_{(>v)}$ can be improved by taking into consideration the symmetry properties of BP trees. Distances between pairs of atoms belonging to two symmetric branches coincide, and therefore their LJ energy is the same. Once one of the two branches has been explored, and its total energy has been computed, it is known in advance that the second branch has the same energy. This property can therefore help in finding better approximations of the lower bound $L_{(>v)}$. Unfortunately, when using this strategy in BP with exact distances, the trade-off between increased cost and performance improvement is not relevant.

VI. Acknowledgments

We wish to thank Thérèse Malliavin for the fruitful comments on this paper. We are also thankful to Brittany Region (France), for funding a 1-year postdoc for DG, and to UNICAMP, CNPq, FAPESP and INRIA, for partial financial support.

References

[1] G. Crippen and T. Havel, *Distance Geometry and Molecular Conformation*. New York: John Wiley & Sons, 1988.

[2] A. Mucherino, C. Lavor, L. Liberti, and N. M. (Eds.), *Distance Geometry: Theory, Methods and Applications*. 410 pages, Springer, 2013.

[3] J. B. Saxe, "Embeddability of weighted graphs in k-space is strongly np-hard," in *Proceedings of* 17^{th} *Allerton Conference in Communications, Control and Computing*, Monticello, IL, 1979, pp. 480–489.

[4] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *to appear in SIAM Review*, 2013.

[5] C. Lavor, L. Liberti, N. Maculan, and A. Mucherino, "The discretizable molecular distance geometry problem," *Computational Optimization and Applications*, vol. 52, pp. 115–146, 2012.

[6] L. Liberti, C. Lavor, A. Mucherino, and N. Maculan, "Molecular distance geometry methods: from continuous to discrete," *International Transactions in Operational Research*, vol. 18, pp. 33–51, 2010.

[7] C. Lavor, L. Liberti, N. Maculan, and A. Mucherino, "Recent advances on the discretizable molecular distance geometry problem," *European Journal of Operational Research*, vol. 219, pp. 698—706, 2012.

[8] L. Liberti, C. Lavor, and N. Maculan, "A branch-and-prune algorithm for the molecular distance geometry problem," *International Transactions in Operational Research*, vol. 15, pp. 1–17, 2008.

[9] C. Lavor, L. Liberti, A. Mucherino, and N. Maculan, "On a discretizable subclass of instances of the molecular distance geometry problem," in *ACM Conference Proceedings, 24th Annual ACM Symposium on Applied Computing*, Hawaii, USA, 2009, pp. 804–805.

[10] A. Mucherino, C. Lavor, T. Malliavin, L. Liberti, M. Nilges, and N. Maculan, "Influence of pruning devices on the solution of molecular distance geometry problems," in *Proceedings of the* 10^{th} *International Symposium on Experimental Algorithms (SEA11)*, ser. Lecture Notes in Computer Science, P. M. Pardalos and S. Rebennack, Eds., Crete, Greece, 2011, vol. 6630, pp. 206–217.

[11] A. Bondi, "van der waals volumes and radii," *Journal of Physical Chemistry*, vol. 68, no. 3, pp. 441–451, 1964.

[12] J. E. L. Jones, "Cohesion," in *Proceedings of the Physical Society*, vol. 43, 1931, pp. 461–482.

[13] A. Mucherino, C. Lavor, and L. Liberti, "A symmetry-driven bp algorithm for the discretizable molecular distance geometry problem," in *IEEE Conference Proceedings, Computational Structural Bioinformatics Workshop (CSBW11), International Conference on Bioinformatics & Biomedicine (BIBM11)*, Atlanta, GA, USA, 2011, pp. 390–395.

[14] ——, "Exploiting symmetry properties of the discretizable molecular distance geometry problem," *Journal of Bioinformatics and Computational Biology*, vol. 10, no. 1242009, pp. 1–15, 2012.

[15] L. Liberti, B. Masson, J. Lee, C. Lavor, and A. Mucherino, "On the number of solutions of the discretizable molecular distance geometry problem," in *Proceedings of the* 5^{th} *Annual International Conference on Combinatorial Optimization and Applications (COCOA11)*, ser. Lecture Notes in Computer Science, vol. 6831, 2011, pp. 322–342.

[16] J. P. Linge and M. Nilges, "Influence of non-bonded parameters on the quality of nmr structures: A new force field for nmr structure calculation," *Journal of Biomolecular NMR*, vol. 13, no. 1, pp. 51–59, 1999.

[17] H. C. Hamaker, "The london–van der waals attraction between spherical particles," *Physica*, vol. 4, no. 10, pp. 1058–1072, 1937.

[18] T. E. Malliavin, A. Mucherino, and M. Nilges, "Distance geometry in structural biology: New perspectives," in *Distance Geometry: Theory, Methods and Applications*, L. L. N. M. A. Mucherino, C. Lavor, Ed. Springer, 2013, pp. 329–350.

[19] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 2000.

[20] C. Lavor, L. Liberti, and A. Mucherino, "The interval branch-and-prune algorithm for the discretizable molecular distance geometry problem with inexact distances," *Journal of Global Optimization*, vol. 56(3), pp. 855–871, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 347–354

A Maximum Matching Based Heuristic Algorithm for Partial Latin Square Extension Problem

Kazuya Haraguchi*, Masaki Ishigaki and Akira Maruoka†
Department of Information Technology and Electronics
Faculty of Science and Engineering
Ishinomaki Senshu University
Ishinomaki, Miyagi 986-8580, Japan
Email: kzyhgc@gmail.com*, amaruoka@isenshu-u.ac.jp†

***Abstract*—A partial Latin square (PLS) is an assignment of n symbols to an $n \times n$ grid such that, in each row and in each column, each symbol appears at most once. The partial Latin square extension (PLSE) problem asks to find such a PLS that is a maximum extension of a given PLS. The PLSE problem is NP-hard, and in this paper, we propose a heuristic algorithm for this problem. To design a heuristic, we extend the previous $\frac{1}{2}$-approximation algorithm that utilizes the notion of maximum matching. We show the empirical effectiveness of the proposed algorithm through computational experiments. Specifically, the proposed algorithm delivers a better solution than the original one and local search. Besides, when computation time is limited due to an application reason, it delivers a better solution than IBM ILOG CPLEX, a state-of-the-art optimization solver, especially for large scale "hard" instances.**

I. Introduction

THROUGHOUT the paper, we consider the *partial Latin square extension* (PLSE) problem. Let n denote a natural number. Suppose that we are given an $n \times n$ grid. A *partial Latin square* (*PLS*) is a partial assignment of n *symbols* to the grid so that the *Latin square condition* is satisfied. The Latin square condition requires that, in each row and in each column, each symbol should appear at most once. Given a PLS, the PLSE problem asks to find such a PLS that is a maximum extension of the given one in terms of the number of the filled cells. Specifically, we are asked to assign symbols additionally to as many empty cells as possible so that the Latin square condition is satisfied. The PLSE problem is NP-hard since its decision problem version is NP-complete [1]. The problem has various applications (e.g., scheduling, optical routers [2]) and was first introduced by Kumar, Russel and Sundaram [3], where they proposed several constant-factor approximation algorithms.

In this paper, we propose a heuristic algorithm for the PLSE problem. Assuming practical use, we aim to develop such a heuristic algorithm that delivers better solutions for more instances. We are interested just in empirical performance. We do not make theoretical analysis on approximation ratio, as previous studies for the PLSE problem do.

Let us explain the reason why we dare to propose such a heuristic algorithm. Researchers from discrete optimization or operations research may see that the problem can be formulated as a 0-1 integer programming (IP) problem. The global optimal solution can be found within practical time by the state-of-the-art optimization solvers (e.g., Gurobi optimizer [6], IBM ILOG CPLEX [7]) only when n is moderately small (e.g., no more than 30 in our experience). When n is larger, the solvers may not find even a feasible solution within practical time. For such large scale instances, a heuristic algorithm that delivers a good solution quickly is an alternative. For another application example, a heuristic algorithm may assist the work of the solvers. Many modern solvers admit us to input an initial solution. The solvers utilize the input solution as the first incumbent solution in the branch-and-bound tree. During the solution search, they may prune the solution subspaces where there is no hope of finding a better solution than the incumbent one, which may improve the efficiency of the search. Then the solver that is given a good initial solution is expected to find a better solution than the solver that is given no initial solution (or a poor initial solution) within the same time limit. An effective heuristic algorithm may help us generate a good initial solution.

Our heuristic algorithm is based on the $\frac{1}{2}$-approximation algorithm proposed in [3] that determines the assignment of symbols by solving the n relevant maximum matching problem instances iteratively. We try to improve its empirical performance by extending the model and by giving a reasonable scheme to decide the order in which the maximum matching problem instances are solved.

There are two approximation algorithms that have better approximation ratios than $\frac{1}{2}$. The best bound is $\frac{2}{3}-\varepsilon$, achieved by Hajirasouliha, Jowhari, Kumar and Sundaram [4], where ε is any positive constant. Their algorithm is based on local search, and the constant ε is automatically determined by the parameter on the neighborhood scale. The smaller ε we wish, the longer the running time becomes since we need to set the neighborhood scale large. For example, to beat the second best bound $1-\frac{1}{e}$, achieved by Gomes, Regis and Shmoys [5], the running time becomes $O(n^{26})$, which is surely polynomial but not practical. The second best algorithm [5] employs the linear programming (LP) relaxation approach. The idea is sophisticated but it is not easy to realize the mechanism to improve the empirical performance. On the other hand, the empirical performance of the maximum matching based

This work is supported by JSPS KAKENHI Grant Number 25870661.

978-83-60810-53-8/$25.00

$\frac{1}{2}$-approximation algorithm can be improved by a smaller ingenuity. This is the highlight of the paper.

We show the empirical effectiveness of the proposed algorithm through computational experiments. Specifically, the proposed algorithm tends to deliver a better solution than the original one and the local search based approximation algorithm. Besides, when computation time is limited, as is often the case in practice, the proposed algorithm tends to deliver a better solution than IBM ILOG CPLEX for large scale "hard" instances.

The paper is organized as follows. In Section II, we prepare terminologies and notations. Then we present the proposed heuristic algorithm in Section III. We report the results of the computational experiments in Section IV, and then give concluding remarks in Section V.

II. Preliminaries

A. The PLSE Problem

First we introduce notations on the $n \times n$ grid. The grid consists of n^2 cells. Let us denote $[n] = \{1, 2, \ldots, n\}$. For any $i, j \in [n]$, we denote the cell in the row i and in the column j by (i, j). We denote the set of the n cells in the row i by R_i, and the set of the n cells in the column j by C_j, i.e.,

$$R_i = \{(i, 1), (i, 2), \ldots, (i, n)\} \text{ and}$$
$$C_j = \{(1, j), (2, j), \ldots, (n, j)\}.$$

We also denote the family of R_i'-s by $\mathcal{R}$, and the family of C_j'-s by $\mathcal{C}$, i.e.,

$$\mathcal{R} = \{R_1, R_2, \ldots, R_n\} \text{ and } \mathcal{C} = \{C_1, C_2, \ldots, C_n\}.$$

Clearly we have $\bigcup_{R_i \in \mathcal{R}} R_i = \bigcup_{C_j \in \mathcal{C}} C_j = [n]^2$.

Next we introduce notations on assignment of symbols to the grid. For simplicity, we represent the n symbols by the integers $1, 2, \ldots, n$ in the set $[n]$. We represent an assignment of symbols by an $n \times n$ array, denoted by L. For each cell (i, j), we denote the assigned symbol by $L_{i,j} \in [n] \cup \{0\}$, where $L_{i,j} = 0$ indicates that (i, j) is empty. We define the *domain of* L as $\mathrm{Dom}(L) = \{(i, j) \in [n]^2 \mid L_{i,j} \neq 0\}$. The cardinality $|\mathrm{Dom}(L)|$ equals to the number of the cells that are assigned symbols by L. For simplicity, we may write $|\mathrm{Dom}(L)|$ as $|L|$. An assignment L is an *extension* of L' if $\mathrm{Dom}(L) \supseteq \mathrm{Dom}(L')$ and $L_{i,j} = L'_{i,j}$ holds for any $(i, j) \in \mathrm{Dom}(L')$. When L is an extension of L', we write $L \geqslant L'$. We call L a *partial Latin square* (*PLS*) if, in each row and in each column, every symbol appears at most once. In particular, if all the cells are assigned symbols (i.e., $\mathrm{Dom}(L) = [n]^2$), then we simply call L a *Latin square* (*LS*). We call a PLS L *extensible* (resp., *blocked*) if there exists (resp., does not exist) a PLS $L' \geqslant L$ $(L' \neq L)$. Two PLSs L and L' are *compatible* if the following two conditions hold:

(i) For each (i, j), at least one of $L_{i,j} = 0$ and $L'_{i,j} = 0$ holds.

(ii) The assignment $L \oplus L'$ defined as follows is a PLS;

$$(L \oplus L')_{i,j} = \begin{cases} L_{i,j} & \text{if } L_{i,j} \neq 0 \text{ and } L'_{i,j} = 0, \\ L'_{i,j} & \text{if } L_{i,j} = 0 \text{ and } L'_{i,j} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

An assignment of symbols can be also represented by a set of triples. Let T be a subset of $[n]^3$. The membership $(i, j, k) \in T$ represents that the symbol k is assigned to (i, j). In order to avoid duplicate assignments on the same cell, we assume that, for any different triples (i, j, k) and (i', j', k') in T, at least one of $i \neq i'$ and $j \neq j'$ holds. To convert the triple set representation into the array representation, we define the array $A(T)$ as follows;

$$A(T)_{i,j} = \begin{cases} k & \text{if there is } k \in [n] \text{ such that } (i, j, k) \in T, \\ 0 & \text{otherwise.} \end{cases}$$

For $(i, j, k), (i', j', k') \in T$, when at least two of $i \neq i'$, $j \neq j'$ and $k \neq k'$ hold, we say that they are compatible. Then T is a PLS if any two triples in T are compatible. Let us simplify some notations. When L and $A(T)$ are compatible, we may say that L and T are compatible, and use the expression $L \oplus T$ instead of $L \oplus A(T)$. When a singleton $T = \{(i, j, k)\}$ is compatible with L, we say that (i, j, k) (instead of $\{(i, j, k)\}$) is compatible with L.

We summarize the PLSE problem as follows.

The Partial Latin Square Extension (PLSE) Problem	
Input:	A PLS L_0.
Output:	A PLS $L \geqslant L_0$ that attains the maximum $\lvert L \rvert$.

A polynomial time algorithm for the PLSE problem is called a *ρ-approximation algorithm* if, for any input PLS L_0, it finds $L \geqslant L_0$ such that;

$$\frac{|L| - |L_0|}{|L^*| - |L_0|} \geq \rho,$$

where L^* denotes a global optimal solution. The bound ρ is called the *approximation ratio*.

B. Graph, Maximum Matching, Independent Set

An *undirected graph* (or simply a *graph*) $G = (V, E)$ consists of a set V of *nodes* and a set E of unordered pairs of nodes, where each element in E is called an *edge*. The *degree* of a node $v \in V$ is the number of the edges incident to v. A graph is *bipartite* when V can be partitioned into two disjoint nonempty sets, say V_1 and V_2, so that every edge joins a node in V_1 and a node in V_2. When we emphasize that G should be bipartite, we may write $G = (V_1 \cup V_2, E)$.

A *matching* E' is a subset of E such that no two edges in E' have a node in common. A *maximum matching* is such a matching that attains the largest cardinality. In particular, a maximum matching is called a *perfect matching* if it covers all the nodes in the graph. The size of maximum matching is called a *matching number*, and is denoted by $\nu(G)$. Suppose that the graph $G = (V, E)$ is bipartite. We can find a maximum matching in $O(\sqrt{|V|}|E|)$ time [8]. Note that maximum matching is not necessarily unique. Any edge e in E is classified into one of the following three classes with respect to how it appears in the possible maximum matchings:

Mandatory edge:	e appears in every maximum matching.
Admissible edge:	e appears in at least one (but not every) maximum matching.
Forbidden edge:	e appears in no maximum matching.

The sets of mandatory, admissible and forbidden edges in G are denoted by $ME(G)$, $AE(G)$ and $FE(G)$, respectively. The edge set E of a bipartite graph G can be decomposed into three disjoint sets $ME(G)$, $AE(G)$ and $FE(G)$ by the Dulmage-Mendelsohn decomposition technique [9]. The computation time is dominated by finding a maximum matching, and thus the decomposition can be made in $O(\sqrt{|V|}|E|)$ time. The proposed algorithm repeatedly solves maximum matching problems on bipartite graphs such that $|V_1| = |V_2| = n$. We denote the upper bound on the computation time for one bipartite graph by τ_n, i.e., $\tau_n = O(n^{5/2})$.

Let $G = (V, E)$ be a general graph (not necessarily bipartite). An *independent set* is a subset of V such that any two nodes in the subset are not adjacent. A *maximum independent set* is such an independent set that attains the largest cardinality. The problem of finding a maximum independent set is known as NP-hard [10]. The proposed algorithm solves this problem for a certain purpose. To construct a nearly maximal independent set, we employ the $(\Delta + 2)/3$-approximation algorithm [11], where Δ denotes the maximum degree in the graph. The algorithm is a greedy method such that, starting from $S = \emptyset$, it inserts a node of the smallest degree into S and removes the inserted node and all the adjacent nodes (along with all the incident edges) from the graph. The process is repeated until all nodes are removed from the graph, and finally S is output. The computation time is linear in the numbers of nodes and edges [11].

III. A Heuristic Algorithm for the PLSE Problem

In this section, we propose a heuristic algorithm for the PLSE problem. The algorithm runs like a typical heuristic algorithm for a packing problem; we are given several containers (what we call *compatibility graphs*), each of which is given its capacity (matching number). Then we are asked to pack as many objects (matching edges) in the containers as possible, where packing an object in a container may decrease the capacities of other containers. To construct an approximate solution, we repeat choosing a certain container and packing the objects up to the capacity. The proposed algorithm is inspired by the $\frac{1}{2}$-approximation algorithm that was given in [3]. First we describe the $\frac{1}{2}$-approximation algorithm in Section III-A. Then in Section III-B, we present the proposed algorithm.

A. The Maximum Matching Based $\frac{1}{2}$-Approximation Algorithm

The $\frac{1}{2}$-approximation algorithm works as follows; starting from $L = L_0$, we repeat choosing a PLS L' that is compatible with L and updating $L \leftarrow L \oplus L'$ iteratively. For convenience, when L is updated to $L \oplus L'$, we say that L' is *added* to the PLS L. To decide L', the algorithm utilizes a maximum matching of what we call a compatibility graph. Given a PLS L, the compatibility graph is constructed for any symbol $k \in [n]$. The *compatibility graph for the symbol* k is denoted by $G_L^{\mathrm{symb},k} = (\mathcal{R} \cup \mathcal{C}, E_L^{\mathrm{symb},k})$, where $\mathcal{R}$ and $\mathcal{C}$ are the node sets and $E_L^{\mathrm{symb},k}$ is the edge set such that;

$$E_L^{\mathrm{symb},k} = \big\{\{R_i, C_j\} \subseteq \mathcal{R} \cup \mathcal{C} \mid (i,j,k) \text{ is compatible with } L\big\}.$$

Let $M \subseteq E_L^{\mathrm{symb},k}$ denote any matching of $G_L^{\mathrm{symb},k}$. We define the triple set $T(M)$ as follows;

$$T(M) = \big\{(i,j,k) \mid \{R_i, C_j\} \in M\big\}.$$

Any two triples (i,j,k) and (i',j',k') in $T(M)$ satisfy $i \neq i'$ and $j \neq j'$ since M is a matching of $G_L^{\mathrm{symb},k}$. Thus $T(M)$ is a PLS. Each $(i,j,k) \in T(M)$ is compatible with L from the definition of $E_L^{\mathrm{symb},k}$. Then $T(M)$ and L are compatible.

In the order $k = 1, 2, \ldots, n$, the algorithm finds a maximum matching M^* of $G_L^{\mathrm{symb},k}$ and adds $T(M^*)$ to L iteratively. We show the algorithm in Algorithm 1. Naïvely implemented, the algorithm runs in $O(n\tau_n) = O(n^{7/2})$ time.

Algorithm 1 The $\frac{1}{2}$-approximation algorithm given in [3]

```
L ← L_0
for k = 1, 2, ..., n do
    M* ← a maximum matching of G_L^{symb,k}
    L ← L ⊕ T(M*)
end for
output L
```

B. The Proposed Algorithm

Let us describe our idea for how we improve the empirical performance of the above approximation algorithm. We have denoted the available symbols by integers $1, 2, \ldots, n$ only for convenience. Their numerical order does not have any significant meaning, whereas the above algorithm chooses k in that order. Thus there is room for developing a smart criterion to choose a compatibility graph. (The approximation ratio is still guaranteed even if we choose k arbitrarily; see the proof in Section 5.2 of [3] for detail.)

Besides, the compatibility graphs for the symbols are not the only "containers" in which we can "pack" the maximum matching to increase the objective value. Motivated by the fact that the dimensions of PLS in the triple set representation are symmetric, we introduce the compatibility graph also for each row and for each column. The *compatibility graph for the row* i is denoted by $G_L^{\mathrm{row},i} = (R_i \cup [n], E_L^{\mathrm{row},i})$, where R_i and $[n]$ denote the node sets and $E_L^{\mathrm{row},i}$ is the edge set such that;

$$E_L^{\mathrm{row},i} = \big\{\{(i,j),k\} \subseteq R_i \cup [n] \mid (i,j,k) \text{ is compatible with } L\big\}.$$

Similarly, the *compatibility graph for the column* j is denoted by $G_L^{\mathrm{col},j} = (C_j \cup [n], E_L^{\mathrm{col},j})$, where C_j and $[n]$ denote the

node sets and $E_L^{\mathrm{col},j}$ is the edge set such that;

$$E_L^{\mathrm{col},j} = \big\{\{(i,j),k\} \subseteq C_j \cup [n] \mid (i,j,k) \text{ is compatible with } L\big\}.$$

We show an example of the compatibility graphs in Fig. 1. Let us denote any matching of $G_L^{\mathrm{row},i}$ or $G_L^{\mathrm{col},j}$ by M. We define the triple set $T(M)$ as follows;

$$T(M) = \big\{(i,j,k) \mid \{(i,j),k\} \in M\big\}.$$

Analogously with the case of $G_L^{\mathrm{symb},k}$, the triple set $T(M)$ is a PLS and compatible with L. Thus $T(M)$ can be added to L.

Finally, given a PLS L, we have $3n$ compatibility graphs. Using all of them as containers, we expect to obtain a better solution than the case when we use only the n compatibility graphs for symbols. We denote the set of the n compatibility graphs for symbols by $\mathcal{G}_L^{\mathrm{symb}} = \{G_L^{\mathrm{symb},1}, \ldots, G_L^{\mathrm{symb},n}\}$, the set of the n compatibility graphs for rows by $\mathcal{G}_L^{\mathrm{row}} = \{G_L^{\mathrm{row},1}, \ldots, G_L^{\mathrm{row},n}\}$ and the set of the n compatibility graphs for columns by $\mathcal{G}_L^{\mathrm{col}} = \{G_L^{\mathrm{col},1}, \ldots, G_L^{\mathrm{col},n}\}$. We also denote the union of these graph sets by $\mathcal{G}_L = \mathcal{G}_L^{\mathrm{symb}} \cup \mathcal{G}_L^{\mathrm{row}} \cup \mathcal{G}_L^{\mathrm{col}}$.

Let us introduce the criterion by which we choose one of the $3n$ compatibility graphs. Recall that any edge corresponds to a triple $(i,j,k) \in [n]^3$. When $L_{i,j} = k$, no edge corresponding to (i,j,k) appears in any compatibility graph. For any compatibility graph $G_L \in \mathcal{G}_L$, we denote by $\rho(G_L)$ the number of node pairs that do not appear as edges due to $L_{i,j} = k$. For example, when G_L is the compatibility graph $G_L^{\mathrm{symb},k}$ for the symbol k, $\rho(G_L^{\mathrm{symb},k})$ is defined as follows;

$$\rho(G_L^{\mathrm{symb},k}) = \big|\big\{\{R_i, C_j\} \subseteq \mathcal{R} \cup \mathcal{C} \mid L_{i,j} = k\big\}\big|.$$

In other words, $\rho(G_L^{\mathrm{symb},k})$ indicates the number of cells to which L already assigns the symbol k. Analogously, $\rho(G_L^{\mathrm{row},i})$ and $\rho(G_L^{\mathrm{col},j})$ indicate the numbers of cells in the row i and in the column j to which L assigns certain symbols, respectively. Now we define the criterion function $f(G_L)$ as follows.

$$f(G_L) = \begin{cases} \nu(G_L) + \rho(G_L) & \text{if } \nu(G_L) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Basically, our heuristic algorithm runs as follows; starting from $L = L_0$, the algorithm chooses the compatibility graph $G_L \in \mathcal{G}_L$ such that $f(G_L)$ is the largest. Finding a maximum matching M^* of the chosen graph, the algorithm updates the solution by $L \leftarrow L \oplus T(M^*)$. The above operation is repeated while L is extensible. The repetition is at most $3n$ times; suppose that a compatibility graph G_L^s is chosen at a certain step of the algorithm, where s denotes the superscript of the compatibility graph. Let us denote the updated PLS by $L' = L \oplus T(M^*)$. For any extension $L'' \geqslant L'$, there is no edge in the compatibility graph $G_{L''}^s$. Then we have $\nu(G_{L''}^s) = 0$ and $f(G_{L''}^s) = 0$. Thus $G_{L''}^s$ is never chosen in the rest execution of the algorithm. Finally, when $f(G_L) = 0$ holds for all the $3n$ G_L'-s, L is blocked and the algorithm halts.

The algorithm utilizes the function f in (1) as the criterion to choose a compatibility graph. It may prefer a compatibility graph G_L such that more triples can be added (i.e., larger $\nu(G_L)$) and/or more cells are already assigned symbols (i.e., larger $\rho(G_L)$). The only matching number $\nu(G_L)$ appears to be a more natural criterion. However, it did not yield good results in our preliminary experiments.

We summarize the heuristic algorithm in Algorithm 2. The structure is as above stated, but in Line 3, we introduce the subroutine named COLLECTME for further improvement. The subroutine constructs a PLS that is compatible with L, from the triples of mandatory edges over the $3n$ compatibility graphs. Then it adds the PLS to L, before adding the triple set of a maximum matching of a certain compatibility graph to L.

Let us describe the motivation. Recall that a maximum matching is not necessarily unique in a graph. For example, given the PLS L in Fig. 1, there are 4 maximum matchings in the compatibility graph $G_L^{\mathrm{symb},1}$. We denote these 4 maximum matchings by $M_1^*, \ldots, M_4^*$, where we define;

$$M_1^* = \big\{\{R_1, C_4\}, \{R_2, C_1\}, \{R_3, C_2\}, \{R_4, C_3\}\big\},$$
$$M_2^* = \big\{\{R_1, C_2\}, \{R_2, C_1\}, \{R_3, C_4\}, \{R_4, C_3\}\big\},$$
$$M_3^* = \big\{\{R_1, C_2\}, \{R_2, C_3\}, \{R_3, C_4\}, \{R_4, C_1\}\big\},$$
$$M_4^* = \big\{\{R_1, C_4\}, \{R_2, C_3\}, \{R_3, C_2\}, \{R_4, C_1\}\big\}.$$

The greedy method would choose $G_L^{\mathrm{symb},1}$ for the container since $f(G_L^{\mathrm{symb},1})$ is the largest (which is 4) among all the compatibility graphs. Then the maximum matching algorithm finds arbitrary one M_x^* $(x = 1, \ldots, 4)$ and $T(M_x^*)$ is added to L. No matter which $T(M_x^*)$ is added, the symbol 1 is assigned to exactly 4 cells. Then the symbols 2 and 3 are assigned to the remaining empty cells, and the final solution is obtained. How many cells are filled in the final solution depends on which $T(M_x^*)$ is added. More precisely, when either $T(M_1^*)$ or $T(M_2^*)$ is added, the final solution becomes worse than the case when either $T(M_3^*)$ or $T(M_4^*)$ is added. To analyze the reason, see Fig. 2 for the final solutions that are obtained by extending $L \oplus T(M_1^*)$, $\ldots$, $L \oplus T(M_4^*)$. For $L \oplus T(M_3^*)$, there is another final solution such that the symbol 2 is assigned not to $(4,3)$ but to $(4,4)$, but it does not matter since we discuss how many cells are filled. We claim that the problem should come from the number of the bold cells occupied by the symbol 1. The bold cells in each grid correspond to the mandatory edges in $G_L^{\mathrm{symb},2}$ and $G_L^{\mathrm{symb},3}$, that is, $\{R_3, C_4\}, \{R_4, C_3\} \in E_L^{\mathrm{symb},2}$ and $\{R_1, C_4\}, \{R_2, C_1\} \in E_L^{\mathrm{symb},3}$; see also these graphs in Fig. 1. The matching number $\nu(G_L^{\mathrm{symb},k})$ gives an upper bound on the number of cells that the symbol k is assignable. Occupying bold cells by the symbol 1 may diminish the matching numbers of $G_L^{\mathrm{symb},2}$ and $G_L^{\mathrm{symb},3}$. When $T(M_1^*)$ or $T(M_2^*)$ is added, 3 out of the 4 bold cells are occupied by the symbol 1, while only 1 bold cell is occupied when $T(M_3^*)$ or $T(M_4^*)$ is added.

Based on the above, we consider that mandatory edges should be treated with a greater care in order to approach a

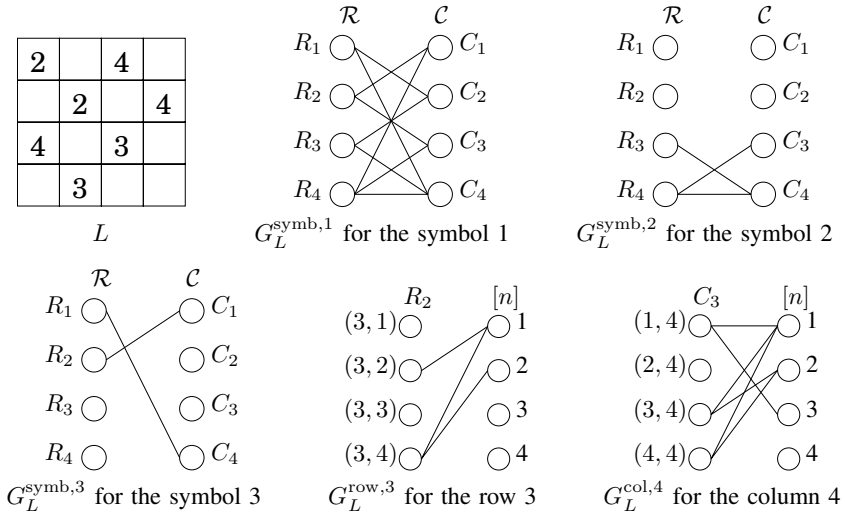


Fig. 1. An example of compatibility graphs for a PLS L

Algorithm 2 The proposed heuristic algorithm for the PLSE problem

```
1: L ← L_0
2: while L is extensible do
3:     L ← CollectME(L)
4:     construct the set G_L of the compatibility graphs for the PLS L
5:     G_L ← a compatibility graph in G_L such that f(G_L) in (1) is the largest
6:     M* ← a maximum matching of G_L
7:     L ← L ⊕ T(M*)
8: end while
9: output L
```

better solution. Our idea is to add as many triples coming from mandatory edges to L as possible. Let ME_L denote the set of all the mandatory edges over the $3n$ compatibility graphs in $\mathcal{G}_L$. We would like to add all the triples in $T(ME_L)$ if possible, but we cannot always do so since there are incompatible triples in $T(ME_L)$ in general. Let us define the graph $H_L = (T(ME_L), F_L)$ such that the triple set $T(ME_L)$ is the node set and the edge set F_L is defined as follows;

$$F_L = \big\{\{(i,j,k),(i',j',k')\} \subseteq T(ME_L) \mid (i,j,k) \text{ and } (i',j',k') \text{ are incompatible}\big\}.$$

Clearly, any independent set of H_L is a PLS and compatible with L. This motivates us to solve the maximum independent set problem on H_L. Since the problem is NP-hard, we find a nearly maximum solution by the greedy algorithm [11] that we described in Section II-B.

We summarize the subroutine CollectME in Algorithm 3. The subroutine repeats enumerating all mandatory edges over the $3n$ compatibility graphs, computing a nearly maximum independent set, and adding the independent set to L until there is no mandatory edge in any compatibility graph.

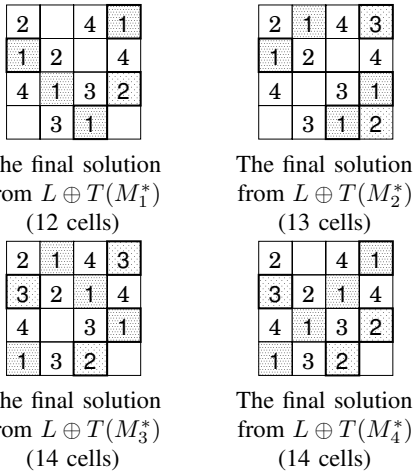


Fig. 2. The final solutions obtained by extending $L \oplus M_1^*, \ldots, L \oplus M_4^*$: the number of the filled cells is indicated in the parentheses

Algorithm 3 The subroutine CollectME

```
1: loop
2:     construct the set G_L of compatibility graphs
3:     ME_L ← the set of the mandatory edges over the 3n compatibility graphs in G_L
4:     if ME_L = ∅ then
5:         break the loop
6:     end if
7:     construct the graph H_L = (T(ME_L), F_L)
8:     I ← an independent set in the graph H_L
9:     L ← L ⊕ I
10: end loop
11: return L
```

C. Computational Complexity

We claim that the proposed algorithm should run in $O(n^{11/2})$ time. The algorithm consists of the main routine in Algorithm 2 and the subroutine COLLECTME in Algorithm 3. In the main routine, the most time-consuming task is to find maximum matchings of all the compatibility graphs. Since the iteration times of the main routine is at most $3n$, the running time is $3n \times 3n \times \tau_n = O(n^{9/2})$.

The iteration times of the subroutine amounts to $O(n^2)$ over the whole execution of the algorithm since there are at most $3n^2$ mandatory edges; there are $3n$ compatibility graphs, and in each graph, there are at most n mandatory edges. The most time-consuming task is not the greedy algorithm for the maximum independent set problem but to find maximum matchings of all the compatibility graphs; one can see that the former takes $O(n^3)$ time, whereas the latter takes $O(n^{7/2})$ time. Then the running time is $O(n^2) \times O(n^{7/2}) = O(n^{11/2})$.

IV. COMPUTATIONAL EXPERIMENTS

In this section, we present some experimental results to show how the proposed algorithm is effective in practice. Specifically, we show that the proposed algorithm tends to deliver a good solution quickly in comparison with other approximation algorithms. We also show that, when computation time is limited, the proposed algorithm tends to deliver a better solution than IBM ILOG CPLEX [7], a state-of-the-art optimization solver.

A. Experimental Settings

We describe how to generate problem instances. This issue has been studied in the field of *constraint programming*. In our experiments, we generate PLSE instances based on the QCP (quasigroup completion problem) framework that was discussed in [12]. An instance is characterized by two parameters. One is a natural number n, the side length of the grid, and the other is p $(0 \leq p \leq 1)$, which is the ratio of the number of pre-assigned cells over n^2. The QCP framework starts with the empty $n \times n$ grid and assigns a random symbol to a random empty cell so that the resulting assignment is PLS. The assignment is repeated until $\lfloor pn^2 \rfloor$ cells are assigned symbols. In the context of the decision problem, when the pre-assigned ratio p is small (resp., large), an instance is more (resp., less) likely to be satisfiable since there are less (resp., more) constraints. Many papers have reported the easy-hard-easy phase transition with respect to p (e.g., [13]); the instances are less computationally tractable when p is intermediate, e.g., $0.4 \leq p \leq 0.7$.

We refer to the proposed algorithm as OURS. For comparison, we will consider the algorithm that does not call the subroutine COLLECTME. We refer to this version of the algorithm to OURS-NOSUB. We also use 4 methods: MATCHING, LS, CPMATH and CPCP. The point is that MATCHING and LS are approximation algorithms that do not necessarily provide a global optimal solution, whereas CPMATH and CPCP are such softwares that retain exact algorithms.

a) MATCHING: The $\frac{1}{2}$-approximation algorithm [3] that forms the basis of the proposed algorithm and was explained in Section III-A.

b) LS: The $\frac{2}{3} - \varepsilon$ approximation algorithm [4] based on local search that achieves the best approximation ratio among those studied so far. Let T_0 denote the triple set representation of the given PLS L_0. Suppose that a non-negative number r is given. For any set $T \subseteq [n]^3$ of triples such that $T \geqslant T_0$, we denote the *neighborhood of* T by $\mathcal{N}_r(T)$ that is defined as follows;

$$\mathcal{N}_r(T) = \{T' = (T \setminus S) \cup S' \mid S \subseteq T \setminus T_0,\ |S| \leq r,\ S' \subseteq [n]^3,\ |S'| = |S| + 1,\ T' \text{ is a PLS}\}.$$

We call r the *radius* of neighborhood. LS starts with $T = T_0$ and iterates choosing a solution $T' \in \mathcal{N}_r(T)$ and updating $T \leftarrow T'$ until no solution exists in the neighborhood, i.e., $\mathcal{N}_r(T) = \emptyset$. The larger r is, the better the approximation ratio is, but at the same time, the more the computation time may become. Although $r \geq 7$ defeats the second best bound $1 - \frac{1}{e}$ given in [5], the time bound becomes $O(n^{26})$ in naïve implementation $(r = 7)$, which is not practical. We use $r \leq 4$ since any larger r was too time consuming in our preliminary experiments.

c) CPMATH: IBM ILOG CPLEX Optimizer (version 12.4) that solves the PLSE problem by means of mathematical programming. The PLSE problem can be formulated as a 0-1 integer programming problem. See [3] for the formulation. We set the time limit parameter to 6.0×10^2 seconds, i.e., if the computation time exceeds 6.0×10^2 seconds, the computation is terminated and the best solution found so far is output. We set all the other parameters to default values.

d) CPCP: IBM ILOG CPLEX CP Optimizer (version 12.4) that solves the PLSE problem by means of constraint programing. We formulate the PLSE problem as a constraint optimization problem as follows.

$$\begin{array}{ll} \text{maximize} & |L| \\ \text{subject to} & \forall i \in [n],\ \text{all-different\_except\_0}(L_{i,1}, \ldots, L_{i,n}), \\ & \forall j \in [n],\ \text{all-different\_except\_0}(L_{1,j}, \ldots, L_{n,j}), \\ & \forall (i,j) \in [n]^2,\ (L_0)_{i,j} \neq 0 \Rightarrow L_{i,j} = (L_0)_{i,j}, \\ & \forall (i,j) \in [n]^2,\ L_{i,j} \in [n] \cup \{0\}. \end{array}$$

In the above formulation, the all-different_except_0 constraint [14] is a special case of the typical all-different constraint, requiring that the variables should take all-different values except the variables assigned 0. We set the level of default inference (`DefaultInferenceLevel`) and the level of all-different inference (`AllDiffInferenceLevel`) to `extended`, i.e., the most sophisticated constraint propagation technique is used. We set the time limit parameter to 6.0×10^2 seconds. We set all the remaining parameters to default values.

All the experiments are conducted by our PC that carries 2.80 GHz CPU and 4GB main memory.

TABLE I
THE NUMBER OF WINS, TIES AND LOSTS OF OURS AGAINST OURS-NOSUB AND MATCHING OVER THE 1000 INSTANCES

			the pre-assigned ratio p			
			0.2	0.4	0.6	0.8
$n = 10$	vs. OURS-NOSUB	(win)	34	273	136	17
		(tie)	957	577	376	976
		(lost)	9	150	488	7
	vs. MATCHING	(win)	744	966	633	190
		(tie)	247	24	217	804
		(lost)	9	10	150	6
$n = 80$	vs. OURS-NOSUB	(win)	809	915	913	809
		(tie)	73	23	20	31
		(lost)	118	62	67	160
	vs. MATCHING	(win)	998	1000	1000	999
		(tie)	1	0	0	1
		(lost)	1	0	0	0

TABLE II
THE NUMBER OF WINS, TIES AND LOSTS OF OURS AGAINST 10 LS'-S OVER THE 100 INSTANCES

vs. LS		the pre-assigned ratio p						
		0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 10$	(win)	570	815	839	412	145	32	0
$r = 4$	(tie)	413	176	116	218	281	689	962
	(lost)	17	9	45	370	574	279	38
$n = 20$	(win)	920	975	994	999	796	157	88
$r = 3$	(tie)	61	16	4	1	76	143	508
	(lost)	19	9	2	0	128	700	404
$n = 30$	(win)	1000	1000	1000	1000	1000	774	283
$r = 2$	(tie)	0	0	0	0	0	69	173
	(lost)	0	0	0	0	0	157	544

B. Results

First, we compare OURS with OURS-NOSUB and MATCHING. We generate 1000 instances for each pair (n, p) such that $n \in \{10, 20, \ldots, 80\}$ and $p \in \{0.1, 0.2, \ldots, 0.9\}$. We solve all the generated instances by each algorithm and compare the algorithms in terms of $|L|$. We show the typical results in Table I. The table shows the numbers of wins, ties and losts of OURS against the other algorithms over the 1000 instances. When n is small (i.e., $n = 10$), OURS is competitive with OURS-NOSUB in general, except that it is rather worse for $p = 0.6$. OURS outperforms MATCHING for smaller p, and for larger p, they are rather competitive. When n is larger, OURS is more likely to outperform the rest two algorithms. In particular, when $n = 80$, OURS wins over OURS-NOSUB in more than 80% of the instances and wins over MATCHING in almost all the instances. Based on these, we claim that the subroutine COLLECTME should play a significant role in improving the solution especially when n is large. We also claim that OURS should be superior to MATCHING, the original approximation algorithm. Let us discuss the computation time. MATCHING takes less than 1 seconds in all the instances from $n = 10$ to 80. Given the side length n, OURS takes more computation time for the instances that are generated by smaller pre-assigned ratio p. For every $n = 10, 20, \ldots, 80$, the average computation time for $p = 0.1$ is twice to three times of that for $p = 0.9$ approximately. For example, when $n = 10$ (resp., 80), the average computation time for $p = 0.1$ is 7.7×10^{-3} (resp., 2.0×10^{1}) seconds, whereas that for $p = 0.9$ is 4.1×10^{-3} (resp., 7.5×10^{0}) seconds. We consider the reason as follows; when p is smaller, there are more edges in the compatibility graphs in the earlier steps of the algorithm. Then it should take more time to compute a maximum matching since the algorithm runs in $O(\sqrt{|V|}|E|)$ time, which is proportional to the number of edges, while $|V| = 2n$ holds for any compatibility graph. The current implementation of OURS is rather naïve, and we believe that it is possible to improve the computation time to some extent by devising the data structure. This is left for future work.

Next, we compare OURS with LS. This time we generate 100 instances for given n and p. We solve each instance by OURS once, whereas we solve it by LS 10 times; in our implementation, LS proceeds to a solution randomly chosen from the neighborhood. Changing the seed of pseudo random numbers, we solve the instance by executing LS 10 times. We examine the $100 \times 10 = 1000$ cases to count the number of wins, ties and losts of OURS. We show the results in Table II, along with the values of the radius r. We describe the reason why we set r to the indicated values. The computation time of LS is affected by r significantly. LS is just a heuristic, and its computation time should be shorter than the time needed for the optimization solvers (i.e., CPMATH or CPCP) to find global optimum solutions. We set the radius r based on this philosophy. For example, when $n = 30$ and $p = 0.2$, the solvers find global optimum solutions in only 1.0×10^{1} seconds for any instance. However, when $r = 3$, LS takes at least 3.0×10^{2} seconds until it outputs a local optimal solution, which we infer from our preliminary experiments. Thus we set $r = 2$ even though the average computation time is still 2.4×10^{1} seconds. Note that OURS is much faster than LS; it takes less than 1 second to solve any instance. As expected from Table II, OURS should outperform LS in this perspective for $n \geq 30$ and $p \leq 0.7$. Note that this range includes "hard" instances in the context of the phase-transition.

Finally, we compare OURS with CPMATH and CPCP. When n gets larger, the solvers are less likely to find global optimum solutions in practical time. To illustrate this, in Table III, we show the number of instances such that the solvers can find global optimum solutions in 6.0×10^{2} seconds. In this experiment, we generate 100 instances for each (n, p). It is shown that, when p is smaller (resp., larger), CPCP finds global optimal solutions in more (resp., less) instances than CPMATH. When p is intermediate ($p = 0.6$ and 0.7 in particular), the solvers hardly find global optimum solutions. When the solvers cannot find a global optimum solution within the time limit, they output the best solution among those searched. In such cases, OURS yields better solutions than the solvers although the computation time is much shorter. We illustrate this for $n = 60$ in Fig. 3. The figure shows the average of $|L|$ (vertical axis) with respect to the change of p (horizontal axis). OURS+CPCP represents CPCP that is given an incumbent solution generated by OURS. In the

TABLE III
THE NUMBER OF INSTANCES THAT THE OPTIMIZATION SOLVERS CAN FIND A GLOBAL OPTIMAL SOLUTION WITHIN 6.0×10^2 SECONDS

		the pre-assigned ratio p						
		0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 40$	CPMATH	0	1	3	71	7	0	100
	CPCP	100	100	99	89	3	0	0
$n = 50$	CPMATH	0	0	0	0	0	0	100
	CPCP	99	94	72	18	0	0	0
$n = 60$	CPMATH	0	0	0	0	0	0	11
	CPCP	92	66	25	1	0	0	0

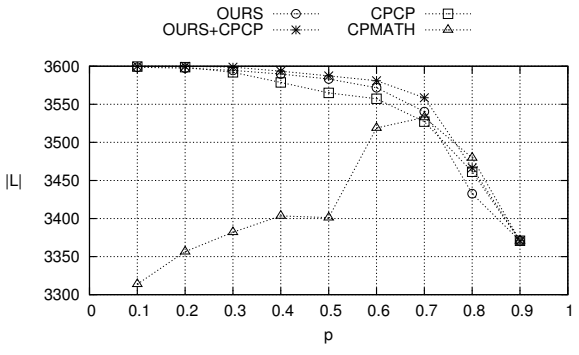


Fig. 3. The averaged objective value $|L|$ with respect to the change of p ($n = 60$)

solution search, CPCP retains the incumbent solution to prune solution subspaces such that there is no hope of finding a better solution. Giving a good incumbent solution may improve the efficiency of the search, or equivalently, it may result a better solution than the default CPCP within the same time limit. When $p \leq 0.2$, OURS and OURS+CPCP are competitive with CPCP and are much better than CPMATH. In this case, CPCP finds a global optimal solution within the time limit in almost all instances (see Table III), and thus OURS is also expected to do so. Surprisingly, when $0.3 \leq p \leq 0.7$ (i.e., "hard" instances), OURS is better than the solvers, and OURS+CPCP is even better. Note that OURS solves an instance with $n = 60$ within only 6.0×10^0 seconds on average, while the solvers take at most 6.0×10^2 seconds. We claim that these results should show the empirical effectiveness of the proposed algorithm.

V. CONCLUDING REMARKS

In this paper, we proposed a heuristic algorithm for the PLSE problem by extending the $\frac{1}{2}$-approximation algorithm [3] that utilizes the notion of maximum matching. We showed how the proposed algorithm is effective in practice through computational experiments.

This paper just shows the possibilities of our approach for the PLSE problem. We believe that the algorithm can be improved further by analyzing its behavior for various instances and the structures of compatibility graphs carefully. It is an alternative to apply metaheuristic techniques such as *genetic algorithm* and *simulated annealing* (*SA*). In fact, SA was applied to *sudoku* problem [15] that asks to find a PLS that is a maximum extension of a given PLS and that satisfies additional constraints. Different from SA, there is no adjustable parameter in the proposed algorithm whose tuning is often exhaustive. Being rather simple, the proposed algorithm delivers a good solution quickly. We can say that the proposed algorithm is a $\frac{1}{3}$-approximation algorithm since it delivers a blocked PLS and any blocked PLS is a $\frac{1}{3}$-factor solution [3]. It is interesting and challenging future work to analyze a nontrivial approximation ratio of the proposed algorithm.

The decision problem version of the PLSE problem has been studied in the field of constraint programming intensively. The problem is whether there is a Latin square that is an extension of a given PLS L_0. The answer is yes only when there is a perfect matching for every compatibility graph G_{L_0}. Any forbidden edge can be ignored in the solution search and the typical constraint programming technique eliminates all of the forbidden edges [16]. Our algorithm is different from this in that ours does not utilize forbidden edges but utilizes mandatory edges.

We have considered a heuristic algorithm for the PLSE problem, assuming practical use. We hope that other researchers take interest in this problem and work on it in the nearest future.

REFERENCES

[1] C. J. Colbourn, "The complexity of completing partial Latin squares," *Discrete Applied Mathematics*, vol. 8, 1984, pp. 25–30.
[2] R. A. Barry and P. A. Humblet, "Latin routers, design and implementation," *IEEE/OSA Journal of Lightwave Technology*, vol. 11-5, 1993, pp. 891–899.
[3] R. Kumar, A. Russel, and R. Sundaram, "Approximating Latin square extensions," *Algorithmica*, vol. 24-2, 1999, pp. 128–138.
[4] I. Hajirasouliha, H. Jowhari, R.Kumar, and R. Sundaram, "On completing Latin squares," In *Proceedings of STACS 2007*, Lecture Notes in Computer Science vol. 4393, 2007, pp. 524–535.
[5] C. P. Gomes, R. G. Regis, and D. B. Shmoys, "An improved approximation algorithm for the partial Latin square extension problem," *Operations Research Letters*, vol. 32-5, 2004, pp. 479–484.
[6] Gurobi Optimizer, http://www.gurobi.com/
[7] IBM ILOG CPLEX, http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/
[8] J. E. Hopcroft and R. M. Karp, "An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs," *SIAM Journal on Computing*, vol. 2-4, 1973, pp. 225–231.
[9] R. Cymer, "Dulmage-Mendelsohn canonical decomposition as a generic pruning technique," *Constraints*, vol. 17, 2012, pp. 234–272.
[10] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York and Oxford; 1979.
[11] M. M. Halldórsson and J. Radhakrishnan, "Greed is good: approximating independent sets in sparse and bounded-degree graphs," *Algorithmica*, vol. 18, 1997, pp. 145–163.
[12] R. Barták, "On generators of random quasigroup problems," In *Proceedings of CSCLP 2005*, 2006, pp. 164–178.
[13] C. P. Gomes and D. Shmoys, "Completing quasigroups or Latin squares: a structured graph coloring problem," In *Proceedings of Computational Symposium on Graph Coloring and Generalizations*, 2002.
[14] N. Beldiceanu, M. Carlsson and, J. X. Rampson, "Global constraint catalog," *Technical Report Swedish Institute of Computer Science*, T2005-08, 2006.
[15] R. Lewis, "Metaheuristics can solve sudoku puzzles," *Journal of Heuristics*, vol. 13-4, 2007, pp. 387–401.
[16] J. C. Régin, "A filtering algorithm for constraints of difference in CSPs," In *Proceedings of the twelfth national conference on artificial intelligence*, 1994, pp. 362–367.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 355–362

Fair optimization with advanced aggregation operators in a multicriteria facility layout problem

Jarosław Hurkała
Warsaw University of Technology
Institute of Control & Computation Engineering
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
Email: J.Hurkala@elka.pw.edu.pl

Adam Hurkała
Warsaw University of Technology
Institute of Control & Computation Engineering
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
Email: A.Hurkala@elka.pw.edu.pl

***Abstract*—In this paper we address a mining operation problem that is a special case of Quadratic Assignment Problem and which belongs to the class of facility layout problems. The considered problem is static and discrete, but the set of possible locations is larger than the set of facilities. We distinguish multiple types of equal-area facilities (mines, processing and auxiliary facilities). Mines can be placed only on selected locations (deposits of various resources), and the production volume of each type of facility depends on the adjacency of other facilities. We examine two situations: when the number of each type of facility is given, and when only the total number of facilities is specified. The goal is to maximize the production. This problem is multi-objective and we use advanced aggregation operators (OWA/WOWA) to achieve fair solutions. A comparison of results obtained with list-based threshold accepting meta-heuristic and simulated annealing algorithm is presented.**

I. Introduction

THE facility layout is a broad class of problems that has been a subject of interest for researches around the world due to many real-life applications as well as significant scientific value. Layout problems are generally NP-Hard [1], which makes them perfect test ground for various optimization algorithms and heuristic approaches. Because of different assumptions and vast variety of considerations, it is difficult to arrive at one, common definition that would encompass all possible kinds of problems associated with finding an optimal placement of facilities in some predefined area.

In terms of layout evolution, the problem can be static, i.e. the key information about the problem (e.g. parameters and relationship between facilities) are constant, or dynamic, when possible changes over subsequent time periods additionally have to be taken into account (see [2] for both formulations).

By analyzing the layout formulations, we can classify the problems as continuous, in which the facilities can be placed anywhere within the designated area [3], or discrete, when the facilities can be placed only in specified locations [4]. The discrete formulation can be considered as a Quadratic Assignment Problem (QAP) [5], in which the assignment function is a bijection, i.e. the number of facilities is equal to the number of possible locations. Moreover, in QAP for each pair of locations a distance is specified and for each pair of facilities a weight is given.

In the literature we can find many different models with single criterion, e.g. total material handling cost, travel time of parts, travel distance, and problems with multiple criteria, for which at the same time different objectives are minimized, e.g. material handling, tools and information flow. For the latter problems, most researchers use a simple aggregation function, e.g. a weighted sum ([6], [7]).

In a great number of articles about the facility layout problems a meta-heuristic is chosen to solve the underlaying optimization problem. The evolutionary algorithms, and most notably genetic algorithms, are the most popular choice [3], [4], [8], [9], [10], [11]. The other group of approximated approaches that attract the interest of researchers are the random search methods - tabu search ([12], [13]) and simulated annealing ([14], [15]).

In this paper we address a layout problem that is static and discrete, but contrary to QAP the considered assignment function is not a bijection. Moreover, we define the relationship between the facilities and their location in a more complex way.

The goal of the facility layout problem considered in this paper is to find a non-overlapping planar arrangement of n rectangular facilities within a given rectangular site that maximizes the production of facilities, and unlike [11], [13] we assume equal-area facilities.

Furthermore, we distinguish multiple types of facilities (and therefore multiple types of products), hence the problem is multicriterial. We show that the application of the ordered weighted averaging (OWA) aggregation introduced by [16] is a consistent, reasonable and fairness-preserving approach to model a multicriteria facility layout problem.

To solve the problem, we propose a novel (not mentioned in the recent reviews and surveys on facility layout problems [17], [18]) heuristic approach based on the threshold accepting algorithm – a list-based threshold accepting meta-heuristic, that was successfully applied in a job-shop scheduling problem [19]. We compare the effectiveness of this algorithm with our design of simulated annealing ([20], [27]).

The paper is organized as follows. In Section 2, we present the definition of the multicriteria facility layout problem that we discuss in this article. Sections 3 briefly portrays the fair aggregation operators. In Section 4, we describe list-based threshold accepting and simulated annealing algorithms, and explain the implementation details for the problem at hand.

978-83-60810-53-8/$25.00

The results of the experiments are shown in Section 5, and the concluding remarks are presented in Section 6.

II. Problem definition

A. Basic notation and definitions

The facility layout problem that we discuss in this paper can be simply described as a problem of choosing a location $l \in \mathcal{L}$ for each facility $f \in \mathcal{F}$ so that the objective function (that will be defined later on) is maximized.

Remark 2.1: We assume equal-area facilities which are placed within a rectangular $n \times m$ grid. Each location is thus a cell in the grid and the considered problem is combinatorial in nature.

Remark 2.2: We assume $|\mathcal{L}| > |\mathcal{F}|$ so that the problem does not boil down to simply finding a correct permutation of facility locations. In consequence, the assignment function is not a bijection, as in QAP formulation.

We distinguish three main classes of facilities : $\mathcal{F} = \mathcal{P} \cup \mathcal{A} \cup \mathcal{M}$, where $\mathcal{P} = \bigcup_{t \in \mathcal{T}} \mathcal{P}_t$ is the set of $|\mathcal{T}|$ types of processing facilities, $\mathcal{A} = \bigcup_{v \in \mathcal{V}} \mathcal{A}_v$ is the set of $|\mathcal{V}|$ types of auxiliary facilities, and $\mathcal{M} = \bigcup_{u \in \mathcal{U}} \mathcal{M}_u$ denotes the set of $|\mathcal{U}|$ types of mines.

Remark 2.3: We assume $|\mathcal{T}| > 1$ and $|\mathcal{U}| > 1$, i.e. there are at least one processing facility and at least one mine.

Corollary 1: The problem is multicriterial in nature.

Remark 2.4: Throughout this paper we will commonly denote $\mathcal{C} = \mathcal{T} \cup \mathcal{U}$ as the set of all the facility types that are responsible for production, and $\mathcal{D} = \mathcal{T} \cup \mathcal{V} \cup \mathcal{U}$ as the set of all possible facility types.

The set of locations is divided into two subsets $\mathcal{L} = \mathcal{S} \cup \mathcal{E}$, respectively locations $\mathcal{S}$ for processing/auxiliary facilities, and extraction sites $\mathcal{E} = \bigcup_{r \in \mathcal{R}} \mathcal{E}_r$ where deposits of $|\mathcal{U}|$ types of resources $\mathcal{R} = \bigcup_{u \in \mathcal{U}} \mathcal{R}_u$ are located, and where the corresponding types of mines can be placed.

Remark 2.5: We assume that $\mathcal{S} \cap \mathcal{E} = \emptyset$ (i.e. a processing/auxiliary facility cannot be placed on an extraction site and a mine can be located only at one of the resource deposits).

In our facility layout problem the objective is to maximize the production output of processing facilities and mines.

Definition 1 (Basic production output): The basic production output denoted by $O_c^B, c \in \mathcal{C}$ is a production volume of a specified type of facility (processing facility or mine) regardless of its location and the locations of other facilities.

The auxiliary facilities by definition are not producing anything themselves, nevertheless they play an important role by affecting production of other facilities. Moreover, we assume that resource deposits have an impact on the facilities production as well and that in general the facilities themselves can positively affect each other.

Definition 2 (Distance function): The relationship between facilities is based on a binary distance function $d : \mathcal{L} \times \mathcal{L} \to \{0, 1\}$, i.e. either a pair of locations is adjacent or not. The facilities are placed in cells of a rectangular grid, and thus we assume that a relationship exists only between two adjacent facilities and facilities adjacent to resource deposits.

Definition 3 (Adjacent location): The set of all locations adjacent to a location $l \in \mathcal{L}$ will be denoted by $\delta_l(l) \subset \mathcal{L}$.

Definition 4 (Adjacent facility): The set of all facilities of type $d \in \mathcal{D}$ adjacent to a facility placed on location l will be denoted by $\delta_d^F(l) \subset \mathcal{F}$.

Definition 5 (Adjacent resource deposit): The set of all resource deposits of type $u \in \mathcal{U}$ adjacent to a facility f placed on location l will be denoted by $\delta_u^R(l) \subset \mathcal{R}$.

Definition 6 (Extra production output - facility): The extra production output denoted by $O_{cd}^F : \mathbb{N} \to \mathbb{R}$ is an additional production volume of a processing facility/mine $c \in \mathcal{C}$ proportional to the number of surrounding (adjacent) facilities $d \in \mathcal{D}$.

Definition 7 (Extra production output - resource deposit): The extra production output denoted by $O_{cu}^R : \mathbb{N} \to \mathbb{R}$ is an additional production volume of a processing facility/mine $c \in \mathcal{C}$ proportional to the number of surrounding (adjacent) resource deposits $u \in \mathcal{U}$.

Definition 8 (Processing facility production output): The (final) production output of a processing facility is defined as follows:

$$o_c(l) = O_c^B \left(\sum_{d \in \mathcal{D}} O_{cd}^F(\delta_d^F(l)) + \sum_{u \in \mathcal{U}} O_{cu}^R(\delta_u^R(l)) \right) \quad c \in \mathcal{C} \tag{1}$$

B. Optimization model

The mining operation problem considered in this article can be formulated as follows:

$$\max_x \ [o_1, o_2, \ldots, o_{|\mathcal{C}|}] \tag{2}$$

$$o_c^B = \sum_{l \in \mathcal{L}} O_c^B x_{lc} \quad c \in \mathcal{C} \tag{3}$$

$$o_c^F = \sum_{l \in \mathcal{L}} \sum_{i \in \delta(l)} \sum_{j \in \mathcal{D}} O_{cj}^F x_{ij} \quad c \in \mathcal{C} \tag{4}$$

$$o_c^R = \sum_{l \in \mathcal{L}} \sum_{i \in \delta(l)} \sum_{j \in \mathcal{U}} O_{cj}^R x_{ij} \quad c \in \mathcal{C} \tag{5}$$

$$o_c = o_c^B(o_c^F + o_c^R) \quad c \in \mathcal{C} \tag{6}$$

$$\sum_{c \in \mathcal{C}} x_{cl} = 1 \quad l \in \mathcal{L} \tag{7}$$

$$\sum_{l \in \mathcal{L}} \sum_{c \in \mathcal{C}} x_{cl} = |\mathcal{F}| \tag{8}$$

$$\sum_{l \in \mathcal{L}} x_{cl} = \alpha_c \quad c \in \mathcal{C} \tag{9}$$

$$x_{cl} \in \{0, 1\} \quad c \in \mathcal{C}, l \in \mathcal{L} \tag{10}$$

where α_c is a parameter that denotes the designated number of each facility type (the constraint (9) is optional).

III. Ordered weighted averaging

In the ordered weighted averaging aggregation of outcomes (OWA) $\mathbf{y} = (y_1, \ldots, y_m)$ the weights $\mathbf{w} = (w_1, w_2, \ldots, w_m)$ are assigned to the ordered values (i.e., to the smallest value,

the second smallest and so on) rather than to the specific criteria:

$$A_w = \sum_{i=1}^{m} w_i \theta_i(\mathbf{y}) \tag{11}$$

where $(\theta_1(\mathbf{y}), \theta_2(\mathbf{y}), \ldots, \theta_m(\mathbf{y}) = \Theta(\mathbf{y})$ is the ordering map $R^m \to R^m$ with $\theta_1(\mathbf{y}) \leq \theta_2(\mathbf{y}) \leq \ldots \leq \theta_m(\mathbf{y})$ and there exists a permutation τ of set I such that $\theta_i(\mathbf{y}) = y_{\tau(i)}$ for $i = 1, 2, \ldots, m$. The OWA operator provides a parameterized family of aggregation operators, which include many of the well-known operators such as the maximum, the minimum, the k-order statistics (including Conditional Value at Risk), the median and the arithmetic mean. The OWA satisfies the properties of strict monotonicity, impartiality and, in the case of monotonic increasing weights $w_1 > w_2 > \ldots > w_{m-1} > w_m$, the property of equitability [21] (satisfies the principle of transfers – equitable transfer of an arbitrary small amount from the larger outcome to a smaller outcome results in a more preferred achievement vector). Every solution maximizing the OWA function is then an equitably efficient solution to the original multiple criteria problem. Moreover, for linear multiple criteria problems every equitably efficient solution can be found as an optimal solution to the OWA aggregation with appropriate weights. Thus the OWA-based optimization generates the so-called equitably efficient solutions (cf. [22] for the formal axiomatic definition). According to [22] and [23], equitable efficiency expresses the concept of fairness, in which all system entities have to be treated equally and in the stochastic problems equitability corresponds to the risk aversion [24].

For the facility layout problem (2)–(10) the final OWA aggregation of the outcomes p_i for all types of production facilities $i \in \mathcal{C}$ can be stated as the following LP model:

$$\max \sum_{k=1}^{|\mathcal{C}|} k w'_k t_k - \sum_{k=1}^{|\mathcal{C}|} \sum_{c \in \mathcal{C}} w'_k d_{ck} \tag{12}$$

subject to

$$d_{ck} \geq t_k - o_c,\ d_{ck} \geq 0 \quad k = 1, 2, \ldots, |\mathcal{C}|,\ c \in \mathcal{C} \tag{13}$$

$$\mathbf{o} \in O \tag{14}$$

where coefficients w'_i are defined as $w'_m = w_m$ and $w'_i = w_i - w_{i+1}$ for $i = 1, 2, \ldots, m-1$, $\mathbf{o} = [o_c]_{c \in \mathcal{C}}$ and O is a feasible set of production output vectors defined by (3)–(10).

The weighted ordered weighted averaging (WOWA) aggregation is a generalization of the OWA aggregation, that allows assigning importance weights to specific criteria [25]. Those weights could express, for example, relative importance of different facility types. The weights assigned to ordered values will be further called preferential weights.

Let $\mathbf{p} = (p_1, \ldots, p_m)$ be an m-dimensional vector of importance weights such that $p_i \geq 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^{m} p_i = 1$. The corresponding Weighted OWA aggregation of vector $\mathbf{y}$ is defined [11] as follows:

$$A_{w,p} = \sum_{i=1}^{m} \omega_i \theta_i(\mathbf{y}) \tag{15}$$

with

$$\omega_i = w^*\Big(\sum_{k \leq i} p_{\tau(k)}\Big) - w^*\Big(\sum_{k<i} p_{\tau(k)}\Big), \tag{16}$$

where w^* is an increasing function interpolating points $(i/m, \sum_{k \leq i} w_k)$ together with the point $(0, 0)$ and τ representing the ordering permutation for $\mathbf{y}$ (i.e. $y_{\tau(i)} = \theta(\mathbf{y})$). Moreover, function w^* is required to be a straight line when the points can be interpolated in this way. We assume the piecewise linear interpolation function w^* which is the simplest form of the required interpolation.

Note, that the piecewise linear functions may be built with various number of breakpoints, not necessarily equal to number of criteria m [25]. Thus, any nonlinear function can be well approximated by a piecewise linear function with appropriate number of breakpoints. Therefore, we will consider weights vectors $\mathbf{w}$ of dimension n not necessarily equal to m. It is even possible to define a generalized WOWA aggregation where the preferential weights w_k are allocated to an arbitrarily defined grid of ordered outcomes defined by quantile breakpoints (see [25] and references therein).

As shown in [25], maximization of an equitable WOWA aggregation with decreasing preferential weights $w_1 \geq w_2 \geq \ldots \geq w_n$ may be implemented as the LP expansion of the original problem. In the case of the facility layout problem (2)–(10), this can be stated as follows:

$$\max \sum_{k=1}^{n} w'_k \left[\frac{k}{n} t_k - \sum_{c \in \mathcal{C}} p_c d_{ck} \right] \tag{17}$$

subject to

$$d_{ck} \geq t_k - o_c,\ d_{ck} \geq 0 \quad k = 1, 2, \ldots, n,\ c \in \mathcal{C} \tag{18}$$

$$\mathbf{o} \in O \tag{19}$$

If the importance weights are equal $p_c = 1/|\mathcal{C}|$, the model reduces to the OWA aggregation.

IV. Algorithms

A. List-based threshold accepting

List-based threshold accepting algorithm (LBTA) [19] is an extent of threshold accepting meta-heuristic, which belongs to the randomized search class of algorithms. The search trajectory crosses the solution space by moving from one solution to a random neighbor of that solution, and so on. Unlike the greedy local search methods which consist of choosing a better solution from the neighborhood of the current solution until such can be found (hill climbing), the threshold accepting allows choosing a worse candidate solution based on a threshold value. In the general concept of the threshold accepting algorithm it is assumed that a set of decreasing threshold values is given before the computation or an initial threshold value and a decrease schedule is specified.

Algorithm 1 Creating the list of threshold values

Require: Initial solution s_1, list size S, set of move operators $m \in M$

$i \leftarrow 0$
while $i < N$ **do**
  $m \leftarrow random(M)$
  $s_2 \leftarrow m(s_1)$
  if $C(s_1) \leq C(s_2)$ **then**
    $\Delta \leftarrow (C(s_2) - C(s_1))/C(s_1)$
    $list \leftarrow list \cup \{\Delta\}$
    $i \leftarrow i + 1$
  else
    $s_1 \leftarrow s_2$
  end if
end while
return $list$

The rate at which the values decrease controls the trade-off between diversification (associated with large threshold values) and intensification (small threshold values) of the search. It is immensely difficult to predict how the algorithm will behave when a certain decrease rate is applied for a given problem without running the actual computation. It is also very common that the algorithm with the same parameters works better for some problem instances and significantly worse for others. These reflections led to the list-based threshold accepting branch of threshold accepting meta-heuristic.

In the list-based threshold accepting approach, instead of a predefined set of values, a list is dynamically created during a presolve phase of the algorithm. The list, which in a way contains knowledge about the search space of the underlying problem, is then used to solve it.

1) Creating the list of threshold values: The first phase of the algorithm consists of gathering information about the search space of the problem that is to be solved. From an initial solution a neighbor solution is created using a move function (perturbation operator) chosen at random from a predefined set of functions. If the candidate solution is better than the current one, it is accepted and becomes the current solution. Otherwise, a threshold value is calculated as a relative change between the two solutions:

$$\Delta = (C(s_2) - C(s_1))/C(s_1) \tag{20}$$

and added to the list, where $C(s_i)$ is the objective function value of the solution $s_i \in S$, and S is a set of all feasible solutions. For this formula to work, it is silently assumed that $C : S \rightarrow \mathbb{R}_+ \cup \{0\}$. This procedure is repeated until the specified size of the list is reached. For the algorithm overview see Algorithm 1.

2) Optimization procedure: The second phase of the algorithm is the main optimization routine, in which a solution to the problem is found. The algorithm itself is very similar to that of the previous phase. We start from an initial solution, create new solution from the neighborhood of current one using one of the move function, and compare both solutions.

Algorithm 2 LBTA optimization procedure

Require: Initial solution s_1, thresholds list L, set of move operators $m \in M$

$i \leftarrow 0$
$s^* \leftarrow s_1$
while $i \leq N$ **do**
  $m \leftarrow random(M)$
  $s_2 \leftarrow m(s_1)$
  $i \leftarrow i + 1$
  if $C(s_2) \leq C(s_1)$ **then**
    if $C(s_2) \leq C(s^*)$ **then**
      $s^* \leftarrow s_2$
    end if
    $s_1 \leftarrow s_2$
    $i = 0$
  else
    $\Delta_{new} \leftarrow (C(s_2) - C(s_1))/C(s_1)$
    if $\Delta_{new} < max(list)$ **then**
      $list \leftarrow list \setminus \{max(list)\}$
      $list \leftarrow list \cup \{\Delta_{new}\}$
      $s_1 \leftarrow s_2$
      $i = 0$
    end if
  end if
end while
return $list$

If the candidate solution is better, it becomes the current one. Otherwise a relative change is calculated. To this point algorithms in both phases are identical. The difference in the optimization procedure is that we compare the threshold value with the largest value from the list. If the new threshold value is larger, then the new solution is discarded. Otherwise, the new threshold value replaces the value from the list, and the candidate solution is accepted to next iteration. The best solution found during the optimization process is considered final.

The list-based threshold accepting algorithm also incorporates early termination mechanism: after a (specified) number of candidate solutions is subsequently discarded, the optimization is stopped, and the best solution found so far is returned.

The optimization procedure of the list-based threshold accepting algorithm is shown in Algorithm 2.

B. Simulated annealing

The optimization process of the simulated annealing algorithm can be described in the following steps. At the start, an initial solution is required. Then, repeatedly, a candidate solution is randomly chosen from the neighborhood of the current solution. If the candidate solution is the same or better than the current one, it is accepted and replaces the current solution. Even if the generated solution is worse than the current one, it still has a chance to be accepted with, so called, acceptance probability. This probability is a function of difference between objective value of the current and

Algorithm 3 Simulated Annealing

Require: Initial solution s_1
1: $s^* \leftarrow s_1$
2: **for** $i = 1$ **to** N **do**
3: **for** $t = 1$ **to** N_{const} **do**
4: $s_2 \leftarrow perturbate(s_1)$
5: $\delta \leftarrow C(s_2) - C(s_1)$
6: **if** $\delta \leq 0$ or $e^{-\delta/k\tau} > random(0,1)$ **then**
7: $s_1 \leftarrow s_2$
8: **end if**
9: **if** $C(s_2) < C(s^*)$ **then**
10: $s^* \leftarrow s_2$
11: **end if**
12: **end for**
13: $\tau \leftarrow \tau * \alpha$
14: **end for**
15: **return** s^*

the candidate solution and depends on a control parameter taken from the thermodynamics, called temperature. The temperature is decreased after a number of iterations, and the process continues as described above. The optimization is stopped either after a maximum number of iterations or when a minimum temperature is reached. The best solution found during the annealing process is considered final.

For the algorithm overview see Algorithm 3.

In order to apply the simulated annealing algorithm to the facility layout problems, the annealing process must be adapted and the parameters adjusted appropriately. Similarly to the threshold decrease rate in LBTA, the temperature decrease (also known as the cooling process) in simulated annealing consists of decreasing the temperature by a so called reduce factor. The parameters associated with this mechanism are:

1) Initial temperature.
2) Function of temperature decrease in consecutive iterations.
3) The number of iterations at each temperature (Metropolis equilibrium).
4) Minimum temperature at which the algorithm terminates or alternatively the maximum number of iterations as the stopping criterion.

Let τ be the temperature and α be the reduce factor. Then the annealing scheme can be represented as the following recursive function:

$$\tau^{i+1} = \alpha * \tau^i, \tag{21}$$

where i is the number of current iteration in which the cooling schedule takes place.

Second building block of SA that has to be customized is the acceptance probability function, which determines whether to accept or reject candidate solution that is worse than the current one. The most widely used function is:

$$p(\delta, \tau) = e^{-\delta/k\tau}, \tag{22}$$

TABLE I
SIMULATED ANNEALING PARAMETERS

Parameter	Description	Value
α	Reduce factor	$1 - \frac{5}{N}$
τ^0	Initial temperature	0.99
δ^0	Minimal difference between solutions	1
p^0	Initial acceptance probability	1
N_{const}	Number at each constant temperature	10
N	Number of SA iterations	100000

where $\delta = E(s_2) - E(s_1)$ is the difference between the objective value (denoted by E) of the candidate (s_2) and the current solution (s_1), and k is the Boltzmann constant found by:

$$k = \frac{\delta^0}{\log \frac{p^0}{\tau^0}}, \tag{23}$$

where δ^0 is an estimated difference between objective values of two solutions, p^0 is the initial value of the acceptance probability and τ^0 is the initial temperature. Notice that we use decimal logarithm rather than natural, which is most widely seen in the literature and, rather than average, we use minimal difference between solutions.

For the overview of the parameters applied in the facility layout problem, see Table I.

C. Neighborhood function

The most problem-specific mechanism of both the SA and the LBTA algorithm that always needs a different approach and implementation is the procedure of generating a candidate solution from the neighborhood of the current one, which is called a perturbation scheme, transition operation/operator or a move function. Although there are many ways to accomplish this task, we have examined the following techniques:

1) Interchange two adjacent processing facilities.
2) Interchange two processing facilities at random locations.
3) Move a single processing facility to an adjacent, empty location.
4) Change type of the facility.
5) Move a mine from current resource deposit to another deposit that is not occupied.

In order to generate a new solution, the LBTA algorithm applies one of the aforementioned operators chosen at random to the current solution. SA on the other hand uses only one, compound operator (a combination of operators which together allow to make a transition from initial to any feasible solution) during the whole optimization procedure.

D. Implementation details

1) Zero elements: In the first phase of the list-based threshold accepting algorithm the list is populated with values of relative change between two solutions $\Delta \geq 0$. After careful consideration, we believe that including zeros in the list is a misconception. In the actual optimization procedure, i.e. the second phase, the threshold value is computed only if the new solution is worse than the current one, which means that the

calculated relative change will always have a positive value ($\Delta_{new} > 0$). The new threshold value is compared with the largest value from the list (T_{hmax}). Thus, we can distinguish three cases:

1) $T_{hmax} = 0$: since thresholds are non-negative from definition, in this case the list contains all zero elements and it will not change throughout the whole procedure (T_{hmax} is constant). Comparing a positive threshold value Δ_{new} against zero yields in discarding the candidate solution. The conclusions are as follows:
   a) it does not matter how many zeros are in the list, the effective size of the list is equal to one,
   b) the algorithm is reduced to hill climbing algorithm that accepts candidate solutions which are at least as good as the current one.
2) $T_{hmax} > 0$ and $\Delta_{new} < T_{hmax}$: the largest (positive) threshold value from the list T_{hmax} is replaced by a smaller (positive) threshold value Δ_{new}. The number of zero elements in the list remains the same throughout the whole procedure and therefore is completely irrelevant to the optimization process. The effective list size is equal to the number of positive elements.
3) $T_{hmax} > 0$ and $\Delta_{new} \geq T_{hmax}$: the new solution is discarded and the list remains unchanged.

The main idea behind the list is to control the diversification and intensification of the search process. In the early stage of the search the algorithm should allow to cover as much solution space as possible, which means that the thresholds in the list are expected to be large enough to make that happen. In the middle stage, the algorithm should slowly stop fostering the diversification and begin to foster the intensification of the search. In the end stage, the intensification should be the strongest, i.e. the list is supposed to contain smaller and smaller threshold values, which induces discarding of worse solution candidates. In consequence, the algorithm is converging to a local or possibly even a global optimum.

2) Stopping criterion: Even though equipped with an early-termination mechanism, the LBTA algorithm does not have a solution space independent stopping criterion. If the number of subsequently discarded worse solutions is set too high, the algorithm will run for an unacceptable long time (it has been observed during preliminary tests). Hence, we propose to use a global counter of iterations so that when a limit is reached, the algorithm will terminate gracefully.

V. Numerical experiments

The numerical experiments were performed on a number of randomly generated different size problems. Each instance is defined by the number of locations $|\mathcal{L}|$, which in turn determines the number of facilities: $|\mathcal{F}| = |\mathcal{L}|/2$, and the number of resource deposits: $|\mathcal{R}| = |\mathcal{L}|/5$ (the number of resource deposits of each type are equal or differ by one). However, the following settings were the same for every instance, regardless of the problem size:

1) the number of types of processing facilities $|\mathcal{T}| = 2$,
2) the number of types of auxiliary facilities $|\mathcal{V}| = 2$,
3) the number of types of mines/resources $|\mathcal{U}| = 2$,
4) the extra production output depending on adjacent facilities is defined for pairs of processing facility type t_i and auxiliary facility type v_i: $O^F_{t_i v_i} = 0.4, i = 1, 2$ and for one pair of processing facility types: $O^F_{t_1 t_2} = 0.2$ (the relationship is not symmetrical),
5) the extra production output depending on adjacent resource deposits is defined for pairs of processing facility type t_i and resource deposit type u_i: $O^R_{t_i u_i} = 0.5, i = 1, 2$,
6) the total basic production output is constant, and is randomly distributed among the (basic) production output of processing facilities and mines.

Algorithm 4 Random distribution algorithm

Require: Number of elements N, value to distribute v
1: $list \leftarrow v$
2: **for** $i = 1$ **to** N **do**
3: $x_0 \leftarrow max(list)$
4: $list \leftarrow list \setminus \{x_0\}$
5: $x_1 \leftarrow random(x_0)$
6: $x_2 \leftarrow x_0 - x_1$
7: $list \leftarrow list \cup \{x_1\} \cup \{x_2\}$
8: **end for**
9: $list \leftarrow shuffle(list)$
10: **return** $list$

The algorithm of random distribution of total number of facilities among the facilities of each type, and distribution of total basic production output among the basic production output of different types of processing facilities and mines has been based on bisecting technique described in [26]. For the overview see Algorithm 4.

For the WOWA aggregation operator we have chosen already examined weights generation methodology ([27]): all the weights, except two, are strictly decreasing numbers with the step 0.1, while the two selected weights ($k = \lfloor n/3 \rfloor$ and $k = \lfloor 2n/3 \rfloor$) differ from the previous ones by 0.5.

After preliminary tests, based on the objective value and iteration number at which the best solution was found, we have arrived at the list size of 2000, for which the LBTA algorithm works at peak performance.

The algorithms were implemented in Java, and all the experiments were performed on a 3.1 GHz processor. The results are the average of 30 tries for each instance.

We have examined two situations: when the number of each type of facility is given (constraint (9)) – see Table II, and when only the total number of facilities is specified – the results are presented in Table III.

Both algorithms produced similar results in terms of both the solution quality (mean value and standard deviation) and computation time. For instances with fixed number of each facility type, SA has smaller standard deviation, but when it comes to instances with fixed total number of facilities, it is the other way around.

TABLE II
LAYOUT PROBLEMS WITH THE NUMBER OF EACH TYPE OF FACILITY GIVEN

$\|\mathcal{L}\|$	$\|\mathcal{F}\|$	$\|\mathcal{R}\|$	LBTA			SA			$LBTA/SA$
			mean objective value	standard deviation	time[s]	mean objective value	standard deviation	time[s]	
100	50	20	18719	19.1	1.745	18713	26.9	1.810	1.0004
100	50	20	11150	84.8	1.288	11205	76.8	1.824	0.9952
100	50	20	15406	6.4	1.688	15404	10.1	1.872	1.0001
100	50	20	11426	223.3	1.977	11612	142.3	2.039	0.9840
100	50	20	17599	48.6	1.700	17605	15.7	1.925	0.9997
144	72	29	27889	74.8	2.039	27882	80.7	2.261	1.0002
144	72	29	23739	139.4	2.008	23747	122.1	2.220	0.9997
144	72	29	27050	73.8	2.249	27055	72.0	2.438	0.9998
144	72	29	28103	49.8	2.114	28116	47.4	2.139	0.9995
144	72	29	19224	214.9	1.881	19626	205.5	2.405	0.9795
196	98	39	32146	237.0	2.572	32325	101.0	2.601	0.9945
196	98	39	23012	97.6	2.908	22993	70.0	3.281	1.0008
196	98	39	36957	81.6	2.346	36913	112.2	2.393	1.0012
196	98	39	34567	51.0	2.285	34563	51.2	2.690	1.0001
196	98	39	39903	90.7	2.943	39921	80.3	3.005	0.9995
256	128	51	57569	93.9	3.483	57629	46.8	4.155	0.9990
256	128	51	50916	167.1	3.763	51006	117.8	3.839	0.9982
256	128	51	59488	264.7	3.267	59658	189.7	3.565	0.9971
256	128	51	40648	141.4	2.061	40740	80.1	2.855	0.9977
256	128	51	33602	252.1	2.851	33744	189.2	3.022	0.9958
324	162	65	71843	122.3	4.149	71756	217.2	4.187	1.0012
324	162	65	26893	316.5	1.169	27174	148.2	2.455	0.9896
324	162	65	49630	126.3	3.617	49672	105.7	3.618	0.9992
324	162	65	43264	240.4	2.658	43347	144.9	2.908	0.9981
324	162	65	57886	401.5	3.583	58280	256.7	3.616	0.9933
400	200	80	92696	438.0	4.571	93413	267.3	5.027	0.9923
400	200	80	48824	288.3	3.068	49195	141.7	3.619	0.9925
400	200	80	66068	148.0	3.828	66403	172.8	3.874	0.9950
400	200	80	51355	382.8	3.612	51800	259.4	4.240	0.9914
400	200	80	72824	185.5	4.824	72848	168.5	5.159	0.9997

VI. CONCLUSION

We are not convinced that one algorithm supersedes the other, which only means that both heuristics are equally good and can be successfully applied to difficult, combinatorial problems, like the one considered in this paper.

By considering only the idea behind the algorithm, we find the LBTA meta-heuristic a little bit more appealing than SA because of the concept of the list. The list holds the key to control the optimization process. The larger the list, the longer the diversification stage will last, because there will be more threshold values to replace and higher probability of that happening. On the other hand, the smaller the list, the stronger the intensification, which in some cases can be more desired (e.g. when the algorithm cannot find a good solution within the given number of iterations). This gives the unique possibility to the algorithm designer to control the whole optimization process with just one parameter.

REFERENCES

[1] Garey, M. R.,& Johnson, D. S. "Computers and intractability: A guide to the theory of NP-completeness". *A Series of Books in the Mathematical Sciences*. W. H. Freeman and Company, 1979.

[2] Kouvelis, P., Kurawarwala, A. A., & Gutierrez, G. J. "Algorithms for robust single and multiple period layout planning for manufacturing systems". *European Journal of Operations Research*, 63(2), pp. 287–303, 1992.

[3] Dunker, T., Radonsb, G., & Westkämpera, E. "Combining evolutionary computation and dynamic programming for solving a dynamic facility layout problem". *European Journal of Operational Research*, 165(1), pp. 55–69, 2005.

[4] Fruggiero, F., Lambiase, A., & Negri, F. "Design and optimization of a facility layout problem in virtual environment". *In Proceeding of ICAD 2006*, pp. 2206–, 2006.

[5] Kaku, Bharat K.; Thompson, Gerald Luther; and Carnegie Mellon University Design Research Center., "An exact algorithm for the general quadratic assignment problem". *Tepper School of Business.* Paper 944. 1983.

[6] Chen, C. W., & Sha, D. Y. "Heuristic approach for solving the multi-objective facility layout problem". *International Journal of Production Research*, 43(21), pp. 4493–4507, 2005.

[7] Ye M., Zhou G. "A local genetic approach to multiobjective, facility layout problems with fixed aisles". *International Journal of Production Research*, 45, pp. 5243–5264, 2007.

[8] Pierreval, H., Caux, C., Paris, J. L., & Viguier, F. "Evolutionary approaches to the design and organization of manufacturing systems". *Computers & Industrial Engineering*, 44(3), pp. 339–364, 2003.

[9] Rajasekharan, M., Peters, B.A. and Yang, T. "A genetic algorithm for facility layout design in flexible manufacturing systems", *Int. J. Production Research*, Vol. 36, No. 1, pp. 95–110, 1998.

[10] Wang, M. J., Hu, M. H., & Ku, M. H. "A solution to the unequal area facilities layout problem by genetic algorithm". *Computers in Industry*, 56(2), pp. 207–220, 2005.

[11] Lee, Y. H., & Lee, M. H. "A shape-based block layout approach to facility layout problems using hybrid genetic algorithm". *Computers & Industrial Engineering*, 42, pp. 237–248, 2002.

[12] Abdinnour-Helm, S. and Hadley, S.W. "Tabu search based heuristics for multi-floor facility layout". *Int. J. Production Research*, Vol. 38, No. 2, pp. 365–383, 2000.

[13] McKendall, A.R. and Hakobyan, A. "Heuristics for the dynamic facility layout problem with unequal-area departments". *European Journal of Operational Research*, Vol. 201, No. 1, pp. 171–182, 2010.

[14] McKendall Jr., A.R., Shanga, J. and Kuppusamy, S. "Simulated annealing heuristics for the dynamic facility layout problem". *Computers and Operations Research*, Vol. 33, pp. 2431–2444, 2006.

TABLE III
LAYOUT PROBLEMS WITH THE TOTAL NUMBER OF FACILITIES GIVEN

$\|\mathcal{L}\|$	$\|\mathcal{F}\|$	$\|\mathcal{R}\|$	LBTA			SA			$LBTA/SA$
			mean objective value	standard deviation	time[s]	mean objective value	standard deviation	time[s]	
100	50	20	26406	41.5	2.298	26409	44.0	2.411	0.9999
100	50	20	24227	34.5	2.346	24235	27.3	2.400	0.9997
100	50	20	25111	32.6	2.344	25126	25.2	2.373	0.9994
100	50	20	26617	15.6	2.325	26599	40.2	2.397	1.0007
100	50	20	23024	41.2	2.286	23030	37.7	2.382	0.9998
144	72	29	38084	42.7	3.007	38120	41.0	3.032	0.9990
144	72	29	33482	23.6	2.969	33462	24.7	3.055	1.0006
144	72	29	34542	76.6	2.867	34490	92.8	3.178	1.0015
144	72	29	38675	56.3	2.877	38742	64.3	2.931	0.9982
144	72	29	39691	65.5	2.831	39754	63.1	2.928	0.9984
196	98	39	53543	112.8	3.153	53511	139.5	3.316	1.0006
196	98	39	51694	40.5	3.686	51748	71.0	3.742	0.9990
196	98	39	59454	51.9	3.689	59509	39.9	3.803	0.9991
196	98	39	52764	206.0	3.670	52714	207.3	3.912	1.0009
196	98	39	54205	64.0	3.538	54234	93.3	3.604	0.9995
256	128	51	68487	202.1	4.755	68456	227.9	4.827	1.0004
256	128	51	67708	49.7	4.532	67779	71.7	4.611	0.9990
256	128	51	92840	173.7	4.752	93035	204.4	4.838	0.9979
256	128	51	70598	71.8	4.533	70678	108.1	4.586	0.9989
256	128	51	56229	157.2	4.802	56245	143.4	4.821	0.9997
324	162	65	86166	110.4	5.564	86315	162.3	5.531	0.9983
324	162	65	77472	102.0	5.551	77480	118.2	5.476	0.9999
324	162	65	69985	51.4	5.562	70020	68.7	5.572	0.9995
324	162	65	90832	148.8	5.337	90801	192.5	5.378	1.0003
324	162	65	88273	185.5	5.412	88452	195.7	5.418	0.9980
400	200	80	146217	195.3	6.982	146485	299.3	6.992	0.9982
400	200	80	132636	273.7	7.096	132663	326.2	7.105	0.9998
400	200	80	124096	227.0	7.042	124082	402.6	7.063	1.0001
400	200	80	98076	155.2	6.243	98079	147.8	6.246	1.0000
400	200	80	86521	165.1	6.532	86574	94.7	6.648	0.9994

[15] Ioannou, G. "An integrated model and a decomposition-based approach for concurrent layout and material handling system design". *Computers and Industrial Engineering*, Vol. 52, pp.459–485, 2007.

[16] Yager, R.R. "On ordered weighted averaging aggregation operators in multicriteria decision making". IEEE Trans. Systems, Man and Cybernetics 18, pp. 183–190, 1988.

[17] Drira, A., Pierreval, H. and Hajri-Gabouj, S., "Facility layout problems: A survey". *Annual Reviews in Control*, 31, pp. 255–267, 2007.

[18] Kundu, A. and Dan, P.K., "Metaheuristic in facility layout problems: current trend and future direction". *Int. J. Industrial and Systems Engineering*, Vol. 10, No. 2, 2012.

[19] Lee, D.S., Vassiliadis, V.S., Park, J.M. "A novel threshold accepting meta-heuristic for the job-shop scheduling problem". *Computers & Operations Research*, 31, pp. 2199–2213, 2004.

[20] Hurkała, J. and Hurkała, A. "Effective Design of the Simulated Annealing Algorithm for the Flowshop Problem with Minimum Makespan Criterion". *9th International Conference on Decision Support for Telecommunications and Information Society*, September 5-6, 2011, Warsaw, Poland [Journal of Telecommunications and Information Technology, no. 2, 2012].

[21] Ogryczak, W., Śliwiński, T. "On Solving Linear Programs with the Ordered Weighted Averaging Objective". *European Journal of Operational Research*, 148, pp. 80–91, 2003.

[22] Kostreva, M.M., Ogryczak, W., "Linear optimization with multiple equitable criteria". *RAIRO Operations Research*, 33, pp. 275–297, 1999.

[23] Rawls, J., "The Theory of Justice". *Cambridge: Harvard Univ Press*, 1971.

[24] Bell, D.E, Raiffa, H,. "Risky choice revisited, in Bell at all, Decision Making Descriptive, Normative and Prescriptive Interactions". *Cambridge University Press*, Cambridge, pp. 99–112, 1988.

[25] Ogryczak, W., Śliwiński, T., "On Efficient WOWA Optimization for Decision Support under Risk". *International Journal of Approximate Reasoning*, 50, pp. 915–928, 2009.

[26] M. Steinbach, G. Karypis, V. Kumar, "A Comparison of Document Clustering Techniques". *KDD workshop on text mining*, 34, pp. 109–111, 2000.

[27] Hurkała, J. and Śliwiński, T., "Fair flow optimization with advanced aggregation operators in Wireless Mesh Networks", *Federated Conference on Computer Science and Information Systems*, Wrocław, Poland, 9–12 September, 2012 [IEEE, Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 415–421, 2012].

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 363–370

Time dependent global optimization via Bayesian inference and Sequential Monte Carlo sampling

Piotr Kopka
1) National Centre for Nuclear Research, Świerk- Otwock, Poland
2) Institute of Computer Science of the Polish Academy of Sciences, Warsaw, Poland
Email: piotr.kopka@ncbj.gov.pl

Anna Wawrzynczak
1) National Centre for Nuclear Research, Świerk-Otwock, Poland
2) Institute of Computer Sciences, Siedlce Univesity, Poland
Email: a.wawrzynczak@ncbj.gov.pl

Mieczyslaw Borysiewicz
1) National Centre for Nuclear Research, Świerk-Otwock, Poland
Email: manhaz@ncbj.gov.pl

***Abstract*—In many areas of application it is important to estimate unknown model parameters in order to model precisely the underlying dynamics of a physical system. In recent years, Sequential Monte Carlo (SMC) methods have become a very popular tool for Bayesian parameter estimation. In this case, the problem of finding the best parameters configuration comes to the optimization issue which is to determine the best fit. In this paper, the application of this approach to the classical global optimization problem is described. We consider the situation when optimized functions are dynamical i.e. the global extremum is changing in time. For this purpose, we adapt two dimensional Ackley and four-dimensional Wood functions. Our aim is to find the most probable localization of the extremum in each time with the use of the Bayesian approach joined with the Markov Chain Monte Carlo (MCMC) and SMC algorithms. We propose a mechanism for dynamic tuning of the proposal distribution in SMC. The approach is based on the Metropolis-Hastings algorithm, combined with a resampling mechanism to achieve better results. We have examined different version of the proposed SMC and MCMC algorithms in terms of effectiveness to estimate the probabilistic distributions. The effect is demonstrated using two benchmark optimization problems. Computed results show that the proposed mechanisms can significantly improve optimization results compared to standard MCMC.**

I. Introduction

CONSIDER the general optimization problem (OP) designed with a time aspect i.e. the global extremum is changing its position with time (see e.g. Fig. 1). Suppose that various possibilities for a OP are defined by some parameters $\phi \in \Phi$, where Φ denotes the bounded space of parameters. As far as we have uncertainty connected with the best parameters configuration which provides optimum, we can express it in a form of a probability density function $P(\phi|D)$. $P(\phi)$ is the prior probability function that we can estimate based on the currently available information D [1]. Moreover, if new data D, related to the behavior of the optimization function, become available, it can be used for updating the prior distribution of searched parameters value $P(\phi)$ using Bayes theorem. This way we can obtain the posterior distribution $P(\phi|D)$ i.e. the distribution updated by the new information. In design optimization algorithms, the goal is to find the optimal values of the parameters set that minimizes (maximizes) the considered function. We consider that effectiveness of searching for the function extremum in subsequent time step can be increased by taking advantage from the information about the location of extremum in previous stages.

Previously, we have applied the methodology combining Bayesian inference with Markov Chain Monte Carlo (MCMC) methods to the problem of the contaminant source localization based only on the substance concentrations registered by distributed sensors network ([2] and [3]).

In this paper, we propose the application of the Sequential Monte Carlo (SMC) methods combined with the Bayesian inference to the global optimization problem. We present the possibility to connect MCMC and SMC to provide additional benefit in the process of event reconstruction. Proposed algorithms were tested on two benchmark functions where optimum was moving with time.

II. Optimization algorithm theoretical preliminaries

A. Bayesian inference

A good introduction to Bayesian theory can be found in [4] and [5]. Bayes' theorem, as applied to optimization problem:

$$P(\phi|D) = \frac{P(D|\phi)P(\phi)}{P(D)} \tag{1}$$

where ϕ represents possible configuration of optimization function parameters and D is the value of the optimized function at given point.

For our problem, Bayes' theorem describes the conditional probability $P(\phi|D)$ of optimum parameters (configurations of variables ϕ) given observed value of function under consideration (D). This conditional probability $P(\phi|D)$ is also known as the posterior distribution and is related to the probability of the D conforming to a given parameters configuration $P(D|\phi)$, and to the possible model configurations $P(\phi)$, before updating by new information D. The probability $P(D|\phi)$, for fixed D, is called the likelihood function, while $P(\phi)$ is the prior distribution. $P(D)$ is the marginal distribution of D and is called prior predictive distribution. $P(D)$ serves as a scaling factor and is this case is equal 1. So, in our case the

978-83-60810-53-8/$25.00

Bayes theorem can be written as follows:

$$P(\phi|D) \propto P(D|\phi)P(\phi) \tag{2}$$

To estimate the unknown function's optimum parameters ϕ using (2), the posterior distribution $P(D|\phi)$ must be sampled. $P(D|\phi)$ quantifies the likelihood of a set of measurements D given the function's optimum parameters ϕ.

We use a sampling procedure with the Metropolis-Hastings algorithm to obtain the posterior distribution $P(\phi|D)$. This way we completely replace the Bayesian formulation with a stochastic sampling procedure to explore the optimized function parameters' space and to obtain a probability distribution for the optimum location.

B. The likelihood function

A measure indicating the quality of the current state of Markov chain is expressed in terms of a likelihood function. This function is proportional to considered global optimization function $H(.)$:

$$ln[P(D|\phi)] = ln[\lambda(\phi)] \propto H(\phi) \tag{3}$$

After calculating value of the likelihood function for the proposed state its acceptance is performed as follows:

$$\frac{ln(\lambda_{prop})}{ln(\lambda)} \geq U(0,1) \tag{4}$$

where λ_{prop} is the likelihood value of the proposal state, λ is the previous likelihood value, and $U(0,1)$ is a random number generated from a uniform distribution in the interval $(0,1)$.

It is important to note that condition (4) is more likely to be satisfied if the likelihood of the proposal is only slightly lower than the previous likelihood value. It gives a chance to choose even a little "worse" state, because the probability of acceptance depends directly on the quality of proposed state.

C. Posterior distribution

The posterior probability distribution (2) is computed directly from the resulting samples defined by the algorithm described below and is estimated with

$$P(\phi|D) \equiv \hat{\pi}^N(\phi) = \frac{1}{N}\sum_{i=1}^{N}\delta(\phi_i - \phi). \tag{5}$$

$P(\phi|D)$ represents the probability of a particular parameters configuration ϕ. Equation (5) is a sum over the entire samples set of length N of all the sampled values ϕ_i. Thus $\delta(\phi_i - \phi) = 1$ when $\phi_i = \phi$ and 0 otherwise. Consequently, if a Markov chain spends several iterations at the same location value of $P(\phi|D)$ increases through the summation (increasing the probability for those optimum parameters).

D. Sequential Monte Carlo

Sequential Monte Carlo (SMC) is designed to sample from dynamic posterior distributions. The SMC methods are easy to parallelize - the different Monte Carlo proposals can be generated and evaluated in parallel. A good introduction to SMC is present in [6], [7], [8] .

E. Sequential importance resampling

Sequential importance resampling (SIR) is a sequential version of importance sampling (IS) and combines IS with resampling procedure [9] . At the center of the SMC approach in our case is the generation of a weighted sample using IS method. IS uses a proposal distribution $q(.)$, that is close to target distribution $\pi(.)$ and from which it is easy to generate samples. The basic methodology is given below.

1) Generate a sample of size N from the proposal distribution $q(\phi)$:
$$\phi_{(i)} \sim q(\phi), i = 1, ..., N \tag{6}$$
2) Compute the importance weights:
$$\breve{w}(\phi_{(i)}) \propto \frac{\pi(\phi_{(i)})}{q(\phi_{(i)})}, i = 1, ..., N \tag{7}$$
and define
$$w(\phi_{(i)}) = \frac{\breve{w}(\phi_{(i)})}{\sum_{j=1}^{N}\breve{w}(\phi_{(j)})} \tag{8}$$
3) The distribution $\pi(\cdot)$ is then approximated by
$$\breve{\pi}^N(\phi) \equiv \sum_{i=1}^{N} w(\phi_{(i)})\delta(\phi_i - \phi) \tag{9}$$
which places the probability mass $w(\phi_{(1)}), ..., w(\phi_{(N)})$ on the support points $\phi_{(1)}, ..., \phi_{(N)}$.

Hence, the weights would be proportional to the value of likelihood. In our case to calculate the weight we use of the following formula,which is related to the likelihood function (3):

$$\breve{w}(\phi_{(i)}) \propto \frac{1}{ln[\lambda(\phi_{(i)})]}, i = 1, ..., N \tag{10}$$

Resampling is used to avoid the situation when almost all (except only a few) of the importance weights are close to zero (problem of degeneracy of the algorithm). Basic idea of resampling methods is to eliminate samples which have small normalized importance weights and to concentrate upon samples with large weights. So,:

1) for $i = 1, ..., N$ are chosen samples with indexes $k(i)$ distributed according to the discrete distribution with N elements satisfying
$$P(k(i) = l) = w(\phi_{(i)}) \tag{11}$$
for $l = 1, ..., N$,
2) then for $i = 1, ..., N$ for samples $M_{k(i)}$ are assigned the weights
$$w(\phi_{k(i)}) = \frac{1}{N}. \tag{12}$$

A sufficient number of draws is called Effective Sample Size (ESS) and is equal:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} w(\phi_{(i)})^2}. \tag{13}$$

where $w(\phi_{(i)})$ are normalized weights. If all weights are equal $1/N$ then effective sample size is N. In the contrast to a situation where all weights $= 0$, except for one weight $= 1$, effective sample size is equal 1.

F. MCMC prior to SMC

The SMC algorithm needs some set of samples to be initialized. An ideal way to generate this initial sample is using MCMC data from first K iterations in all time steps. The resulting equally weighted MCMC set of samples can then be passed on to SMC for processing in the subsequent iteration.

First, the scanning algorithm starts from the randomly chosen values of parameters ϕ (i.e. first we start from the "flat" priori). This assumption reflects lack of knowledge about the function optimum parameters. For the actual state ϕ likelihood function λ is calculated. Then we apply random walk procedure "moving" our Markov chain to the new position. Precisely, we change each model ϕ parameter by the value draw from the Gaussian distribution with the zero mean and variance σ_ϕ^2 each parameter. Standard deviations for sampling parameters are determined by the problem's domain size and refined with a trial and error procedure to ensure that the Markov chains had access to realistic ranges with minimal occurrences of stuck problem. Problem of stuck in chains can occur when the standard deviations chosen for the next iteration lead to a large number of rejected samples, causing that the chain remains in a given position for many iterations. For the proposal state the likelihood function λ_{prop} is again estimated. We compare this two values λ and λ_{prop} according to (4). If comparison is more favorable than the previous chain location, the proposal is accepted (Markov chain "moves" to the new location). If the comparison is "worse", new state is not immediately rejected. Random variable from binomial distribution is used to decide whether or not to accept the new state of chain. After K iteration we pass all the samples (from all m chains)to the sequential procedure. We compute importance weights by (10) and normalize them. Next we use roulette procedure to draw N samples from the set generated by Markov Chain.

This random component is important because it prevents the chain from becoming trapped in a local minimum. The pseudo code for one time step of the algorithm is given below.

One of the important aspects of stochastic procedure of calculating the posterior distribution is choosing burn-in phase. The burn-in factor represents the number of samples needed at the beginning for the Markov chain to actually reach the search state where it is sampling from the target distribution.

Statistical convergence (to the posterior distribution) is monitored by computing between-chain variance and within-chain variance [4]. If there are m Markov chains of length N, then we can compute between-chain variance B with

$$B = \frac{N}{m-1}\sum_{j=1}^{m}(\bar{\phi_j} - \bar{\phi})^2 \tag{14}$$

where $\bar{\phi_j}$ is the average value along each Markov and $\bar{\phi}$ is the average of the values from all Markov chains. The within-chain variance W is

$$W = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \tag{15}$$

where

$$s_i^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_{ij} - \bar{\phi_i})^2 \tag{16}$$

The convergence parameter R is then computed as

$$R = \frac{var(\phi)}{W} \tag{17}$$

where $var(\phi)$ is estimate variance of ϕ and is computed as

$$var(\phi) = \frac{N-1}{N}W + \frac{1}{N}B. \tag{18}$$

In this paper, we consider the following variants of scanning algorithms:

1) **Classic MCMC**
   In this algorithm, the parameter space scan in each time step t is independent form the previous ones. So, in this case we don't use information from past calculations. Classic MCMC don't use sequential mechanism.

2) **SMC via Maximal Weights**
   As the first location of Markov chain ϕ_0^t it select the set of ϕ parameters for which weight in previous time step procedure was the highest. So, for $t > 1$:
   $$\phi_0^t \sim \ arg\ (\phi \in \{\phi_0^{t-1}, ..., \phi_n^{t-1}\})\ max\{w(\phi_i^{t-1})\} \tag{19}$$
   With this approach, we always start with the best values found so far.

3) **SMC via Rejuvenation and Extension**
   In contrast to SMC via Maximal Weights this algorithm as the first location of Markov chain ϕ_0^t at the time $t > 1$ chooses the set of parameters ϕ selected randomly from previous realization of resampling procedure in $t-1$ with use of the uniform distribution:
   $$\phi_0^t \sim U(\phi_0^{t-1}, \phi_1^{t-1}, ..., \phi_n^{t-1}) \tag{20}$$
   $$a\ uniform\ distribution\ \{1, ..., n\}$$
   Applying the new knowledge (new measurements) the current chain is "extended" starting from selected position with use of the new information in the likelihood function calculation.

III. Algorithms results for selected optimization problem

A. Two-dimensional (2D) Ackley function

We have benchmark the proposed global optimization algorithms with use of the 2D version of Ackley function Fig. 2.

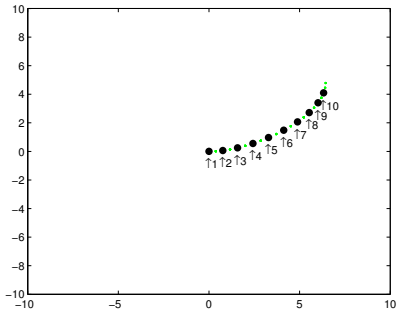


Fig. 1. Trajectory of the optimum of Ackley function

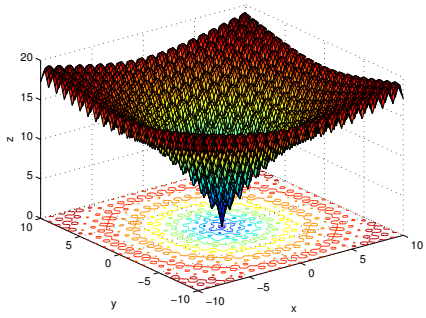


Fig. 2. Surf of two-dimensional Ackley benchmark function

$$H_1(x,y) = -20\exp(-0.2\sqrt{0.5(x^2+y^2)}) - \exp(0.5(cos(2\pi x)+\cos(2\pi y))) + 20 + e \quad (21)$$

The proposed optimization algorithms are designed to search for the optimum of the dynamical functions. To achieve the dynamical nature of the process described by the considered function we have ascribe the displacement of the optimum in 10 subsequent time steps. The assumed trajectory of the searched function optimum is presented in Fig. 1.

Based on the dynamical Ackley function we would like to compare the performance of two described in previous chapter SMC algorithms (i.e. SMC via Maximal Weights and SMC via Rejuvenation and Extension) in compare with a well-known stochastic simulation method i.e. classic MCMC. Since we are interested in runtime of all algorithms for optimization problems we use exactly the same parameters. The number of iteration for each algorithm is equal $K = 2000$. This number was chosen based on the numerical experiments as the number of iteration needed to reach convergence for each sampled dimension ($R \approx 1$) Fig. 3. The same way we tuned the rest of the algorithm parameters which adequately are equal: number of chains M=10; burn-in factor=500.

Fig. 5, and 6 presents the probability distributions of x and y optimum parameters in each time step for classic MCMC algorithm. Fig. 7 and 8 presents the same results for SMC via Maximal Weights and Figs. 10 and 11 for SMC via

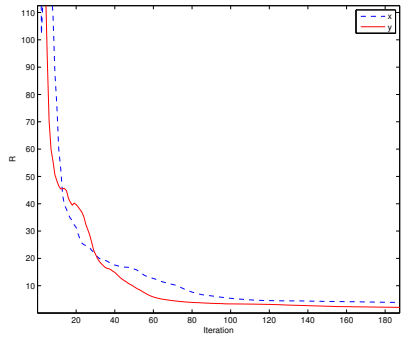


Fig. 3. R convergence parameter for x and y. The samples came from results of MCMC algorithm.

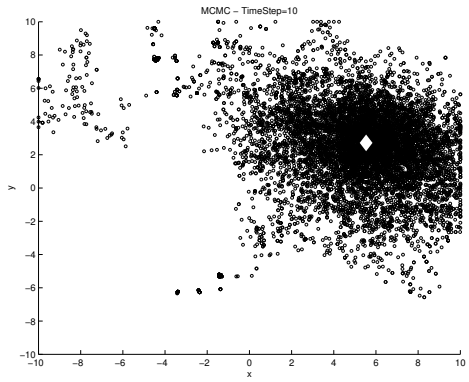


Fig. 4. The traces of three Markov chains in the x,y space. The global minimum is marked by diamond. The samples came from results of MCMC algorithm.

Rejuvenation and Extension, respectively. The target value of search optimum parameters are denoted by vertical lines. One can see that for the 2D Ackley dynamical function all algorithms successfully generate samples in the vicinity of the optimal solution. If we look carefully we can denote the difference of its probability values for the search parameters x and y. Moreover, the posteriori distributions of MCMC algorithm are much flatter than for SMC algorithms. For SMC via Maximal Weights and SMC via Rejuvenation and Extension the maximum value of the probability distribution x and y is close to 0.05 while for MCMC it is ≈ 0.027.

In both SMC algorithms transmission of the information in subsequent time steps about "fleeing minimum" effect enlarge the concentration of the samples around the target optimum. Resampling mechanism can be seen in Fig. 4 and Fig. 9. Fig. 4 presents the traces of the Markov chains for classic MCMC and Fig. 9 for SMC via Rejuvenation and Extension in the last time step. One can see that MCMC algorithm consider samples spread out far from the the searched optimum values, at the same time the SMC method in subsequent time steps choose samples close to the target value, which results in the increase of its probability.

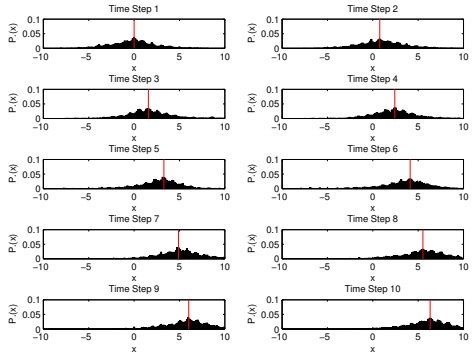


Fig. 5. Posterior distribution of x parameter in subsequent time steps for MCMC algorithm. Vertical line represents the target value of x.

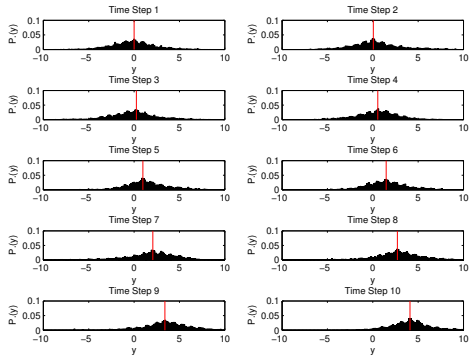


Fig. 6. Posterior distribution of y parameter in subsequent time steps for MCMC algorithm. Vertical line represents the target value of y.

B. Optimization problem - 4D Wood function

The previous example showed that the SMC algorithms give impute to the value of the probability of the searched optimum parameters (the probability is doubled). However, the classic MCMC also reached the target value of optimum. We would like to check if the proposed algorithms increase their efficiency in the case of the multidimensional space.

To test the effectiveness of SMC for optimization problem

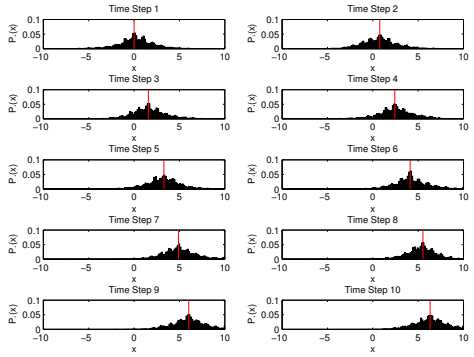


Fig. 7. Posterior distribution of x parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of x.

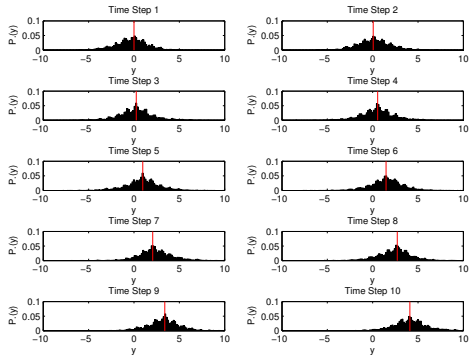


Fig. 8. Posterior distribution of y parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of y.

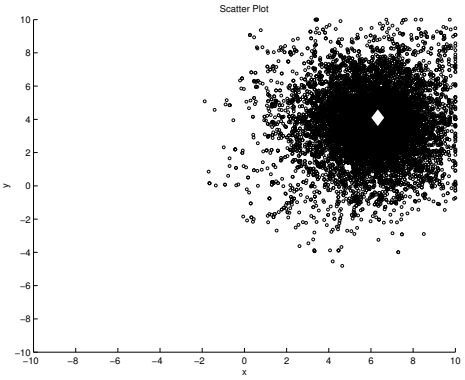


Fig. 9. Scatter plot of samples in the x,y space. The global minimum is marked by diamond. The samples came from results of SMC via Rejuvenation and Extension

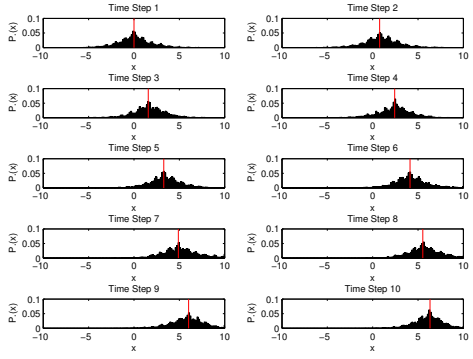


Fig. 10. Posterior distribution of x parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of x.

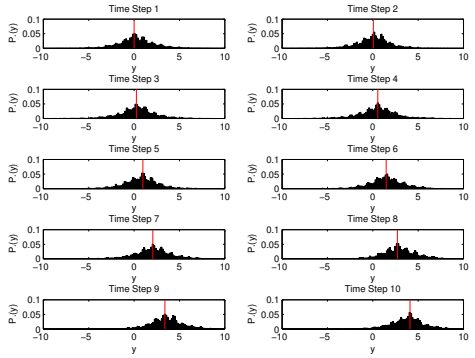


Fig. 11. Posterior distribution of y parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of y.

with higher dimensions, we consider the four-dimensional (4D) Wood function in our second benchmark test:

$$H_2(x_1,x_2,x_3,x_4) = 100({x_1}^2 - x_2)^2 + (x_1-1)^2 + (x_3-1)^2 + 90({x_3}^2 - x_4)^2 + 10.1((x_2-1)^2 + (x_4-1)^2) + 19.8(x_2-1)(x_4-1) \quad (22)$$

In this test we also assume that the optimum initial value $x = (1,1,1,1)$ moves in 6 subsequent time steps reaching at last $x = (3.84, 6.89, 3.84, 6.89)$. In this test for all considered algorithms we take: number of iteration $K = 20000$, number of chains $M = 10$; burn-in factor $= 2000$.

Figs. 12- 22 presents the marginal probability distributions for all four optimum parameters of the considered 4D Wood dynamical function. The target minimum location in each dimension is marked by the vertical red line. One can see that for the 4D Wood function efficiency of the classic MCMC is decreased. This method do not mark the target value of x_2 and x_4 parameters as the values with the highest probability (Figs. 13, 15). At the same time the results obtained from the two SMC algorithms are better than the MCMC algorithm. However, one can note than the SMC via Rejuvenation and Extension algorithm seems to be more efficient than SMC via Maximal Weights. The reason is that the SMC via Maximal Weights results for parameter x_2 (Fig. 17)are a bit worse than obtained from SMC via Rejuvenation and Extension (Fig. 21). Moreover, the SMC via Rejuvenation and Extension denotes the target values of the x_1 ,x_3 and x_4 parameter with higher probabilities than SMC via Maximal Weights.

It is worth to mention that SMC via Maximal Weights, SMC via Rejuvenation and Extension use the probability distributions obtained based on information from previous time steps to update the probability distributions with use of the new information. This causes a significant increase in convergence of the algorithm to the target location of the function's optimum in the subsequent time steps. This methodology makes these algorithms more effective for optimization of multidimensional dynamical functions than classic MCMC.

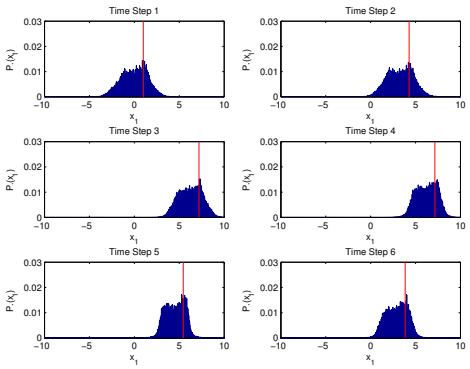


Fig. 12. Posterior distribution of x_1 parameter in subsequent time steps for MCMC. Vertical line represents the target value of x_1.

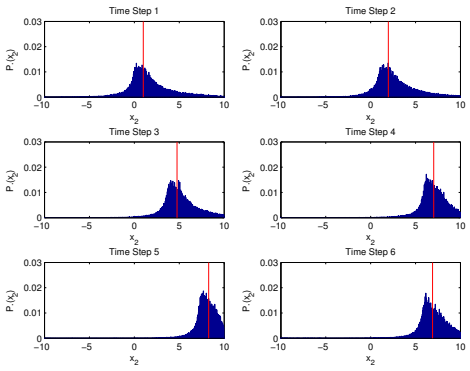


Fig. 13. Posterior distribution of x_2 parameter in subsequent time steps for MCMC. Vertical line represents the target value of x_2.

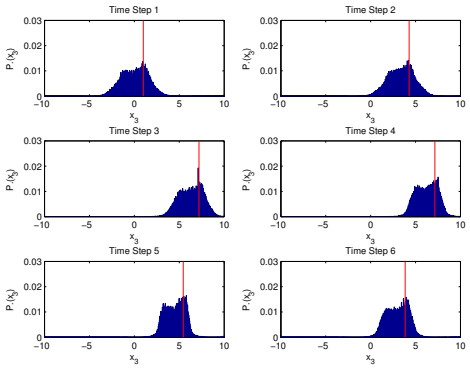


Fig. 14. Posterior distribution of x_3 parameter in subsequent time steps for MCMC. Vertical line represents the target value of x_3.

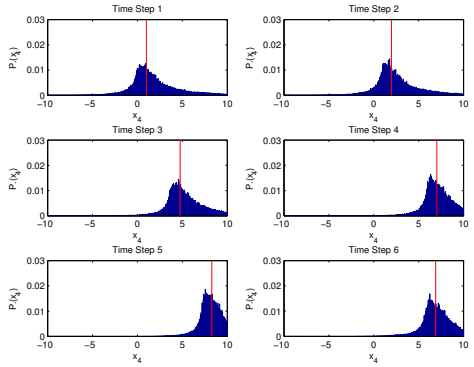


Fig. 15. Posterior distribution of x_4 parameter in subsequent time steps for MCMC. Vertical line represents the target value of x_4.

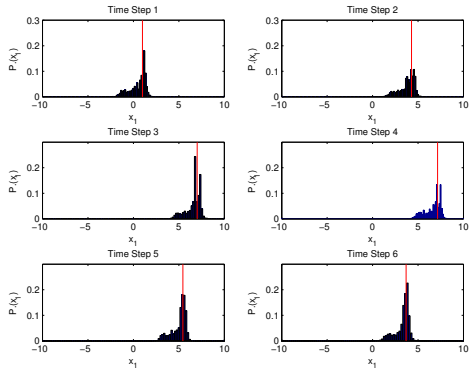


Fig. 16. Posterior distribution of x_1 parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of x_1.

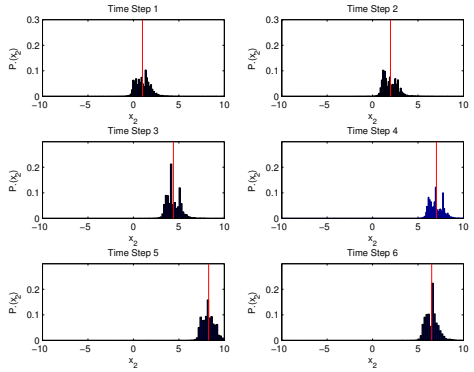


Fig. 17. Posterior distribution of x_2 parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of x_2.

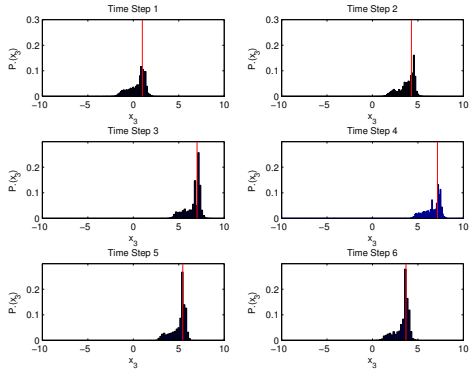


Fig. 18. Posterior distribution of x_3 parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of x_3.

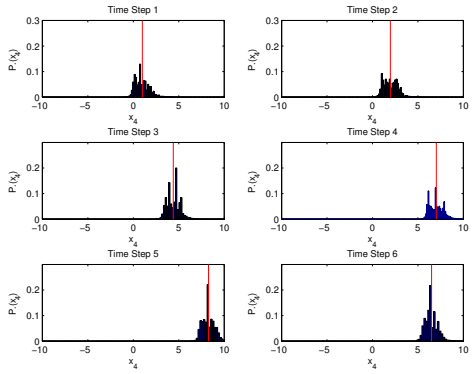


Fig. 19. Posterior distribution of x_4 parameter in subsequent time steps for SMC via Maximal Weights. Vertical line represents the target value of x_4.

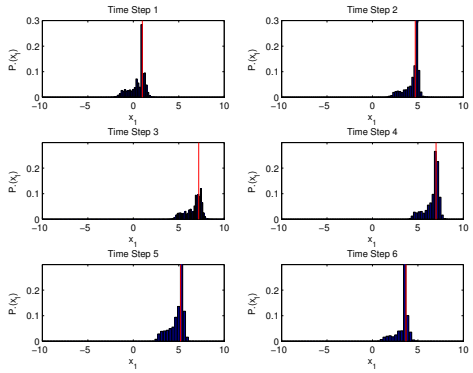


Fig. 20. Posterior distribution of x_1 parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of x_1.

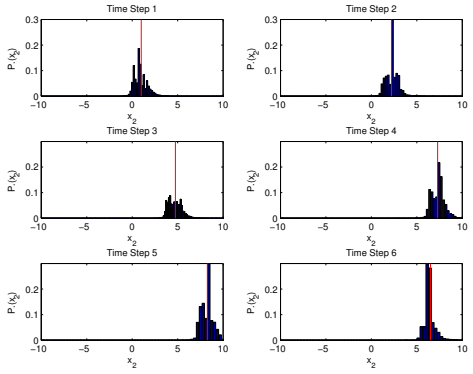


Fig. 21. Posterior distribution of x_2 parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of x_2.

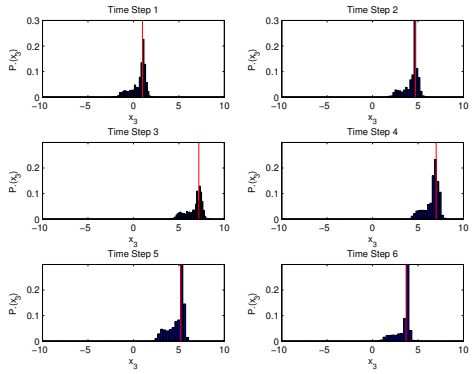


Fig. 22. Posterior distribution of x_3 parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of x_3.

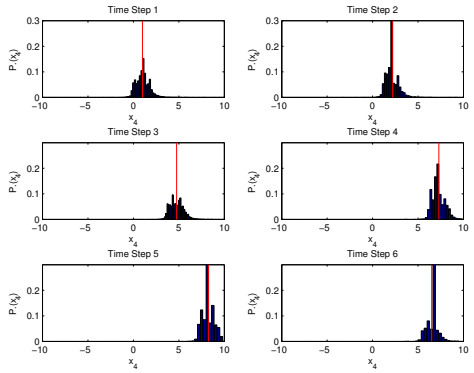


Fig. 23. Posterior distribution of x_4 parameter in subsequent time steps for SMC via Rejuvenation and Extension. Vertical line represents the target value of x_4.

IV. Conclusion

We have presented a methodology to solve the global optimization problem of dynamical functions with use of the Bayesian approach joined with SMC algorithms. The presented method combines Bayesian inference with SMC sampling and produces posterior probability distributions of the searches extremum's parameters. We have examined two version of the SMC algorithms i.e. SMC via Maximal Weights, SMC via Rejuvenation and Extension and compare its efficiency to estimate the probabilistic distributions of optimum parameters for 2D and 4D optimization functions. We compared the effectiveness of the proposed SMC algorithms with classic MCMC and have shown the advantage of the SMC algorithms that in different ways use the probability distributions of possible optimum parameters obtained basing on samples generated in previous time steps. We have shown that efficiency of proposed SMC algorithms increases with increasing the dimension of the optimized dynamical function. We conclude that proposed methodology joining the Bayesian inference with SMC algorithms is effective for optimization of multidimensional dynamical functions.

Acknowledgment

This work was supported by the Welcome Programme of the Foundation for Polish Science operated within the European Union Innovative Economy Operational Programme 2007-2013 and by the EU and MSHE grant nr POIG.02.03.00-00-013/09; Project VI.B.08 financed by the National Centre for Research and Development.

References

[1] Zuev, K.M., Beck, J.L., (2013): Global optimization using the asymptotically independent Markov sampling method. *Comput Struct*, http://dx.doi.org/10.1016/j.compstruc.2013.04.005

[2] Borysiewicz, M., Wawrzynczak A.,Kopka P., (2012): Stochastic algorithm for estimation of the model's unknown parameters via Bayesian inference. *Proceedings of the Federated Conference on Computer Science and Information Systems, 501–508, IEEE Press*,

[3] Borysiewicz, M., Wawrzynczak, A., Kopka, P., (2012): Bayesian-Based Methods for the Estimation of the Unknown Model's Parameters in the Case of the Localization of the Atmospheric Contamination Source, *Foundations of Computing and Decision Sciences*, 37, 4, 253–270

[4] Gelman, A., Carlin, J., Stern, H., Rubin, D., (2003): Bayesian Data Analysis. *Chapman & Hall/CRC*, 668 pp.

[5] Gilks, W., Richardson, S., Spiegelhalter ,D., (1996): Markov Chain Monte Carlo in Practice. *Chapman & Hall/CRC*, 486

[6] Doucet, A.,J. F. G.de Freitas, and Gordon, N. J., (2001): Sequential Monte Carlo methods in practice. *New York: Springer-Verlag*

[7] Pitt, M. K. Shephard, N. (2001): Sequential Monte Carlo methods in practice, chapter Auxiliary variable based particle filters. *New-York: Springer-Verlag*

[8] Liu, J. S. (2001): Monte Carlo Strategies in Scientific Computing. *New York: Springer*

[9] Gordon, N. J., Salmond, D. J., Smith, A. F. M., (1993): Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F on Radar and Signal Processing* 140 (2): 107–113,

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 371–376

Influence of the Population Size on the Genetic Algorithm Performance in Case of Cultivation Process Modelling

Olympia Roeva
Institute of Biophysics and
Biomedical Engineering,
Bulgarian Academy of Science,
Acad. G. Bonchev Str., bl. 105,
1113 Sofia, Bulgaria
E-mail: olympia@biomed.bas.bg

Stefka Fidanova
Institute of Information and
Communication Technology,
Bulgarian Academy of Science,
Acad. G. Bonchev Str., bl. 25A,
1113 Sofia, Bulgaria
E-mail: stefka@parallel.bas.bg

Marcin Paprzycki
Systems Research Institute,
Polish Academy of Sciences, Warsaw
and
Management Academy,
Warsaw, Poland
E-mail:marcin.paprzycki@ibspan.waw.pl

***Abstract*—In this paper, an investigation of the influence of the population size on the genetic algorithm (GA) performance for a model parameter identification problem, is considered. The mathematical model of an *E. coli* fed-batch cultivation process is studied. The three model parameters – maximum specific growth rate (μ_{max}), saturation constant (k_S) and yield coefficient ($Y_{S/X}$) are estimated using different population sizes. Population sizes between 5 and 200 chromosomes in the population are tested with constant number of generations. In order to obtain meaningful information about the influence of the population size a considerable number of independent runs of the GA are performed. The observed results show that the optimal population size is 100 chromosomes for 200 generations. In this case accurate model parameters values are obtained in reasonable computational time. Further increase of the population size, above 100 chromosomes, does not improve the solution accuracy. Moreover, the computational time is increased significantly.**

I. Introduction

METAHEURISTICS, such as genetic algorithms (GA), are widely used to solve various optimization problems. The GA are highly relevant for industrial applications, because they are capable of handling problems with non-linear constraints, multiple objectives, and dynamic components – properties that frequently appear in the real-world problems [15]. Since their introduction and subsequent popularization [16], the GA have been frequently used as an alternative optimization tool to the conventional methods and have been successfully applied in a variety of areas, and still find increasing acceptance [1], [3], [7], [11], [23], [28], [29].

The metaheuristic algorithms require of setting the values of several algorithm components and parameters. These parameters values have great impact on performance and efficacy of the algorithm [13], [22], [30], [14]. Therefore, it is important to investigate the algorithm parameters influence on the performance of the developed metaheuristic algorithms. The aim is to find the optimal parameters values for the considered optimization problem. The optimal values for the parameters depend mainly on i) the problem; ii) the instance of the problem to deal with and iii) the computational time that will be spend in solving the problem. Usually in the algorithm parameters tuning a compromise between solution quality and search time should be done.

For the parameter setting of metaheuristics, several automated approaches exist. These methods use i) a single step of parameter tuning (prior to the practical use of the algorithm), or parameter control (self adaptation to the problem being optimized) [19]. Parameter control is well suited when one wants good average performances across diverse problems, but the needed computation overhead leads to less efficiency on specific problems, compared to parameter tuning [9]. Best known parameter tuning techniques are racing [8], sequential parameter optimization [5] and meta-parameter setting (sometimes referred as meta-algorithm [5]).

Population sizing has been one of the important topics to consider in evolutionary computation [2], [12], [31]. Various results about the appropriate population size can be found in the literature[25], [27]. Researchers usually argue that a "small" population size could guide the algorithm to poor solutions [17], [24], [31] and that a "large" population size could make the algorithm expend more computation time in finding a solution [17], [20], [21]. Due to significant influence of population size to the solution quality and search time [27] a more thorough research should be done for this GA parameter.

The main goal of this research is to carry out investigation of the influence of one of the key GA parameters – population size (number of chromosomes) – on the algorithm performance for identification of a cultivation process model. Parameter identification of non-linear cultivation process models is a hard combinatorial optimization problem for which exact algorithms or traditional numerical methods do not work efficiently. A non-linear mathematical model of fed-batch cultivation process of the most important host organism for recombinant protein production — bacteria *Escherichia coli* – is considered [27].

The paper is organized as follows. The problem formulation is given in Section 2. The numerical results and a discussion

978-83-60810-53-8/$25.00

are presented in Section 3. Conclusion remarks are done in Section 4.

II. Problem Formulation

A. E. coli Fed-batch Cultivation Model

Application of the general state space dynamical model [6] to the *E. coli* cultivation fed-batch process leads to the following nonlinear differential equation system [27]:

$$\frac{dX}{dt} = \mu_{max}\frac{S}{k_S + S}X - \frac{F_{in}}{V}X \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu_{max}\frac{S}{k_S + S}X + \frac{F_{in}}{V}(S_{in} - S) \tag{2}$$

$$\frac{dV}{dt} = F_{in} \tag{3}$$

where X is the biomass concentration, [g/l]; S is the substrate concentration, [g/l]; F_{in} is the feeding rate, [l/h]; V is the bioreactor volume, [l]; S_{in} is the substrate concentration in the feeding solution, [g/l]; μ_{max} is the maximum value of the specific growth rate, [h^{-1}]; k_S is the saturation constant, [g/l]; $Y_{S/X}$ is the yield coefficient, [-].

The initial process conditions are [4]:

- t_0 = 6.68 h,
- $X(t_0)$ = 1.25 g/l and $S(t_0)$ = 0.8 g/l,
- S_{in} = 100 g/l.

For the considered non-linear mathematical model of *E. coli* fed-batch cultivation process the parameters that should be identified are:

- maximum specific growth rate (μ_{max}),
- saturation constant (k_S),
- yield coefficient ($Y_{S/X}$).

B. Genetic Algorithm

GA was developed to model adaptation processes mainly operating on binary strings and using a recombination operator with mutation as a background operator. The GA maintains a population of chromosomes, $P(t) = x_1^t$, ..., x_n^t for generation t. Each chromosome represents a potential solution to the problem and is implemented as some data structure S. Each solution is evaluated to give some measure of its "fitness". Fitness of a chromosome is assigned proportionally to the value of the objective function of the chromosomes. Then, a new population (generation $t+1$) is formed by selecting more fit chromosomes (selection step). Some members of the new population undergo transformations by means of "genetic" operators to form new solution. There are unary transformations m_i (mutation type), which create new chromosomes by a small change in a single chromosome ($m_i : S \to S$), and higher order transformations c_j (crossover type), which create new chromosomes by combining parts from several chromosomes ($c_j : S \times \ldots \times S \to S$). After some number of generations the algorithm converges – it is expected that the best chromosome represents a near-optimum (reasonable) solution. The combined effect of selection, crossover and mutation gives so-called reproductive scheme growth equation [15]:

$$\xi(S, t+1) \geq$$

$$\xi(S,t) \cdot eval(S,t) / \bar{F}(t) \left[1 - p_c \cdot \frac{\delta(S)}{m-1} - o(S) \cdot p_m\right]$$

The structure of the herewith used GA is shown by the pseudocode below (Figure 1).

begin
  $i = 0$
  Initial population $P(0)$
  Evaluate $P(0)$
  while (not done) **do**
  (test for termination criterion)
  begin
    $i = i + 1$
    Select $P(i)$ from $P(i-1)$
    Recombine $P(i)$
    Mutate $P(i)$
    Evaluate $P(i)$
  end
end

Fig. 1. Pseudocode for GA

Three model parameters are represented in the chromosome – μ_{max}, k_S and $Y_{S/X}$. The following upper and lower bounds of the model parameters are considered [29]:

$$0 < \mu_{max} < 0.7,$$

$$0 < k_S < 1,$$

$$0 < Y_{S/X} < 30.$$

Roulette wheel, developed by Holland [16] is the herewith used selection method. The probability, P_i, for each chromosome is defined by:

$$P[\text{Individual } i \text{ is chosen}] = \frac{F_i}{\sum_{j=1}^{PopSize} F_j}, \tag{4}$$

where F_i equals the fitness of chromosome i and $PopSize$ is the population size.

To reproduce the chromosomes simple crossover and binary mutation according to [29] are applied. In proposed genetic algorithm fitness-based reinsertion (selection of offspring) is used.

For the considered here model parameter identification, the type of the basic operators in GA are as follows [29]:

- encoding – binary,
- fitness function – linear ranking,
- selection function – roulette wheel selection,
- crossover function – simple crossover,
- mutation function – binary mutation,
- reinsertion – fitness-based.

The values of GA parameters are [29]:

- generation gap, ggap = 0.97,
- crossover probability, xovr = 0.75,
- mutation probability, mutr = 0.01,
- maximum number of generations, maxgen = 200.

C. Optimization Criterion

In practical view, modelling studies are performed to identify simple and easy-to-use models that are suitable to support the engineering tasks of process optimization and, especially of control. The most appropriate model must satisfy the following conditions:

(i) the model structure should be able to represent the measured data in a proper manner;

(ii) the model structure should be as simple as possible compatible with the first requirement.

The optimization criterion is a certain factor, whose value defines the quality of an estimated set of parameters. To evaluate the mishmash between experimental and model predicted data the Least Square Regression is used.

The objective consists of adjusting the parameters (μ_{max}, k_S and $Y_{S/X}$) of the non-linear mathematical model function (Eq. (1) - Eq. (3)) to best fit a data set. A simple data set consists of n points (data pairs) (x_i, y_i), $i = 1, 2, \ldots, n$, where x_i is an independent variable and y_i is a dependent variable whose value is found by observation. The model function has the form $f(x, \beta)$, where the m adjustable parameters are held in the vector β, $\beta = [\mu_{max}\ k_S\ Y_{S/X}]$. The goal is to find the parameter values for the model which "best" fits the data. The least squares method finds its optimum when the sum S of squared residuals:

$$S = \sum_{i=1}^{n} r_i^2$$

is a minimum. A residual is defined as the difference between the actual value of the dependent variable and the value predicted by the model. A data point may consist of more than one independent variable. For an example, when fitting a plane to a set of height measurements, the plane is a function of two independent variables, x and z, say. In the most general case there may be one or more independent variables and one or more dependent variables at each data point.

$$r_i = y_i - f(x_i, \beta).$$

III. Numerical Results and Discussion

All computations are performed using a PC/Intel Core i5-2320 CPU @ 3.00GHz, 8 GB Memory (RAM), Windows 7 (64 bit) operating system and Matlab 7.5 environment.

A series of numerical experiments are performed to evaluate the influence of the population size in GAs on the accuracy of the obtained solution. Using mathematical model of the *E. coli* cultivation process (Eq. (1) - Eq. (3)) the model parameters – maximum specific growth rate (μ_{max}), saturation constant (k_S) and yield coefficient ($Y_{S/X}$) – are estimated. For the identification procedures consistently different population sizes (from 5 to 200 chromosomes in the population) are used. The number of generations is fixed to 200. Because of the stochastic characteristics of the applied GA series of 30 runs for each population size are performed.

TABLE I
ALGORITHM PERFORMANCE FOR VARIOUS POPULATION SIZES - OBJECTIVE FUNCTION VALUES

Population size	Objective function S		
	Average	Best	Worst
5	6.1200	4.8325	9.2958
10	5.8000	4.8548	9.6175
20	4.7660	4.4753	5.3634
30	4.6519	4.4816	5.0094
40	4.6359	4.4437	4.9669
50	4.6070	4.4488	4.8636
60	4.5886	4.4625	4.8013
70	4.5648	4.4384	4.7357
80	4.5782	4.4474	4.7463
90	4.5711	4.4496	4.7211
100	4.5406	4.4252	4.7017
110	4.5455	4.4332	4.7319
150	4.5511	4.4575	4.6717
200	4.5453	4.4359	4.7206

TABLE II
ALGORITHM PERFORMANCE FOR VARIOUS POPULATION SIZES - COMPUTATIONAL TIME

Population size	Computational time, s		
	Average	Best	Worst
5	4.9457	4.5552	5.6004
10	6.0039	5.6316	6.3648
20	7.6482	7.3008	7.9561
30	11.1115	10.8265	11.5129
40	12.9824	12.4957	13.3537
50	14.9087	14.3989	15.5377
60	17.2766	16.6141	20.3113
70	19.7601	19.1725	20.0617
80	22.1880	21.7153	22.6669
90	24.3414	23.9150	24.8198
100	26.8644	26.4890	27.8306
110	29.7057	29.1878	30.2642
150	39.7273	39.1407	40.3887
200	52.4782	51.3087	55.8952

In the Table I, obtained average, best and worst objective function values for considered population sizes, are presented. The results observed for computational time are listed in Table II.

The numerical experiments show that increasing the size of the population of 5 to 100 chromosomes significantly improves the resulting value of the objective function (average results) – from 6.1200 to 4.5406 (see Table I). The further increase in the size of population (more than 100 chromosomes) does not lead to more accurate results. The subsequent increase in the population size leads only to an increase in computational time without improving the value of the objective function (average results) – from 26.8644 s (100 chromosomes) to 52.4782 s (200 chromosomes) vs. $S = 4.5406$ to $S = 4.5453$ (see Table II).

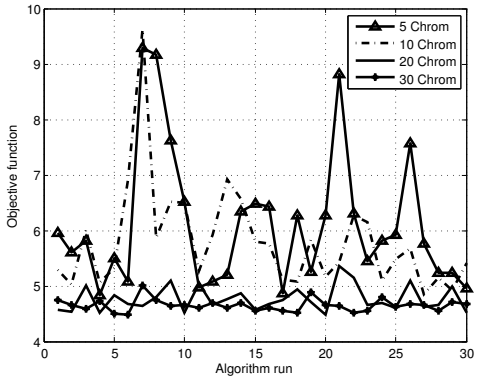


Fig. 2. Objective function values obtained during the 30 algorithm runs for 5, 10, 20 and 30 chromosomes in the population

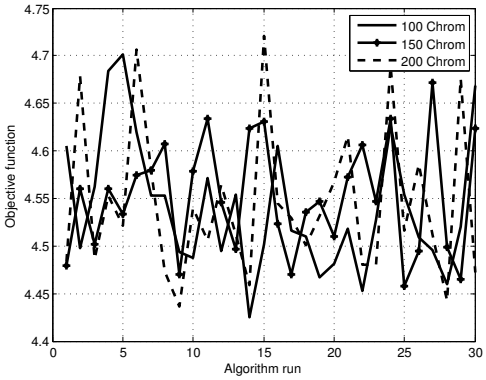


Fig. 3. Objective function values obtained during the 30 algorithm runs for 100, 150 and 200 chromosomes in the population

For better interpretation the obtained numerical results are graphically visualized in the next figures. On Figure 2 the objective function values, obtained during the 30 GA runs for 5, 10, 20 and 30 chromosomes in the population, are shown. The graphical results show that the GA could not find accurate solution using small population size – 5 or 10 chromosomes. It is need at least 20 chromosomes in population for achieving a better solution. On Figure 3 the objective function values, obtained during the 30 algorithm runs for 100, 110, 150 and 200 chromosomes in the population, are shown. Here, it could be seen that using large population size (110, 150 or 200 chromosomes) did not result in an improvement of the objective function values. The ANOVA test is applied and the values of the objective function for population size equal and more than 100 are statistically equal. Moreover, as can be seen from Figure 5 increasing the population size result in an acceleration of computational time. When the population size increases it leads to increase of the needed computational resources like time and memory which can be a problem for large-scale tests. Therefore we can conclude that populations with 100 individuals is optimal with respect to the value of the objective function and the needed computational resources.

All numerical experiments for the influence of the population size on the objective function value and on the computational time are summarized in Figure 4 and Figure 5. It can be concluded that for the considered here non-linear cultivation model parameter identification problem the optimal population size is 100 chromosomes in the population (for 200 generations).

In the Table III the best parameter values (μ_{max}, k_S and $Y_{S/X}$), obtained using GA with 100 chromosomes in the population, are presented. According to [10], [18], [32] the values of the estimated model parameters are in admissible boundaries.

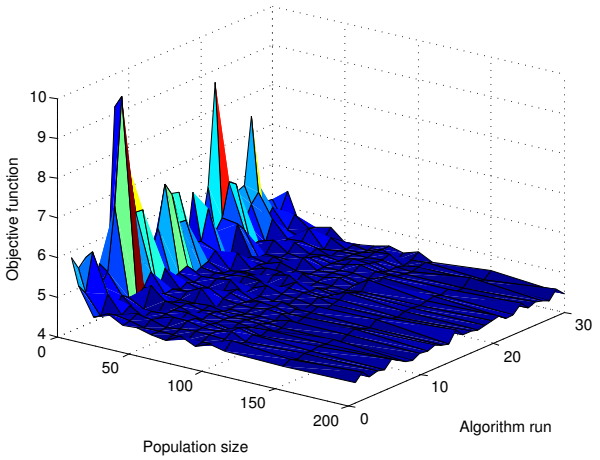


Fig. 4. Influence of the population size on the objective function value

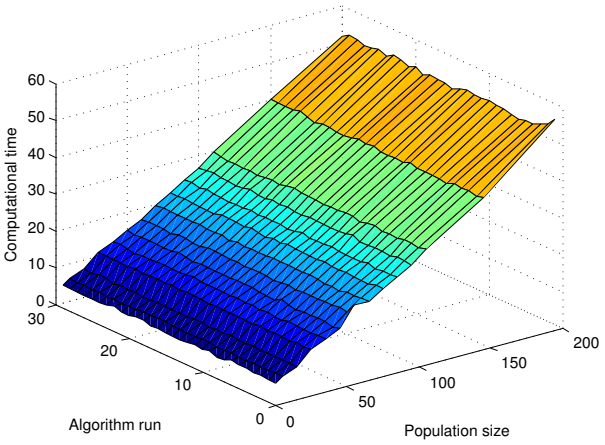


Fig. 5. Influence of the population size on the computational time

IV. Conclusion

A good selection of the GA parameters improve both computation time and solution accuracy. Finding good parameter values is not a trivial task and requires human expertise as

TABLE III
BEST PARAMETER VALUES OF THE MODEL (100 CHROMOSOMES)

Parameter	Value
μ_{max}, [1/h]	0.4881
k_S, [g/l]	0.0120
$Y_{S/X}$, [-]	2.0193

well as time. In this paper, the influence of the one of key GA parameters (population size) on the GA performance, is studied. As a test problem, the *E. coli* fed-batch cultivation model parameter identification, is considered. The three model parameters (maximum specific growth rate (μ_{max}), saturation constant (k_S) and yield coefficient ($Y_{S/X}$)) are identified. For a fixed number of the generations (200) different population sizes of the GA are explored. The numerical experiments are started with 5 chromosomes in the population and consistently increased to 200 chromosomes. The obtained results show that the optimal population size, for the considered here case study, is 100 chromosomes. Thus, accurate model parameters values are obtained with reasonable computational efforts. The use of smaller populations result in lower accuracy of the solution, obtained for a smaller computational time. The further increase of the population size increases the accuracy of solution. This effect is observed to a population size of 100 chromosomes. The use of larger populations does not improve the solution accuracy and only increase the needed computational resources.

ACKNOWLEDGMENT

This work has been partially supported by the Bulgarian National Scientific Fund under the Grants DID 02/29 "Modeling Processes with Fixed Development Rules (ModProFix)" and DMU 02/4 "High quality control of biotechnological processes with application of modified conventional and metaheuristics methods". Work presented here is a part of the Poland-Bulgarian collaborative Grant "Parallel and distributed computing practices" and by European Commission project ACOMIN.

REFERENCES

[1] S. Akpinar and G. M. Bayhan, "A Hybrid Genetic Aalgorithm for Mixed Model Assembly Line Balancing Problem with Parallel Workstations and Zoning Constraints", Engineering Applications of Artificial Intelligence, Vol. 24, No. 3, 2011, pp. 449-457.

[2] J. T. Alander, "On optimal population size of genetic algorithms", In Proceedings of the IEEE Computer Systems and Software Engineering, 1992, pp. 65–69.

[3] H. N. Al-Duwaish, "A Genetic Approach to the Identification of Linear Dynamical Systems with Static Nonlinearities", International Journal of Systems Science, Vol. 31, No. 3, 2000, pp. 307-313.

[4] M. Arndt and B. Hitzmann, "Feed Forward/feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*", 8th IFAC Int. Conf. on Comp. Appl. in Biotechn, Canada, 2001, pp. 425-429.

[5] T. Bartz-Beielstein, "Experimental Research in Evolutionary Computation: The New Experimentalism", Natural Computing Series, Springer, 2006.

[6] G. Bastin and D. Dochain, "On-line Estimation and Adaptive Control of Bioreactors", Els. Sc. Publ, 1991.

[7] K. K. Benjamin, A. N. Ammanuel, A. David and Y. K. Benjamin, "Genetic Algorithm using for a Batch Fermentation Process Identification", J of Applied Sciences, Vol. 8, No. 12, 2008, pp. 2272-2278.

[8] M. Birattari, T. Stützle, L. Paquete and K. Varrentrapp, "A racing algorithm for configuring metaheuristics", In GECCO 02: Proceedings of the Genetic and Evolutionary Computation Conference, 2002, pp. 11–18, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.

[9] J. Clune, S. Goings, B. Punch and E. Goodman, "Investigations in metagas: panaceas or pipe dreams?", In GECCO 05: Proceedings of the 2005 workshops on Genetic and evolutionary computation, 2005, pp. 235–241, New York, NY, USA, ACM.

[10] J. Contiero, C. Beatty, S. Kumari, C. L. DeSanti, W. L. Strohl, A. Wolfe, "Effects of mutations in acetate metabolism on high-cell-density growth of *Escherichia coli*", Journal of Industrial Microbiology and Biotechnology, Vol. 24, 2000, pp. 421-430.

[11] M. F. J. da Silva, J. M. S. Perez, J. A. G. Pulido and M. A. V. Rodriguez, "AlineaGA - A Genetic Algorithm with Local Search Optimization for Multiple Sequence Alignment", Appl Intell, Vol. 32, 2010, pp. 164-172.

[12] P. A. Diaz-Gomez and D. F. Hougen, "Initial Population for Genetic Algorithms: A Metric Approacs", Proceedings of the 2007 International Conference on Genetic and Evolutionary Methods, GEM 2007, June 25-28, 2007, Las Vegas, Nevada, USA, Hamid R. Arabnia and Jack Y. Yang and Mary Qu Yang (Eds), pp. 43-49, CSREA Press.

[13] Á. E. Eiben, R. Hinterding and Z. Michalewicz, "Parameter Control in Evolutionary Algorithms", IEEE Transactions on Evolutionary Computation, Vol. 3, No. 2, 1999.

[14] Fidanova S., "Simulated Annealing: A Monte Carlo Method for GPS Surveying", Computational Science - 2006, Lecture Notes in Computer Science No 3991, 2006, pp. 1009-1012.

[15] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley Longman, London, 2006.

[16] J. H. Holland, "Adaptation in Natural and Artificial Systems", 2nd Edn. Cambridge, MIT Press, 1992.

[17] V. K. Koumousis and C. P. Katsaras, A sawtooth genetic algorithm combining the effects of variable population size and reinitialization to enhance performance, IEEE Transactions on Evolutionary Computation, Vol. 10, No. 1, 2006, pp. 19–28.

[18] D. Levisauskas, V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk, A. Lubbert, "Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*", Bioprocess and Biosystems Engineering, Vol. 25, 2003, pp. 255-262.

[19] F. G. Lobo, C. F. Lima and Z. Michalewicz, (Ed.), "Parameter Setting in Evolutionary Algorithms", Studies in Computational Intelligence, Vol. 54, 2007.

[20] F. G. Lobo and D. E. Goldberg, "The parameterless genetic algorithm in practice, Information SciencesInformatics and Computer Science, Vol. 167, No. 1-4, 2004, pp. 217–232.

[21] F. G. Lobo and C. F. Lima, A review of adaptive population sizing schemes in genetic algorithms, In Proceedings of the Genetic and Evolutionary Computation Conference, 2005, pp. 228–234.

[22] R. Nowotniak and J. Kucharski, "GPU-based Tuning of Quantum-Inspired Genetic Algorithm for a Combinatorial Optimization Problem", In Proceedings of the XIV International Conference System Modeling and Control, 2011, ISSN 978-83-927875-1-8.

[23] J. P. Paplinski, "The Genetic Algorithm with Simplex Crossover for Identification of Time Delays", *Intelligent Information Systems*, 2010, pp. 337-346.

[24] M. Pelikan, D. E. Goldberg, and E. Cantu-Paz, "Bayesian optimization algorithm, population sizing, and time to convergence", Illinois Genetic Algorithms Laboratory, University of Illinois, Tech. Rep., 2000.

[25] C. R. Reeves, "Using Genetic Algorithms With Small Populations", In Proceedings of the Fifth International Conference on Genetic Algorithms, 1993, pp. 92-99.

[26] E. Ridge, "Design of Experiments for the Tuning of Optimisation Algorithms", PhD Thesis, The University of York, Department of Computer Science, 2007.

[27] O. Roeva, "Improvement of Genetic Algorithm Performance for Identification of Cultivation Process Models", Advanced Topics on Evolutionary Computing, Book Series: Artificial Intelligence Series-WSEAS, 2008, pp. 34-39.

[28] O. Roeva and Ts. Slavov, "Fed-batch Cultivation Control based on Genetic Algorithm PID Controller Tuning", Lecture Notes on Computer Science, Springer-Verlag Berlin Heidelberg, Vol. 6046, 2011, pp. 289-296.

[29] O. Roeva and S. Fidanova, "Chapter 13. A Comparison of Genetic Algorithms and Ant Colony Optimization for Modeling of *E. coli* Cultivation Process", In book Real-World Application of Genetic Algorithms, In Tech, 2012, pp. 261-282.

[30] A. Saremi, T. Y. ElMekkawy and G. G. Wang, "Tuning the Parameters of a Memetic Algorithm to Solve Vehicle Routing Problem with Backhauls Using Design of Experiments", International Journal of Operations Research, Vol. 4, No. 4, 2007, pp. 206-219.

[31] A. Piszcz and T. Soule, "Genetic programming: Optimal population sizes for varying complexity problems", In Proceedings of the Genetic and Evolutionary Computation Conference, 2006, pp. 953–954.

[32] B. Zelic, D. Vasic-Racki, C. Wandrey, R. Takors, "Modeling of the pyruvate production with *Escherichia coli* in a fed-batch bioreactor", Bioprocess and Biosystems Engineering, Vol. 26, 2004, pp. 249-258.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 377–384

Quadratic TSP: A lower bounding procedure and a column generation approach

Borzou Rostami, Federico Malucelli
Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
Email: rostami@elet.polimi.it, malucell@elet.polimi.it

Pietro Belotti
Department of Mathematical Sciences, Clemson University, Clemson, USA

Stefano Gualandi
Dipartimento di Matematica, Università degli Studi di Pavia, Pavia, Italy

***Abstract*—In this paper we present a Column Generation approach to the Quadratic Traveling Salesman Problem. Given a graph and a function that maps every pair of edges to a cost, the problem consists in finding a cycle that visits every vertex exactly once and such that the sum of the costs over all pairs of consecutive edges of the cycle is minimum. We propose a Linear Programming formulation that has a variable for each cycle in the graph. Since the number of cycles is exponential in the graph size, we solve our formulation via column generation. Computational results on some set of instances used in the literature show that our approach is promising. As it obtains a lower bound close to the optimal solutions for all instances.**

I. Introduction

THE TRAVELING SALESMAN PROBLEM (TSP) is one of the most studied optimization problems. Given an undirected graph $G = (V, E)$ with costs $c_e, e \in E$, the problem consists in finding a cycle C that visits each vertex in V exactly once, and such that the sum of the costs c_e of each edge in C is minimum. In its most common form, the TSP has a linear cost function.

In this paper we study a variant of the TSP that has a quadratic cost function, the so-called Quadratic TSP (QTSP). The input of this problem is an undirected (or directed) graph $G = (V, E)$ and a cost function $r : E \times E \to \Re^+$ that maps every pair of edges (or arcs) to a non-negative cost. The QTSP consists in finding a cycle C of minimum cost that visits every vertex of G exactly once. This problem is NP-hard [4]. We distinguish between Asymmetric QTSP and Symmetric QTSP, depending on the fact that the direction of the cycle matters.

The QTSP was introduced with an application to bioinformatics [4] and also has application in robotics and telecommunications. The QTSP can be viewed as a generalization of the Reload Cost TSP (RTSP) introduced in [1]. In the RTSP, one is given a graph whose every edge is assigned a color and there is a *reload* cost when passing through a node on two edges that have different colors.

The QTSP has been tackled in [4] with heuristic algorithms based on well-known heuristics for the TSP, with an ad-hoc branch-and-bound solver, and with a branch-and-cut approach based on a linearization of a 0-1 Quadratic Programming formulation of the problem. In [2] and [3], a polyhedral study is used to develop a branch-and-cut algorithm for symmetric and asymmetric QTSP respectively that are the current state-of-the-art for QTSP.

In this paper we present a Linear Programming formulation of the QTSP with an exponential number of variables that is solved via Column Generation (CG). The basic idea is to have a variable for each cycle of G. This yields a pricing subproblem that consists in finding a cycle of minimum quadratic cost. We formulated the pricing subproblem as a 0-1 Quadratic Program, which is linearized and solved with standard techniques. We resort to stabilization techniques to overcome the tailing-off effect of CG approach.

In Section II we present mathematical formulations and some linearized models for the symmetric QTSP and the asymmetric QTSP. In Section III we present our Linear Programming formulation of the problem. In Section IV we solve the LP model presented in Section III by a column-generation technique and present different formulations of the pricing subproblems for both the symmetric and the asymmetric cases. Computational results are discussed in Section V.

II. The Quadratic Traveling Salesman Problem

In this section we present mathematical formulations for both symmetric QTSP (SQTSP) and asymmetric QTSP (AQTSP). We denote the edge incident with vertices i and j as $\{i, j\}$ in the symmetric case while in the asymmetric case the arc from vertex i to vertex j is denoted as (i, j).

A. The Symmetric QTSP and a Linearized formulation

Consider a complete undirected graph G on a vertex set $V = \{1, 2, \ldots, n\}$ and let $r : E \times E \to \Re^+$ be a cost function that maps a pair of edges to a non-negative cost. The cost of a subgraph in G is equal to the sum of the costs of the pairs of edges incident in the same vertex. The SQTSP seeks a tour (i.e., a cycle passing through each vertex exactly once) of minimum cost. We define the binary variable x_{ij} that is equal to 1 if edge $\{i, j\}$ belongs to the minimum cost tour,

978-83-60810-53-8/$25.00

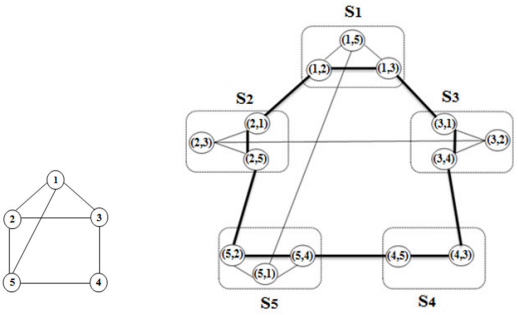


Fig. 1. Graph G and its corresponding *Gadget* graph $\hat{G}$. Note that we assume G to be complete, but this example shows the connections in the gadget graph when G is not complete, or when some of its edges have infinite cost

and 0 otherwise. The SQTSP can be modeled as a integer linear programming (ILP) problem as follows:

$$\min \quad \sum_{\{i,j\}\in E} \sum_{\{j,k\}\in E} r_{ijk} x_{ij} x_{jk} \tag{1}$$

$$\text{s.t.} \quad \sum_{j:\{i,j\}\in E} x_{ij} = 2 \quad \forall i \in V \tag{2}$$

$$\sum_{\substack{\{i,j\}\in E:\\ i\in S, j\in S}} x_{ij} \leq |S| - 1 \quad \emptyset \neq S \subset V \tag{3}$$

$$x_{ij} \in \{0,1\} \quad \forall \{i,j\} \in E. \tag{4}$$

The objective function is straightforward and considers all costs between edges incident on the same vertices. Constraints (2) ensure that each vertex has degree two in the tour, and constraints (3) ensure that no subtour is formed among the subsets of vertices.

A simple linear relaxation of (1)–(4) can be obtained by replacing the term $x_{ij}x_{jk}$ with a binary variable u_{ijk} subject to the following constraints.

$$u_{ijk} \leq x_{ij}, \quad u_{ijk} \leq x_{jk}, \quad \text{and} \quad u_{ijk} \geq x_{ij} + x_{jk} - 1.$$

Motivated by the desire to develop a linearized formulation, we reformulate the SQTSP by creating a new graph $\hat{G} = (\hat{V}, \hat{E})$, called the Gadget graph, as follows:

- For each edge in the graph G, create two nodes $\langle i,j\rangle$ and $\langle j,i\rangle$ and add them to $\hat{V}$.
- For each node $i \in V$ in G, define a super node $S_i = \{\langle i,j\rangle : j = 1,2,\ldots,n, j \neq i\}$. This is only an aggregation of nodes of $\hat{V}$.
- For any $i,j,k \in V$, create an edge between each two nodes $\langle i,j\rangle$ and $\langle i,k\rangle$ in the super node S_i and assign the weight r_{jik} to it.
- For any $i,j \in V$, create an edge between each two nodes $\langle i,j\rangle$ and $\langle j,i\rangle$ in super node S_i and S_j respectively and assign null weight to it.

In Figure 1 we present an example of a graph G and the corresponding *Gadget* graph $\hat{G}$.

By introducing the decision variables u_{jik} to indicate whether edge $\{\langle i,j\rangle, \langle i,k\rangle\}$ in super node S_i is selected or not in the optimal solution, we can rewrite the SQTSP on graph G as the following integer linear problem on the graph $\hat{G}$:

$$\min \quad \sum_{i\in V} \sum_{\substack{j,k\in V:\\ \{\langle i,j\rangle,\langle i,k\rangle\}\in S_i}} r_{jik} u_{jik} \tag{5}$$

$$\text{s.t.} \quad \sum_{\substack{j,k\in V:\\ \{\langle i,j\rangle,\langle i,k\rangle\}\in S_i}} u_{jik} = 1 \quad \forall i \in V \tag{6}$$

$$\sum_{\substack{\langle i,k\rangle\in S_i:\\ k\neq j}} u_{jik} - \sum_{\substack{\langle j,k\rangle\in S_j:\\ k\neq i}} u_{ijk} = 0 \quad \forall \langle i,j\rangle \in \hat{V} \tag{7}$$

$$\sum_{i\in S} \sum_{\substack{\{\langle i,j\rangle,\langle i,k\rangle\}\in \hat{E}:\\ j,k\in S}} u_{jik} \leq |S| - 1 \quad S \subset V \tag{8}$$

$$u_{jik} \in \{0,1\} \quad \forall \{\langle i,j\rangle, \langle i,k\rangle\} \in \hat{E}. \tag{9}$$

Constraints (6) guarantee that within every super node we select exactly one edge, constraints (7) indicate that the number of the edges incident on node $\langle i,j\rangle$ is equal to the number of incident edges on node $\langle j,i\rangle$, and constraints (8) are the subtour elimination constraints. Note that in terms of feasible solutions in the *Gadget graph*, a tour is called *feasible* if it is simple and passes through each super node S_i by visiting exactly one edge of S_i. In the example shown in Figure 1, the bold lines illustrate a feasible tour (a subtour of $\hat{G}$) corresponding to the feasible tour $\{(1,2),(2,5),(5,4),(4,3),(3,1)\}$ in the original graph G.

An alternative model in the case of reload costs spanning tree was studied in [10], where it is assumed that the triangle inequality holds for reload costs at each node of the graph.

B. The Asymmetric QTSP and Linearized formulation

Suppose that the given graph G is a complete directed graph on the vertex set $V = \{1,2,\ldots,n\}$, and let $r : E \times E \to \Re^+$ be a cost function that maps every pair of arcs to a non negative integer cost. The Asymmetric QTSP (AQTSP) can be modeled as follows:

$$\min \quad \sum_{(i,j)\in E} \sum_{(j,k)\in E} r_{ijk} x_{ij} x_{jk} \tag{10}$$

$$\text{s.t.} \quad \sum_{(i,j)\in E} x_{ij} = 1 \quad \forall i \in V \tag{11}$$

$$\sum_{(i,j)\in E} x_{ij} = 1 \quad \forall j \in V \tag{12}$$

$$\sum_{i\in S, j\notin S:(i,j)\in E} x_{ij} \geq 1 \quad S \subset V \tag{13}$$

$$x_{ij} \in \{0,1\} \quad \forall (i,j) \in E, \tag{14}$$

where the binary variable x_{ij} is equal to 1 if the arc (i,j) belongs to the minimum cost tour. Constraints (11) and (12) force to select a single outgoing arc and a single incoming arc for each node, respectively, and constraints (13) are the well known subtour elimination constraints.

In order to linearize the model, we follow the idea of constructing an auxiliary graph as in the symmetric case. We

call this auxiliary graph the *extended graph* and denote it as $\overline{G}$. For each arc (i,j) in the graph G we create a node $\langle i,j\rangle$ in $\overline{G}$, a link between each two nodes $\langle i,j\rangle$ and $\langle j,k\rangle$ in $\overline{G}$ and assign a weight r_{ijk} to each edge $(\langle i,j\rangle,\langle j,k\rangle)$ in $\overline{G}$. If in $\overline{G}$ we partition the set of nodes into n clusters $\mathcal{V}_1,\mathcal{V}_2,\dots\mathcal{V}_n$ such that $\mathcal{V}_i=\{\langle i,j\rangle : j=1,2,\dots,n, j\neq i\}$ for $i=1,2,\dots,n$, then all arcs are defined between nodes $\langle i,j\rangle$ and $\langle j,k\rangle$ from different clusters such that $r_{ijk}>0$; therefore there are no intra-set arcs.

Proposition 2.1: Any feasible tour in G corresponds to a tour in $\overline{G}$ that goes through each cluster exactly once.

Figure 2 represents the extended graph, $\overline{K}_4$, of a directed complete graph, K_4. The bold lines illustrate a feasible tour.

Definition 2.1: Given a directed graph $G=(V,E)$ and a partition $\{\mathcal{V}_i : i=1,\dots,k\}$ of the set V such that $\bigcap_{i=1}^{k}\mathcal{V}_i=\emptyset$ and $\bigcup_{i=1}^{k}\mathcal{V}_i=V$, the *Asymmetric Generalized* TSP (AGTSP) can be stated as the problem of finding a feasible cycle $T\subset E$ which includes exactly one node from each cluster $\mathcal{V}_i$, $i=1,\dots,k$, and whose global cost $\sum_{e\in T}r_e$ is minimum. Therefore the AGTSP involves two related decisions: (i) choosing a node subset $S\subset V$ such that $|S\cap\mathcal{V}_i|=1$ for all $i=1,2,\dots,n$ and (ii) finding a minimum cost Hamiltonian cycle in the subgraph of G induced by S.

Corollary 2.2: Solving the AQTSP on graph G is equivalent to solving the AGTSP on $\overline{G}$.

Using Corollary 2.2, instead of solving the original AQTSP one can solve an AGTSP on graph $\overline{G}$ which is again NP-hard, as it can be reduced to an Asymmetric TSP [8], [9].

We can formulate an integer linear programming model for the AGTSP. Variables u_{ijk} indicate whether arc $(\langle i,j\rangle,\langle j,k\rangle)$ is selected or not in the optimal solution, and variables y_{ij} indicate whether node $\langle i,j\rangle$ is visited or not. The problem is displayed as follows:

GTSP1:

$$\min \sum_{(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}} r_{ijk}u_{ijk} \tag{15}$$

$$\text{s.t.} \sum_{\substack{j\in V:\\ \langle i,j\rangle\in\mathcal{V}_i}} y_{ij}=1 \quad \forall i\in V \tag{16}$$

$$\sum_{\substack{k\in V:\\ \langle j,k\rangle\in\overline{V}}} u_{ijk}=y_{ij} \quad \forall\langle i,j\rangle\in\overline{V} \tag{17}$$

$$\sum_{\substack{k\in V:\\ \langle k,i\rangle\in\overline{V}}} u_{kij}=y_{ij} \quad \forall\langle i,j\rangle\in\overline{V} \tag{18}$$

$$\sum_{i\in S}\sum_{\substack{j\notin S:\\ \langle i,j\rangle\in\mathcal{V}_i}}\sum_{\substack{k\in V:\\ \langle j,k\rangle\in\mathcal{V}_j}} u_{ijk}\geq 1 \quad S\subset V \tag{19}$$

$$u_{ijk}\in\{0,1\} \quad \forall(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E} \tag{20}$$

$$y_{ij}\in\{0,1\} \quad \forall\langle i,j\rangle\in\overline{V}. \tag{21}$$

Constraint (16) guarantees that from every cluster we select exactly one node. Constraints (17) and (18) require a solution to include exactly one of the arcs entering and exactly one of the arcs leaving the node $\langle i,j\rangle$ if this node is visited. Finally, constraint (19) eliminates all subtours.

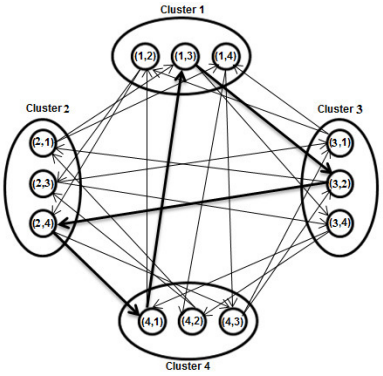


Fig. 2. Extended graph $\overline{K}_4$ of the complete graph K_4

Note that relaxing the integrality requirement on the variables u_{ijk} for all $(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}$ does not have any effect on problem GTSP1.

By eliminating variables y_{ij} we can write the problem as follows:

GTSP2:

$$\min \sum_{(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}} r_{ijk}u_{ijk} \tag{22}$$

$$\text{s.t.} \sum_{j\in V}\sum_{\substack{k\in V:\\ (\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}}} u_{ijk}=1 \quad \forall i\in V \tag{23}$$

$$\sum_{k\in V}\sum_{\substack{j\in V:\\ (\langle k,i\rangle,\langle i,j\rangle)\in\overline{E}}} u_{kij}=1 \quad \forall i\in V \tag{24}$$

$$\sum_{\substack{k\in V:\\ \langle j,k\rangle\in\overline{V}}} u_{ijk}-\sum_{\substack{k\in V:\\ \langle k,i\rangle\in\overline{V}}} u_{kij}=0 \quad \forall\langle i,j\rangle\in\overline{V} \tag{25}$$

$$(19),(20). \tag{26}$$

Constraint (23) requires a solution to include exactly one of the arcs entering a cluster, while constraint (24) requires a solution to include exactly one of the arcs leaving a cluster. Constraint (25) is equivalent to network flow conservation constraints and ensure that a solution tour is uninterrupted and continuous.

Theorem 2.3: Constraints (23) in problem GTSP2 are redundant and can be removed from the model.

Proof: Suppose u^* is a feasible solution for problem GTSP2 after removing constraint (23) and define, for each $i\in V$,

$$\epsilon_i=\sum_{j\in V}\sum_{\substack{k\in V:\\ (\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}}} u^*_{ijk}.$$

We show that $\epsilon_i=1$ for all $i\in V$.

Consider cluster $\mathcal{V}_s$. Then constraint (24) is satisfied for

cluster $\mathcal{V}_s$ by u^*, i.e.,

$$\sum_{k\in V} \sum_{\substack{j\in V \\ (\langle k,s\rangle,\langle s,j\rangle)\in\overline{E}}} u^*_{ksj} = 1.$$

Therefore there exists a node $\langle s,t\rangle \in \mathcal{V}_s$ with in-degree one in the tour, while the in-degree of all other nodes in cluster $\mathcal{V}_s$ is zero. Hence by (25) the out-degree of node $\langle s,t\rangle$ must be equal to one while the out-degree of all the other nodes in the same cluster must be zero. ■

As a consequence of Theorem 2.3 GTSP2 simplifies as follows:

$$\text{GTSP3:} \quad \min \sum_{(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}} r_{ijk}u_{ijk} \tag{27}$$

$$\text{s.t.} \quad (24), (25), (26). \tag{28}$$

III. Cycle Reformulation of QTSP

In this section we present a new formulation of the QTSP which is based on a cycle generation approach in the given graph. Let C be a cycle of G represented by the set of edges (arcs) that appear in the cycle. The cost of cycle C is $r(C) = \sum_{i,j,k\in V:(i,j),(j,k)\in C} r_{ijk}$, i.e. the sum of the costs of the pairs of consecutive edges (arcs) contained in the cycle. Let $\mathcal{C}$ and $\mathcal{T}$ denote the collection of all cycles and all tours of G, respectively. Clearly, we have that $\mathcal{T} \subseteq \mathcal{C}$, and hence $\min_{C\in\mathcal{C}} c(C) \leq \min_{T\in\mathcal{T}} c(T)$.

Following the approach of Held and Karp to the TSP [6], we add a penalty π_i to every vertex i in V, and denote by $\pi(C) = \sum_{i\in C} \pi_i$. Let us consider a new cost function defined as follows: $d(C) = c(C) + \pi(C)$. Let T^* denote an optimal tour of G, i.e. $c(T^*) = \min_{T\in\mathcal{T}} c(T)$. Then, the following relations hold:

$$\min_{C\in\mathcal{C}} d(C) = \min_{C\in\mathcal{C}}\{c(C) + \sum_{i\in C}\pi_i\} \leq d(T^*) = c(T^*) + \sum_{i\in V}\pi_i$$

$$\min_{C\in\mathcal{C}}\{c(C) - \sum_{i\in V\setminus C}\pi_i\} \leq c(T^*).$$

For any vector of penalty terms π we get a lower bound. However, we are interested in finding π that maximizes the lower bound:

$$\max_{\pi\in\Re^n} \min_{C\in\mathcal{C}} \{c(C) - \sum_{i\in V\setminus C}\pi_i\}. \tag{29}$$

We describe an LP equivalent to (29) by introducing a variable z as follows:

$$\text{P: max} \quad z \tag{30}$$

$$\text{s.t.} \quad z + \sum_{i\in V\setminus C}\pi_i \leq c(C) \qquad \forall C\in\mathcal{C} \tag{31}$$

$$z, \pi \text{ unrestricted.} \tag{32}$$

Let λ_C be the dual multiplier of constraint (31). The dual of problem P is called master problem and has the following form:

$$\text{D1: min} \quad \sum_{C\in\mathcal{C}} c(C)\lambda_C \tag{33}$$

$$\text{s.t.} \quad \sum_{C\in\mathcal{C}:i\notin C} \lambda_C = 0 \qquad \forall i\in V \tag{34}$$

$$\sum_{C\in\mathcal{C}} \lambda_C = 1 \tag{35}$$

$$\lambda_C \geq 0 \qquad \forall C\in\mathcal{C}. \tag{36}$$

Since all multipliers in (34) are non-negative, and, in each iteration of a column generation approach, exactly one column is generated (as we explain in Section IV), an optimal solution to the problem must satisfy $\lambda^*_C = 1$ for some $C^* \in \mathcal{C}$ and $\lambda^*_C = 0$ for all $C \in (\mathcal{C} \setminus C^*)$. It follows that the cycle C^* (single or multiple) is optimal and $c(C^*)$ provides a lower bound for the original QTSP.

By subtracting each constraint (34) from (35) and removing (35) from D1, one can find a relaxation of the problem D1 as follows:

$$\text{D2: min} \quad \sum_{C\in\mathcal{C}} c(C)\lambda_C \tag{37}$$

$$\text{s.t.} \quad \sum_{C\in\mathcal{C}:i\in C} \lambda_C = 1 \qquad \forall i\in V \tag{38}$$

$$\lambda_C \geq 0 \qquad \forall C\in\mathcal{C}. \tag{39}$$

Problem D2 seeks a minimum-weight "combination of cycles" such that each vertex appears, on average, in one cycle.

IV. Column Generation Approach

In this section we develop a column generation approach to solve problems D1 and D2. Since the number of cycles in $\mathcal{C}$ is exponential with respect to the number of vertices, we first consider a restricted version of the master problem (RMP) with a feasible subset of cycles, $\bar{\mathcal{C}} \subseteq \mathcal{C}$. Note that the subset $\bar{\mathcal{C}}$ for problem D1 must include at least one tour, while an initial set $\bar{\mathcal{C}}$ for problem D2 must satisfy constraints (38). Let us first start with problem D1 and suppose that $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ and z are the dual variables corresponding to constraints (34) and (35) respectively. The reduced cost of the variable λ_C for each $C \in \mathcal{C}$ is $\overline{r(C)} = r(C) - (\pi(C))' - z$, where $(\pi(C))' = \sum_{i\notin C}\pi_i$. A column to enter the basis can be found by computing a minimum cost cycle with respect to $r_{ijk} + \pi_j$ for each $(i,j) \in E, (j,k) \in E$. Let $\overline{r(C^p)} = \min_{C\in\mathcal{C}} \overline{r(C)}$. If $\overline{r(C^p)} \geq 0$ then the current solution is optimal. Otherwise we select column C^p to enter the basis.

Theorem 4.1: If the column C^p to enter the basis corresponds to a tour, then it is an optimal tour.

Proof: Consider any cycle C. Since column C^p is the selected column to enter the basis we have

$$\overline{r(C^p)} = r(C^p) - (\pi(C^p))' - z \leq r(C) - (\pi(C))' - z. \tag{40}$$

If cycle C is also a tour, then $r(C^p) \leq r(C)$. ■

Note that for problem D2, the reduced cost of the variable λ_C for each $C \in \mathcal{C}$ is modified to:

$$\overline{r(C)} = r(C) - \pi(C). \tag{41}$$

A. Pricing subproblems

A column to enter the basis can be found by computing a minimum cost cycle in the original graph G with respect to $r_{ijk} + \pi_j$. Looking for a cycle having minimum negative cost with respect to a quadratic objective is itself an interesting combinatorial optimization problem, which is NP-hard [5]. In this section we explain how to update the linearized models, presented in Section II, to solve the pricing problems. In order to define a suitable model for the pricing subproblem of the restricted master problem D1, we consider the SQTSP and AQTSP separately.

a) Symmetric case: Consider the pricing problem of D1 in the symmetric case. As we mention in Section II-A, instead of looking for a cycle in the original graph one can easily find a cycle in the *Gadget* graph; i.e., finding a negative reduced cost of problem D1 is the same as finding a negative cost feasible cycle in the Gadget graph. Defining a binary variable w_i to indicate whether the super node S_i is on the cycle or not, the minimum negative cost cycle is then found by solving the following problem:

$$\min \quad \sum_{i\in V} \sum_{\{\langle i,j\rangle,\langle i,k\rangle\}\in S_i} r_{jik}u_{jik} - \sum_{i\in V}\pi_i(1-w_i) - z$$

$$\text{s.t.} \quad \sum_{\substack{j,k\in V:\\ \{\langle i,j\rangle,\langle i,k\rangle\}\in S_i}} u_{jik} = w_i \quad \forall i \in V \tag{42}$$

$$(7)-(9) \tag{43}$$

$$w_i \in \{0,1\} \quad \forall i \in V. \tag{44}$$

Considering the definition of the w variables, one can obtain an equivalent formulation for the pricing problem by replacing constraint (42) with the so-called resource constraint $\sum_{\substack{j,k\in V\\ \{\langle i,j\rangle,\langle i,k\rangle\}\in S_i}} u_{jik} \le 1$ and removing the variables w from the model. Also, it should be observed that by forcing the solution of the pricing problem to be a cycle with negative cost one can easily remove the constraint (8) from the pricing model. Therefore, finding a cycle with negative reduced cost is simply a matter of finding a path between each node of the *Gadget* graph and itself with a negative cost which satisfies the resource constraint. Since the dual variables π are defined on each super node of the *Gadget* graph, the problem of finding a negative reduced cost cycle can be formulated as resource-constrained elementary shortest path problem. Let q_{st} denote the cost of the resource-constrained elementary shortest path from origin node $\langle s,t\rangle$ to itself in the *Gadget* graph. The pricing problem can be written as: $\min_{\langle s,t\rangle\in\hat{V}}\{q_{st}\}$, where q_{st} is the optimal value of the following problem.

$$\min \quad \sum_{i\in V} \sum_{\{\langle i,j\rangle,\langle i,k\rangle\}\in S_i} (r_{jik}+\pi_i)u_{jik} - \sum_{i\in V}\pi_i - z \tag{45}$$

$$\text{s.t.} \quad \sum_{\substack{j\in V:\\ \{\langle s,t\rangle,\langle s,j\rangle\}\in S_i}} u_{tsj} = 1 \tag{46}$$

$$\sum_{\substack{j\in V:\\ \{\langle s,j\rangle,\langle s,t\rangle\}\in S_i}} u_{jst} = 1 \tag{47}$$

$$\sum_{\substack{\langle i,k\rangle\in S_i:\\ k\neq j}} u_{jik} - \sum_{\substack{\langle j,k\rangle\in S_j:\\ k\neq i}} u_{ijk} = 0 \quad \forall \langle i,j\rangle \neq \langle s,t\rangle \tag{48}$$

$$\sum_{\substack{j,k\in V\\ \{\langle i,j\rangle,\langle i,k\rangle\}\in S_i}} u_{jik} \le 1 \quad \forall i \in V \tag{49}$$

$$u_{jik} \in \{0,1\} \quad \forall\{\langle i,j\rangle,\langle i,k\rangle\} \in \hat{E}. \tag{50}$$

Constraints (46), (47), (48) and (50) find a path from the source node $\langle s,t\rangle$ to itself. The resource constraint (49) guarantee that each super node S_i is visited at most once.

b) Asymmetric case: In the asymmetric case, finding a negative reduced cost directed cycle for problem D1 is equivalent to solving the modified version of the GTSP1 or GTSP3 explained in Section II-B. Here we provide a version of the latter so that the resulting problem gives the most negative directed cycle.

$$\min \quad \sum_{(\langle i,j\rangle,\langle j,k\rangle)\in\overline{E}} r_{ijk}u_{ijk} - \sum_{i\in V}\pi_i(1-t_i) - z$$

$$\text{s.t.} \quad \sum_{k\in V} \sum_{\substack{j\in V\\ (\langle k,i\rangle,\langle i,j\rangle)\in\overline{E}}} u_{kij} = t_i \quad \forall i \in V$$

$$(20),(25),(24)$$

$$t_i \in \{0,1\} \quad \forall i \in V.$$

The binary variable t_i is equal to 1 if cluster i is visited, otherwise it is zero. Following the same process as the symmetric case, one can obtain an equivalent formulation based on resource-constrained elementary shortest path problem in the extended graph.

The solution of the subproblems provides either a certificate of optimality of the current solutions (λ,π,z) or a new column C^p that will be added to the master problem. It is worth pointing out that solving the subproblem to optimality is only needed to prove optimality of the current primal and dual solutions; one can stop solving the subproblem whenever a negative reduced cost column is found [16]. This happens because adding this column to $\bar{C}$ ensures that the new dual solution (π,z) will be different, and therefore the termination of the algorithm.

B. Stabilized column generation

Column generation methods usually suffer from slow convergence to the optimal solution. Primal degeneracy, dual degeneracy and instability in the behavior of dual variables are well known to cause slow convergence and *tailing off* effects to column generation procedures [11], [12].

To control the dual variables during the solution process, we use the stabilized column generation approach proposed in [15]. This approach combines the box step method [14] with a

kind of descent method proposed in [13]. The box step method introduces a box around the previous dual vector and modifies the master problem such that the feasible dual space is limited to the area defined by these boxes, while the latter tries to adapt the master problem so that the distance separating a dual solution from the previous optimal dual solution is linearly penalized.

In order to present the idea, let us rewrite the restricted version of the master problem D1, for $\bar{\mathcal{C}} \in \mathcal{C}$, as the following model:

$$\text{RD1: min} \quad \sum_{C \in \bar{\mathcal{C}}} c(C)\lambda_C \tag{51}$$

$$\text{s.t.} \quad \sum_{C \in \bar{\mathcal{C}}: i \in C} \lambda_C \geq 1 \qquad \forall i \in V \tag{52}$$

$$(35), (36).$$

Note that since the set partitioning constraints allow negative dual values which can be problematic for the sub-problem, we used a relaxed version of the problem as the first step of the stabilization approach.

Consider the dual variables π associated with the constraints (52) and bound each π_i in the interval $[\delta_i^-, \delta_i^+]$. These bounds are first given as parameters to the model and then automatically updated during the process. The dual variable π_i can take values outside the given bounds, but the dual objective is then penalized by $\varepsilon_i^-(\delta_i^- - \pi_i)$ if $\pi_i < \delta_i^-$ and by $\varepsilon_i^+(\delta_i^+ - \pi_i)$ if $\pi_i > \delta_i^+$. The dual of the problem RD1 then becomes:

$$\text{SP: max} \quad z + \sum_{i \in V} \pi_i - \varepsilon_i^- w_i^- - \varepsilon_i^+ w_i^+ \tag{53}$$

$$\text{s.t.} \quad z + \sum_{i \in C} \pi_i \leq c(C) \qquad \forall C \in \bar{\mathcal{C}} \tag{54}$$

$$\pi_i + w_i^- \geq \delta_i^- \qquad \forall i \in V \tag{55}$$

$$\pi_i - w_i^+ \leq \delta_i^+ \qquad \forall i \in V \tag{56}$$

$$\pi, w^-, w^+ \geq 0, \ z \text{ unrestricted.} \tag{57}$$

The primal of the stabilized restricted master problem, and hence the dual of SP, is:

$$\text{SD1: min} \quad \sum_{C \in \bar{\mathcal{C}}} c(C)\lambda_C + \sum_{i \in V} -\delta_i^- \mu_i^- + \delta_i^+ \mu_i^+ \tag{58}$$

$$\text{s.t.} \quad \sum_{C \in \bar{\mathcal{C}}: i \in C} \lambda_C - \mu_i^- + \mu_i^+ \geq 1 \quad \forall i \in V \tag{59}$$

$$\sum_{C \in \bar{\mathcal{C}}} \lambda_C = 1 \tag{60}$$

$$\mu_i^- \leq \varepsilon_i^- \quad \forall i \in V \tag{61}$$

$$\mu_i^+ \leq \varepsilon_i^+ \quad \forall i \in V \tag{62}$$

$$\lambda, \mu^-, \mu^+ \geq 0. \tag{63}$$

This method is referred to as BoxPen stabilization since the bounds (δ^-, δ^+) on the original dual variables π can be represented by a bounding box containing the current dual solution. Note that the stabilized version of the problem D2 is the same as SD1 without the convexity constraint (60) and is called SD2. In order to use the stabilized models efficiently, one must initialize and update the parameters correctly. With desire to reduce the dual variables' variations, select $[\delta^-, \delta^+]$ to form a small box containing the current dual solution, and solve the problem SD1 (SD2). At the first iteration, when no solution is available to the problem, the dual variables π can be simply estimated. If the new π lies in the box $[\delta^-, \delta^+]$, reduce its width and augment the penalty given by ε^- and ε^+. Otherwise, enlarge the box and decrease the penalty. The update could be performed in each iteration, or alternatively, each time a dual solution of currently best value is obtained. In Section V we discuss more about the updating process.

V. Computational Experiments

In this section we present our computational experiments on a class of randomly generated instances with size rating from $n = 5$ to $n = 25$ introduced in [4]. All instances consist of complete graphs with n vertices and $m = n(n-1)/2$ edges. The cost function $r : E \times E \to \Re_0^+$ for both symmetric and asymmetric instances is defined as an integer number which is chosen from the set $\{0, 1, ..., 10000\}$ uniformly for all $(i, j), (j, k) \in E$ such that $i \neq k$ and set to infinity for all $(i, j), (j, i) \in E$. We used the AMPL modeling language [17] with GUROBI 5.0.0 [18] as linear solver for the RMP and as a mixed integer linear solver for the pricing problem on an Intel Core i5-2410M CPU with 2.30 GHz and 6 GB RAM in single processor mode.

A. Stabilization

We implemented the stabilized column generation approach using different sets of initial values. In the following we present the results of some preliminary experiments whose purpose was to initialize and update the parameters for both SD1 and SD2.

For the problem SD1, we initialized δ^- and δ^+ at -1000 and 1000 respectively. The vector parameter ε^- and ε^+ were selected as -5 and 5 respectively and were kept fixed throughout the solution process. We updated the parameter (δ^-, δ^+) from $(-1000, 1000)$ to $(\tilde{\pi} - 100, \tilde{\pi} + 100)$, ($\tilde{\pi}$ is the current dual solution), only if the column returned by the subproblem had a non-negative reduced cost and $(\mu^-, \mu^+) \neq (0, 0)$. The stopping criteria of the stabilized column generation algorithm is $\overline{r(C)} \geq 0$ and $(\mu^-, \mu^+) = (0, 0)$.

In order to find potentially good initial values of (δ^-, δ^+) for problem SD2, we first solved the problem D2 with a feasible subset of cycles. By using the dual variables $\tilde{\pi}$ of problem D2, we initialized (δ^-, δ^+) with $(\tilde{\pi} - 10, \tilde{\pi} + 10)$. The vector parameter ε^- and ε^+ were initially set to 0.0001. If the subproblem is able to find negative reduced cost cycle, then the values of ε^- and ε^+ were increased by 10%. However, when there was no more such column and $(\mu^-, \mu^+) \neq (0, 0)$, the values of ε^- and ε^+ were decreased by dividing each one by 100 and the parameter (δ^-, δ^+) were updated to $(\tilde{\pi} - 100, \tilde{\pi} + 100)$, where $\tilde{\pi}$ is the current dual solution of the SD2. The stopping criteria of the stabilized column generation algorithm is the same as case of problem SD1.

TABLE I
COMPUTATIONAL TIME OF COLUMN GENERATION (CG) AND STABILIZED CG APPROACHES FOR BOTH THE SQTSP THE AQTSP INSTANCES

	CPU time (Symmetric)				CPU time (Asymmetric)			
size	CG2	SCG2	CG1	SCG1	CG2	SCG2	CG1	SCG1
10	2.15	1.91	4.36	4.12	3.91	3.26	8.93	7.35
20	43.57	33.34	205.94	60.70	97.43	34.92	508.53	271.81

In order to show the effectiveness of stabilization, we compare the computational time of column generation to its stabilized version on two instances of different dimensions in Table I. Each row of the table reports the average computational time over ten instances of the same size. CG1 and CG2 stand for Column Generation approach to the problem SD1 and SD2 respectively. SCG1 and SCG2 are for the stabilized version of the CG1 and CG2 respectively. The results show that stabilization is effective for both symmetric and asymmetric instances.

B. Lower bound Computation

As we mentioned in Section III, a solution of the pricing problem, which may contain a single or multiple cycles, is optimal for the master problem $RD1$ if it covers all the nodes $i \in V$. Since looking for a single cycle in the pricing problem requires some kind of subtour elimination constraint, we restrict the search to find a cycle (single or multiple) with negative cost. In other words, we allow subtours in the optimal solution of the original QTSP, which is in fact a relaxation of the problem. Therefore, when no more new columns can be priced out, a solution of the master problem RD1 gives a lower bound on the original problem. Note that the optimal value of the problem RD2 always gives a lower bound on the original problem, regardless of the solution being a single cycle or multiple ones.

In Table II we present computational results of the lower bounding schemes for both the symmetric and the asymmetric QTSPs. Each row of the tables reports the average results over ten instances of the same size. The problem size is found in the first column of the table. The second column shows the average optimal values (opt) of the 10 instances for each dimension. We compare three different average lower bounds (LB) on the optimal objective values, their computing time (time), number of iterations (iter), and the average gap. The first lower bound LB(LP) in column three is the lower bound obtained with the linear relaxation of problem (5) – (9) and the linear relaxation of problem GTSP3 for the SQTSP and the AQTSP respectively. The computation time of the LP relaxation is less than two seconds for all instances; therefore we did not mention it in the table. The second lower bound is obtained via the Stabilized Column Generation approach to the problem SD2 (SCG2); and the third lower bound is obtained via the Stabilized Column Generation approach apply to the problem SD1 (SCG1). Columns four to seven and columns eight to eleven represent the optimal value, computation time, number of iterations needed to identify the optimal solution and the gap. The formula we used to compute these gaps is $(opt - LB(SCG))/(opt - LB(LP))$, where opt and $LB()$ stand for the optimal value and the lower bound, respectively. We can see in this table that the bounds obtained by SCG1 outperforms the other two in all instances and are close to the optimal values in both the SQTSP and the AQTSP. Also the lower bond obtained by SCG2 is indeed better than the one obtained with LP relaxation except for the instances of dimension five, for which the LP relaxation gives on average a tighter bound. On average the ratio of gap between the lower bound obtained by SCG2 and the optimal solution, and the gap between the lower bound obtained by linear relaxation and the optimal solution for the symmetric instances is 0.72, while this ratio for the asymmetric instances is 0.73. As we see, on average, the improvement of lower bound by applying the SCG2 for both symmetric and the asymmetric cases is almost the same, while the computational time for the asymmetric instances is more than the computational time for the symmetric ones.

According to Table II, applying SCG1 yields a considerable improvement of the lower bounds in both the symmetric and the asymmetric cases, i.e., on average the ratio of gap between the lower bound obtained by SCG1 and the optimal solution over the gap between the lower bound obtained by linear relaxation and the optimal solution for the symmetric instances is 0.09, and for the asymmetric instances is 0.20. These gaps show the improvement of lower bounds in both the symmetric and asymmetric cases in compare to the SCG2. Also it should be noted that the improvement of the lower bounds and also the computational time in the symmetric case is more attractive than in the asymmetric case.

VI. CONCLUSIONS

In this paper we first proposed two different linearization models to the SQTSP and the AQTSP. We also presented a different cycle formulation for the QTSP (in general) and solved the resulting LP problem by Column Generation approach. We have shown how the linearized formulations can be applied to finding the negative reduced cost in the pricing problem. To overcome the problems of instability in the behavior of dual variables of the presented master problem, we used a stabilized column generation approach. Our experiments show that our column generation approach outperforms the LP relaxation of the QTSP in both the symmetric and the asymmetric cases. The main goal of this paper is to show the advantage of column generation and the weakness of the LP relaxation in finding a good lower bound for the QTSP.

TABLE II
COMPARISON OF THREE DIFFERENT LOWER BONDING APPROACHES

			$SCG2$				$SCG1$			
$size$	$Opt.$	$LB(LP)$	LB	$time$	$iter$	Gap	LB	$time$	$iter$	Gap
Symmetric:										
5	14922.2	13839.5	13606.3	0.62	6	1.21	14922.2	3.89	65	0.0
6	12217.5	11817.1	11816.2	0.53	7	0.99	12160.4	3.35	43	0.14
7	14648.6	13200.4	13747.9	0.85	11	0.62	14356.5	3.51	41	0.20
8	15055.7	12544.1	12623.9	1.12	14	0.96	14542.6	4.06	43	0.20
9	14562.2	11909.4	12710.7	1.91	16	0.73	14225.7	4.12	37	0.12
10	15018.6	11939.9	13104.5	1.77	18	0.62	14718.7	4.86	45	0.09
11	13819.6	10367.8	11175.9	1.75	17	0.76	12958.9	6.37	33	0.24
12	14719.4	10896.2	12036.6	3.15	21	0.70	13740.2	7.96	31	0.25
13	13174.3	9826.6	10954.4	4.90	23	0.66	12705.5	12.11	32	0.14
14	13548.7	9823.7	11089.7	9.34	25	0.66	12974.4	13.83	32	0.15
15	12531.0	9354.7	10697.8	12.38	28	0.57	12219.3	14.20	25	0.09
16	13426.1	9065.7	10253.3	13.24	26	0.72	12573.4	20.56	28	0.19
17	13022.5	9324.8	10420.6	18.55	30	0.70	12735.9	27.35	32	0.07
18	12388.6	8522.8	9831.7	22.35	34	0.66	12137.6	28.38	29	0.06
19	12697.5	8630.6	10036.8	29.98	39	0.65	12394.9	45.92	34	0.07
20	13246.0	8797.3	10092.8	33.34	40	0.70	12491.8	60.70	35	0.16
21	12699.4	8352.0	9868.1	45.71	45	0.65	12260.4	75.16	34	0.10
22	12032.2	8078.2	9519.4	50.81	47	0.63	11859.1	98.45	32	0.04
23	12378.4	8123.9	9376.2	53.98	46	0.70	11911.7	154.01	37	0.10
24	11871.1	7996.5	9250.5	65.78	51	0.67	11480.9	203.40	41	0.10
25	11673.5	7494.2	8717.8	81.18	54	0.70	11187.1	172.77	28	0.11
Asymmetric:										
5	12372.2	10758.2	11126.8	1.80	46	0.77	12373.2	2.10	56	0.0
6	11883.1	10291.6	10596.9	1.79	38	0.80	10684.6	1.51	37	0.75
7	13204.9	10486.3	11369.9	1.95	42	0.67	12746.3	2.98	55	0.16
8	13363.4	10116.3	11194.4	2.08	38	0.66	12878.5	3.24	45	0.14
9	13063.2	9610.1	10546.5	2.52	37	0.72	12135.1	5.24	48	0.26
10	12921.5	9427.1	10461.9	3.26	33	0.70	12500.8	7.35	43	0.12
11	12997.5	8965.8	9891.20	4.41	34	0.77	12089.9	11.99	49	0.22
12	11434.8	8723.3	9682.4	5.79	29	0.64	10897.9	10.87	35	0.19
13	12171.5	8421.6	9608.1	8.55	33	0.68	11584.3	21.11	41	0.15
14	11838.3	8182.7	9005.6	9.95	37	0.77	10635.6	20.66	38	0.32
15	12428.8	8094.6	9564.7	16.83	35	0.66	11748.8	37.05	45	0.15
16	12135.7	7726.8	8764.6	16.64	35	0.76	11119.8	47.34	51	0.23
17	11832.2	7241.3	8682.7	21.69	35	0.68	10094.7	60.02	50	0.37
18	11662.2	7380.9	8544.8	24.40	36	0.72	11104.6	119.24	59	0.13
19	12095.9	7264.6	8560.4	31.22	38	0.73	11207.8	157.51	57	0.18
20	11802.8	6951.1	8200.1	34.92	37	0.74	11162.4	271.81	63	0.13
21	11288.2	6948.5	8140.5	51.44	40	0.72	10819.1	435.246	58	0.10
22	11741.0	6906.8	7950.1	51.61	43	0.78	10517.2	480.15	58	0.25
23	11549.7	6821.3	7688.3	59.05	45	0.81	10549.3	671.15	68	0.21
24	11239.1	6580.4	7716.1	84.03	43	0.75	10180.1	877.83	69	0.22
25	11434.8	6663.1	7785.6	105.36	44	0.76	10422.3	1698.31	70	0.21

REFERENCES

[1] Amaldi, E., G. Galbiati, and F. Maffioli, *On minimum reload cost paths, tours, and flows*, Networks, *57* (2011), 254–260.

[2] Fischer, A., and C. Helmberg, "The symmetric quadratic traveling salesman problem," F. Mathematik, T.U. Chemnitz, Preprint 2011-8, 2011.

[3] Fischer, A., "The asymmetric quadratic traveling salesman problem," F. Mathematik, T.U. Chemnitz, Preprint 2011-19, 2011.

[4] Fischer, F., G. Jäger, A. Lau, and P. Molitor, "Complexity and algorithms for the traveling salesman problem and the assignment problem of second order," F. Mathematik, T.U. Chemnitz, Preprint 2009-16, 2009.

[5] Galbiati, G., S. Gualandi, and F. Maffioli, *On minimum reload cost cycle cover*, Electronic Notes in Discrete Mathematics, *36* (2010), 81–88.

[6] Held, M., and R. M. Karp, *The Traveling-Salesman Problem and Minimum Spanning Trees*, Operations Research, *18* (1970), 1138–1162.

[7] Jäger, G., and P. Molitor, *Algorithms and experimental study for the traveling salesman problem of second order*, Lecture Notes in Computer Science, *5165* (2008), 211–224.

[8] Laporte, G., H. Mercure, Y. Nobert, *Generalized Traveling Salesman Problem through n set of nodes: the assymmetric case*, Discrete Applied Mathematics, *18* (1987), 185–197.

[9] Noon, C.E, and J.C. Bean, *A Lagrangian Based approach for the Assymmetric Generalized Traveling Salesman Problem*, Operations Research, *39* (1991), 623–632.

[10] Gamvros, I., L. Gouveia, and S. Raghavan, *Reload Cost Trees and Network Design*, Networks, *59* (2012), 365–379.

[11] Gilmore, P.C, and R.E. Gomory, *A Linear Programming approach to the Cutting-stock problem*, Operations Research, *9* (1961), 849–859.

[12] Kallehauge, B., J. Larsen, O. B. G. Madsen,"Lagrangian duality applied on wehicle routing with time windows", Technical report IMMM-IR-2001-9, Information and Mathematical Modeling, Technical University of Denmark, KGS, Lyngby, Denmark, 2001.

[13] Kim, S., K.-N. Chang, J.-Y. Lee, *A descent method with linear programming subproblems for nondifferentiable convex optimization*, Math. Programming, *71* (1995), 17–28.

[14] Marsten, R. E., W.W. Hogan, J.W. Blankenship, *The BOXSTEP method for large-scale optimization*, Operations Research, *9* (1975), 389–405.

[15] Du Merle, O., D. Villeneuve, J. Desrosiers, P. Hansen, *Stabilized column generation, Discrete Math. 194* (1999), 229–237.

[16] Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W.P. and Vance, P.H., *Branch-and-Price: Column Generation for Solving Huge Integer Programs*, Operations Research *46* (1998), 316–329.

[17] Fourer, R., D. M. Gay, B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, Duxbury Press, ISBN 978-0-534-38809-6, (2002).

[18] http://www.gurobi.com/documentation/5.5/reference-manual.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 385–392

A hybrid method for modeling and solving constrained search problems

Paweł Sitek
Kielce University of Technology
Al. 1000-lecia PP 7, 25-314
Kielce, Poland, Institute of Management and Control Systems
e-mail:sitek@tu.kielce.pl

Jarosław Wikarek
Kielce University of Technology
Al. 1000-lecia PP 7, 25-314
Kielce, Poland, Institute of Management and Control Systems
e-mail:j.wikarek@tu.kielce.pl

***Abstract*—The paper presents a concept and the outline of the implementation of a hybrid approach to modeling and solving constrained problems. Two environments of mathematical programming (MP) and logic programming (LP) were integrated. The strengths of integer programming (IP) and constraint logic programming (CLP), in which constraints are treated in a different way and different methods are implemented, were combined to use the strengths of both. The proposed approach is particularly important for the decision models with an objective function and many discrete decision variables added up in multiple constraints.**

To validate the proposed approach, two illustrative examples are presented and solved. The first example is the authors' original model of cost optimization in the supply chain with multimodal transportation. The second one is the two-echelon variant of the well-known Capacitated Vehicle Routing Problem, 2E-CVRP.

I. Introduction

THE vast majority of models [1]–[4] of decision support and/or optimization in manufacturing, distribution, supply chain management, etc., have been formulated as the mixed integer linear programming (MILP) or integer programming (IP) problems and solved using the operations research (OR) methods. Their structures are similar and proceed from the principles and requirements of mathematical programming. The constraint-based environments have the advantage over traditional methods of mathematical modeling in that they work with a much broader variety of interrelated constraints (resource, time, technological, and financial) and allow producing "natural" solutions for highly combinatorial problems.

A. Constraint-based environments

We strongly believe that the constraint-based environment [5]–[7] offers a very good framework for representing the knowledge and information needed for the decision support. The central issue for a constraint-based environment is a constraint satisfaction problem. Constraint satisfaction problems (CSPs) are the mathematical problems defined as a set of elements whose state must satisfy a number of constraints. CSPs represent the entities in a problem as a homogeneous collection of finite constraints over variables, which are solved using constraint satisfaction methods. CSPs are the subject of intense study in both artificial intelligence and operations research, since the regularity in their formulation provides a common basis for analyzing and solving the problems of many unrelated families [5]. Formally, a constraint satisfaction problem is defined as a triple (X,D,C), where X is a set of variables, D is a domain of values, and C is a set of constraints. Every constraint is in turn a pair (t,R) (usually represented as a matrix), where t is an n-tuple of variables and R is an n-ary relation on D. An evaluation of the variables is a function from the set of variables to the domain of values, $v:X \rightarrow D$. An evaluation v satisfies constraint $((x_1, \ldots,x_n),R)$ if $(v(x_1),..v(x_n)) \in R$. A solution is an evaluation that satisfies all constraints.

Constraint satisfaction problems on finite domains are typically solved using a form of search. The most widely used techniques include variants of backtracking, constraint propagation, and local search. Constraint propagation embeds any reasoning that consists in explicitly forbidding values or combinations of values for some variables of a problem because a given subset of its constraints cannot be satisfied otherwise [26].

CSPs are frequently used in constraint programming. Constraint programming is the use of constraints as a programming language to encode and solve problems.

Constraint logic programming (CLP) is a form of constraint programming (CP), in which logic programming is extended to include concepts from constraint satisfaction. A constraint logic program is a logic program that contains constraints in the body of clauses. Constraints can also be present in the goal. These environments are declarative.

The declarative approach and the use of logic programming provide incomparably greater possibilities for decision problems modeling than the pervasive approach based on mathematical programming.

B. Paper contents

In this paper we focus on the problem of modeling and solving decision problems using the novel hybrid approach. Having combined the strengths of MILP and CP/CLP (II, III), we developed the environment that ensures the better and easier way of problem modeling and implementation and that provides the more effective search solution (IV). In

978-83-60810-53-8/$25.00

order to verify the proposed approach, two illustrative examples are presented (V).

II. Motivation

Based on [1]–[4], and our previous work [6], [8]–[12], we observed some advantages and disadvantages of these environments.

An integrated approach of constraint programming (CP) and mixed integer programming (MIP) can help to solve optimization problems that are intractable with either of the two methods alone [13]–[16]. Although operations research (OR) and constraint programming (CP) have different roots, the links between the two environments have grown stronger in recent years.

Both MIP/MILP/IP and finite domain CP/CLP involve variables and constraints. However, the types of the variables and constraints that are used, and the way the constraints are solved, are different in the two approaches [16].

MILP relies completely on linear equations and inequalities in integer variables, i.e., there are only two types of constraints: linear arithmetic (linear equations or inequalities) and integrity (stating that the variables have to take their values in the integer numbers). In finite domain CP/CLP, the constraint language is richer. In addition to linear equations and inequalities, there are various other constraints: disequalities, nonlinear, symbolic (*alldifferent, disjunctive, cumulative* etc).

The motivation behind this work was to create a hybrid approach for supply chain modeling and optimization instead of using integer programming or constraint programming separately. We developed the hybrid framework for modeling and optimization of supply chain problems. In both MILP/MIP and CP/CLP, there is a group of constraints that can be solved with ease and a group of constraints that are difficult to solve. The easily solved constraints in MILP/MIP are linear equations and inequalities over rational numbers.

Integrity constraints are difficult to solve using mathematical programming methods and often the real problems of MIP / MILP make them NP-hard.

In CP/CLP, domain constraints with integers and equations between two variables are easy to solve. The system of such constraints can be solved over integer variables in polynomial time. The inequalities between two variables, general linear constraints (more than two variables), and symbolic constraints are difficult to solve, which makes real problems in CP/CLP NP-hard. This type of constraints reduces the strength of constraint propagation. As a result, CP/CLP is incapable of finding even the first feasible solution.

It follows from the above that what is difficult to solve in one environment can be easy to solve in the other.

The motivation was to offer the most effective tools for model–specific constraints and solution efficiency.

III. State Of The Art

As mentioned in Chapter I, the vast majority of decision-making models for the problems of production, logistics, supply chain are formulated in the form of mathematical programming (MIP, MILP, IP).

Due to the structure of these models (summing of discrete decision variables in the constraints and the objective function) and a large number of discrete decision variables (integer and binary) they can only be applied to small problems. Another disadvantage is that only linear constraints can be used. In practice, the issues related to the production, distribution and supply chain constraints are often logical, nonlinear, etc. For these reasons the problem was formulated in a new way,

In our approach to modeling and optimization of constrained search problems we proposed the optimization environment, where:

- knowledge related to the problem can be expressed as linear and logical constraints (implementing all types of constraints of the previous MILP/MIP models [10]–[14] and introducing new types of constraints (logical, nonlinear, symbolic etc.));
- the optimization model solved using the proposed framework can be formulated as a pure model of MILP/MIP or of CP/ CLP, or it can also be a hybrid model;
- the problem is modeled in CP/CLP, which is far more flexible than MIP/MILP/IP;
- the novel method of constraint propagation is introduced (obtained by transforming the optimization model to explore its structure);
- constrained domains of decision variables, new constraints and values for some variables are transferred from CP/CLP into MILP/MIP;
- the efficiency of finding solutions to the problems of larger sizes is increased.

As a result, we obtained the more effective search solution for a certain class of decision and optimization constrained problems.

IV. Hybrid solution environment

Both environments have advantages and disadvantages. Environments based on the constraints such as CLPs are declarative and ensure a very simple modeling of decision problems, even those with poor structures if any. The problem is described by a set of logical predicates. The constraints can be of different types (linear, non-linear, logical, binary, etc.). The CLP does not require any search algorithms. This feature is characteristic of all declarative backgrounds, in which modeling of the problem is also a solution, just as it is in Prolog, SQL, etc. The CLP seems perfect for modeling and solving any decision problem.

In OR numerous models of decision-making have been developed and tested, particularly in the area of decision optimization. Constantly improved methods and mathematical programming algorithms, such as the simplex algorithm,

branch and bound, branch-and-cost [20] etc., have become classics now.

The proposed method's strength lies in high efficiency of optimization algorithms and a substantial number of tested models. The decision problems we deal with in this paper, very common in manufacturing, logistics, supply chain, etc., have a number of decision variables, including binary and integer ones, which are aggregated in the constraints.

Traditional methods when used alone to solve complex problems provide unsatisfactory results. This is related directly to different treatment of variables and constraints in those approaches (II). The proposed hybrid approach, a composition of methods as described in Chapter III offers the optimal system for specific contexts.

A. Architecture and Implementation of Hybrid Solution Environment

This Hybrid Solution Environment (HSE) consists of MIP/MILP/CLP/Hybrid models and Hybrid Solution Framework (FSE) to solve them (Fig. 1). The concept of this framework with its phases (P1 .. P5, G1 .. G3) is presented in Fig. 2.

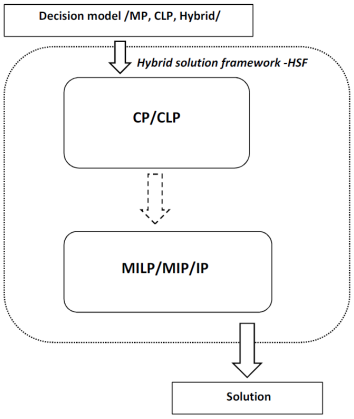


Fig. 1 Scheme of the hybrid solution environment (HSE)

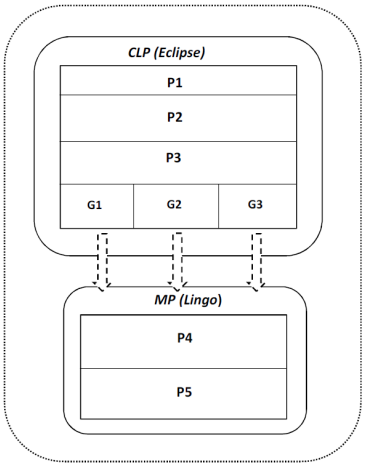


Fig. 2 Detailed scheme of the hybrid solution framework (HSF)

A detailed description of the phases in the order of execution is shown in Table I.

From a variety of tools for the implementation of the CP/CLP in HSE, ECL^{i}PS^e software [21] was selected. ECL^{i}PS^e is an open-source software system for the cost-effective development and deployment of constraint programming applications. Environment for the implementation of MILP/MIP/IP in HSE was LINGO by LINDO Systems. LINGO Optimization Modeling Software is a powerful tool for building and solving mathematical optimization models [22].

TABLE I
DESCRIPTION OF PHASES

Phase	**P1**
Name	Implementation of decision model
Description	The implementation of the model in CLP, the term representation of the problem in the form of predicates.
Phase	**P2**
Name	Transformation of implemented model for better constraint propagation (optional)
Description	The transformation of the original problem aimed at extending the scope of constraint propagation. The transformation uses the structure of the problem. The most common effect is a change in the representation of the problem by reducing the number of decision variables, and the introduction of additional constraints and variables, changing the nature of the variables, etc.
Phase	**P3**
Name	Constraint propagation
Description	Constraint propagation for the model. Constraint propagation is one of the basic methods of CLP. As a result, the variable domains are narrowed, and in some cases, the values of variables are set, or even the solution can be found.
Phase	**G1**
Name	Generation of MILP/MIP/IP model
Description	Generation of the model for mathematical programming. Generation performed automatically using CLP predicate. The resulting model is in a format accepted by the system LINGO.
Phase	**G2**
Name	Generation of additional constraints (optional)
Description	Generation of additional constraints on the basis of the results obtained in step P3
Phase	**G3**
Name	Generation domains of decision variables and other values
Description	Generation of domains for different decision variables and other parameters based on the propagation of constraints. Transmission of this information in the form of fixed value of certain variables and/or additional constraints to the MP.
Phase	**P4**
Name	Merging MILP/MIP/IP model
Description	Merging files generated during the phases G1, G2, G3 into one file. It is a model file format in LINGO system.
Phase	**P5**
Name	Solving MILP/MIP/IP model
Description	The solution model from the previous stage by LINGO. Generation of the report with the results and parameters of the solution.

ECL^{i}PS^e software is the environmental leader in HSE. ECL^{i}PS^e was used to implement the following phases of the framework: P1, P2, P3, G1, G2, G3 (Fig. 2, Table I.). The transformed files of the model were transferred from ECL^{i}PS^e to LINGO where they were merged (P4). Then the complete model was solved using LINGO efficient solvers (P5). Constraint propagation (phase–P3) greatly affected the efficiency of the solution. Therefore phase P2 was introduced. During this phase, the transformation was performed using the structure and properties of the model. This is an optional phase that depends on the modeled problem. The details of this phase will be presented in one of the illustrative examples in Chapter V (cost optimization of supply chain).

V. Illustrative Examples

The proposed HSE environment was verified and tested for two illustrative examples. The first example is the authors' original model of cost optimization of supply chain with multimodal transport (section A). The second is a 2E-CVRP model (section B). It is the known benchmark of a very large number of sets/instances of data and their solutions.

A. Cost optimization of supply chain with multimodal transport

A detailed description of the cost optimization of supply chain models, their constraints, parameters and decision variables etc. are presented in [17] and Table II.

During the first stage, the model was formulated as a MILP problem [9], [10], [17] in order to test the proposed environment (Fig. 1,2) against the classical integer-programming environment [22]. The next step involved the implementation and solving of the hybrid model. Indices, parameters and decision variables in the models together with their descriptions are provided in Table II. The simplified structure of the supply chain network for this model, composed of producers, distributors and customers is presented in Fig.3.

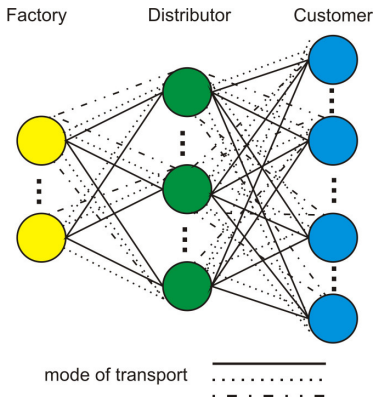


Fig. 3 The simplified structure of the supply chain network

The proposed models are the cost models that take into account three other types of parameters, i.e., the spatial parameters (area/volume occupied by the product, distributor capacity and capacity of transport unit), time (duration of de-

TABLE II
Summary Indices, Parameters and Decision Variables

Symbol	Description
Indices	
k	product type (k=1..O)
j	delivery point/customer/city (j=1..M)
i	manufacturer/factory (i=1..N)
s	distributor /distribution center (s=1..E)
d	mode of transport (d=1..L)
N	number of manufacturers/factories
M	number of delivery points/customers
E	number of distributors
O	number of product types
L	number of mode of transport
Input parameters	
F_s	the fixed cost of distributor/distribution center *s*
P_k	the area/volume occupied by product *k*
V_s	distributor *s* maximum capacity/volume
$W_{i,k}$	production capacity at factory *i* for product *k*
$C_{i,k}$	the cost of product *k* at factory *i*
$R_{s,k}$	if distributor *s* can deliver product *k* then R_{sk}=1, otherwise R_{sk}=0
$Tp_{s,k}$	the time needed for distributor *s* to prepare the shipment of product *k*
$Tc_{j,k}$	the cut-off time of delivery to the delivery point/customer *j* of product *k*
$Z_{j,k}$	customer demand/order *j* for product *k*
Zt_d	the number of transport units using mode of transport *d*
Pt_d	the capacity of transport unit using mode of transport *d*
$Tf_{i,s,d}$	the time of delivery from manufacturer *i* to distributor *s* using mode of transport *d*
$K1_{i,s,k,d}$	the variable cost of delivery of product *k* from manufacturer *i* to distributor *s* using mode of transport *d*
$R1_{i,s,d}$	if manufacturer *i* can deliver to distributor *s* using mode of transport *d* then $R1_{i,s,d}$=1, otherwise $R1_{i,s,d}$=0
$A_{i,s,d}$	the fixed cost of delivery from manufacturer *i* to distributor *s* using mode of transport *d*
$Koa_{i,s,d}$	the total cost of delivery from manufacturer *i* to distributor *s* using mode of transport *d*
$Tm_{s,j,d}$	the time of delivery from distributor *s* to customer *j* using mode of transport *d*
$K2_{s,j,k,d}$	the variable cost of delivery of product *k* from distributor *s* to customer *j* using mode of transport *d*
$R2_{s,jd}$	if distributor *s* can deliver to customer *j* using mode of transport *d* then $R2_{s,j,d}$=1, otherwise $R2_{s,j,d}$=0
$G_{s,j,d}$	the fixed cost of delivery from distributor *s* to customer *j* using mode of transport *d*
$Kog_{s,j,d}$	the total cost of delivery from distributor *s* to customer *j* using mode of transport *d*
Od_d	the environmental cost of using mode of transport *d*
Decision variables	
$X_{i,s,k,d}$	delivery quantity of product *k* from manufacturer *i* to distributor *s* using mode of transport *d*
$Xa_{i,s,d}$	if delivery is from manufacturer *i* to distributor *s* using mode of transport *d* then $Xa_{i,s,d}$=1, otherwise $Xa_{i,s,d}$=0
$Xb_{i,s,d}$	the number of courses from manufacturer *i* to distributor *s* using mode of transport *d*
$Y_{s,j,k,d}$	delivery quantity of product *k* from distributor *s* to customer *j* using mode of transport *d*
$Ya_{s,j,d}$	if delivery is from distributor *s* to customer *j* using mode of transport *d* then $Ya_{s,j,d}$ =1, otherwise $Ya_{s,j,d}$ =0
$Yb_{s,j,d}$	the number of courses from distributor *s* to customer *j* using mode of transport *d*
Tc_s	if distributor *s* participates in deliveries, then Tc_s=1, otherwise Tc_s=0
CW	Arbitrarily large constant

livery and service by distributor, etc.) and the transport mode.

The main assumptions made for the construction of these models were as follows:

- the shared information process in the supply chain consists of resources (capacity, versatility, costs), inventory (capacity, versatility, costs, time), production (capacity, versatility, costs), product (volume), transport (cost, mode, time), demand, etc;
- a part of the supply chain has the structure as in Fig. 3.;
- the transport is multimodal (several modes of transport, a limited number of means of transport for each mode);
- the environmental aspects of use of transport modes are taken into account;
- different products are combined in one batch of transport;
- the cost of supplies is presented in the form of a function (in this approach, linear function of fixed and variable costs);
- models have linear or linear and logical (hybrid model) constraints;
- logical constraints of hybrid model allow the distribution of exclusively one of two selected products in the distribution center and allow the production of exclusively one of two selected products in the factory.

Details of both mathematical models for cost optimization of supply chain are presented in [17].

Objective function

The objective function defines the aggregate costs of the entire chain and consists of five elements. The first element comprises the fixed costs associated with the operation of the distributor involved in the delivery (e.g. distribution centre, warehouse, etc.). The second element corresponds to environmental costs of using various means of transport. Those costs are dependent on the number of courses of the given means of transport, and on the other hand, on the environmental levy, which in turn may depend on the use of fossil fuels and carbon-dioxide emissions.

The third component determines the cost of the delivery from the manufacturer to the distributor. Another component is responsible for the costs of the delivery from the distributor to the end user (the store, the individual client, etc.). The last component of the objective function determines the cost of manufacturing the product by the given manufacturer.

Formulating the objective function in this manner allows comprehensive cost optimization of various aspects of supply chain management. Each subset of the objective function with the same constrains provides a subset of the optimization area and makes it much easier to search for a solution.

Constraints

The model was based on constraints (2) .. (24) Constraint (2) specifies that all deliveries of product k produced by the manufacturer i and delivered to all distributors s using mode of transport d do not exceed the manufacturer's production capacity.

Constraint (3) covers all customer j demands for product k ($Z_{j,k}$) through the implementation of delivery by distributors s (the values of decision variables $Y_{i,s,k,d}$). The flow balance of each distributor s corresponds to constraint (4). The possibility of delivery is dependent on the distributor's technical capabilities - constraint (5). Time constraint (6) ensures the terms of delivery are met. Constraints (7a), (7b), (8) guarantee deliveries with available transport taken into account.

The hybrid model was additionally enriched with logical constraints.

First logical constraint allows the distribution of exclusively one of the two selected products in the distribution center s. Second logical constraint allows the production of exclusively one of the two selected products in the factory i.

These constraints stem from technological, marketing, sales or safety restrictions. Therefore, some products cannot be distributed and/or produced together. The constraint can be re-used for different pairs of product k and for some or all of the distribution centers s and factories i. A logical constraint like this cannot be easily implemented in a MILP model.

Model transformation

Due to the nature of the decision problem (adding up decision variables and constraints involving a lot of variables), the constraint propagation efficiency decreases dramatically. Constraint propagation is one of the most important methods in CLP affecting the efficiency and effectiveness of the CLP and hybrid optimization environment (Fig. 1, Table I). For that reason, research into more efficient and more effective methods of constraint propagation was conducted. The results included different representation of the problem and the manner of its implementation.

The classical problem modeling in the CLP environment consists in building a set of predicates with parameters.

Each CLP predicate has a corresponding multi-dimensional vector representation. While modeling both problems, quantities i, s, k, d and decision variable $X_{i,s,k,d}$ were vector parameters (Fig. 4a). As shown in Fig. 4b, for each vector there were 5 values to be determined, defining the size of the delivery, factories, distributors involved in the delivery and the mode of transport.

[Z_n,P,M,D,F,Tu,Tu,Oq,X,T]

Fig. 4a Representation of the problem in the classical approach-definition

[[z_1,p1,m1,_,_,_,_,10,_,8],
[z_2,p1,m2,_,_,_,_,20,_,6],…]

Fig. 4b Representation of the problem in the classical approach- the process of finding a solution

The process of finding the solution may consist in using the constraint propagation methods, variable labeling and the

backtracking mechanism. The numbers of parameters that must be specified/labeled in the given predicate/vector critically affect the quality of constraint propagation and the number of backtracks. In both models presented above, the classical problem representation included five parameters: *i, s, k, d* and $X_{i,s,k,d}$. Considering the domain size of each parameter, the process was complex and time-consuming. In addition, the above representation (Fig. 4a, Fig. 4b) arising from the structure of the problem is the cause of many backtracks.

Our idea involved the transformation of the problem by changing its representation without changing the very problem. All permissible routes were first generated based on the fixed data and a set of orders, then the specific values of parameters *i, s, k, d* were assigned to each of the routes. In this way, only decision variables $X_{i,s,k,d}$ (deliveries) had to be specified (Fig. 5). This transformation fundamentally improved the efficiency of the constraint propagation and reduced the number of backtracks. A route model is a name adopted for the models that underwent the transformation.

[[name_1,f1,p1,c1,m1,s1,s1,5,12,100,_],
[name_2,f1,p1,c1,m1,s1,s2,6,14,100,_],
[name_3,f1,p1,c1,m1,s2,s1,6,22,100,_],...]

Fig. 5 Representation of the problem in the novel approach- set of feasible routes

Symbols necessary to understand both the representation of the problem and their descriptions are presented in Table III.

TABLE III
SYMBOLS USED IN THE REPRESENTATION OF THE PROBLEM

Symbol	Description
Z_n	order number
P	products, P∈ $\{p_1,p_2, \ldots ,p_o\}$
M	customers, M∈$\{m_1,m_2, \ldots m_m\}$
D	distributors, D∈$\{c_1,c_2, \ldots c_e\}$
F	factories, F∈$\{f_1,f_2, \ldots f_n\}$
Tu	transport unit, Tu∈$\{s_1,s_2, \ldots s_l\}$
T	delivery time/period
Oq	order quantity
X	delivery quantity
Name_	routes name-number

B. Two-Echelon Capacitated Vehicle Routing Problem

The 2E-CVRP is proposed as a benchmark verifying the presented approach. The Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP) is an extension of the classical Capacitated Vehicle Routing Problem (CVRP) where the delivery depot-customers pass through intermediate depots (called satellites). As in CVRP, the goal is to deliver goods to customers with known demands, minimizing the total delivery cost in the respect of vehicle capacity constraints. Multi-echelon systems presented in the literature usually explicitly consider the routing problem at the last level of the transportation system, while a simplified routing problem is considered at higher levels [18], [19], [23].

In 2E-CVRP, the freight delivery from the depot to the customers is managed by shipping the freight through intermediate depots. Thus, the transportation network is decomposed into two levels (Fig. 6): the 1st level connecting the depot (*d*) to intermediate depots (*s*) and the 2nd one connecting the intermediate depots (*s*) to the customers (*c*). The objective is to minimize the total transportation cost of the vehicles involved in both levels. Constraints on the maximum capacity of the vehicles and the intermediate depots are considered, while the timing of the deliveries is ignored.

From a practical point of view, a 2E-CVRP system operates as follows (Fig. 6):

- freight arrives at an external zone, the depot, where it is consolidated into the 1st-level vehicles, unless it is already carried into a fully-loaded 1st-level vehicles;
- each 1st-level vehicle travels to a subset of satellites that will be determined by the model and then it will return to the depot;
- at a satellite, freight is transferred from 1st-level vehicles to 2nd-level vehicles;

The mathematical model (MILP) was taken from [17]. It required some adjustments and error corrections. Table IV shows the parameters and decision variables of 2E-CVRP. Figure 6 shows an example of the 2E-CVRP - transportation network.

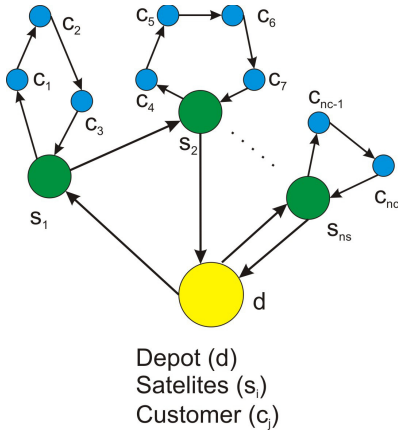


Fig. 6 Example of 2E-CVRP transportation network

The transformation of this model in the hybrid approach focused on the resizing of $Y_{k,i,j}$ decision variable by introducing additional imaginary volume of freight shipped from the satellite and re-delivered to it. Such transformation resulted in two facts. First of all, it forced the vehicle to return to the satellite from which it started its trip. Secondly, it reduced decision variable $Y_{k,i,j}$ to variable $Y_{i,j}$ which decreased the size of the combinatorial problem.

VI. NUMERICAL EXPERIMENTS

A. Cost optimization of supply chain with multimodal transport

In order to verify and evaluate the proposed approach, many numerical experiments were performed. All the examples relate to the supply chain with two manufacturers

TABLE IV
SUMMARY INDICES, PARAMETERS AND DECISION VARIABLES

Symbol	Description
Indices	
n_s	Number of satelites
n_c	Number of customers
$V_0 = \{v_o\}$	Deport
$V_s = \{v_{s1}, \ldots, v_{sns}\}$	Set of satellites
$V_c = \{v_{c1}, \ldots, v_{cnc}\}$	Set of customers
Input parameters	
m_1	Number of the 1st-level satelites
M_2	Number of the 2nd-level satelites
k_1	Capacity of the vehicles for the 1st level
k_2	Capacity of the vehicles for the 2nd level
d_i	Demand required by customer i
$c_{i,j}$	Cost of the arc (i,j)
s_k	Cost of loading/unloading operations of a unit of freight in satelite k
Decision variables	
$X_{i,j}$	Is an integer variable of the 1st-level routing and is equal to the number of 1st-level vehicles using arc (i, j).
$Y_{k,i,j}$	Is a binary variable of the 2nd-level routing and is equal to 1 if a 2nd-level vehicle makes a route starting from satellite k and goes from node i to node j and 0 otherwise
$Q1_{i,j}$	freight flow arc ij for the 1st-level
$Q2_{k,i,j}$	freight arc ij where k represents the satellite where the freight is passing through.
$z_{k,j}$	Binary variable that is equal to 1 if the freight to be delivered to customer j is consolidated in satellite k and 0 otherwise

(i=1..2), three distributors *(s=1..3)*, five customers *(j=1..5),* three modes of transport *(d=1..3),* and ten types of products *(k=1..10).* Other parameter values are shown in Appendix A1 [17].

The first series of experiments was designed to show the advantages of the hybrid approach used.

The experiments began with six examples: E1, E2, E3, E4, E5, E6 for the problem formulated in MILP (V) [17]. Two approaches were used to implement the proposed model: mathematical programming (LINGO) and the hybrid approach (LINGO, Eclipse, transformation). The examples E1 .. E6 varied in terms of the number of orders (*No*). The set of all orders for calculation examples are given in Appendix A.

The experiments were conducted to optimize examples E7, E8, which are implementations of the hybrid model (with logical constraints) in the hybrid approach.

The implementation of logic constraints for the hybrid model was as follows: product $k = 5$ cannot be distributed with product $k= 6$; product $k = 2$ cannot be distributed with product $k = 8$, and these products cannot be produced together. The results in the form of the objective function, the computation time, the number of discrete decision variables and constraints are shown in Table V.

TABLE V
THE RESULTS OF NUMERICAL EXAMPLES FOR BOTH APPROACHES

E(No)	MILP-LINGO				MILP-Hybrid			
	Fc	***T***	***V***	***C***	***Fc***	***T***	***V***	***C***
E1(5)	6680	7	1389	1351	6680	2	117	172
E2(10)	20439	28	1389	1621	20439	3	173	172
E3(15)	29107	55	1389	1891	29107	9	245	172
E4(20)	45710*	600**	1389	2161	45654	18	301	172
E5(25)	46660*	600**	1389	2431	46150	235	376	172
E6(30)	48946*	600**	1389	2701	48006	375	429	172
P(No)	**Hybrid-Hybrid**							
	Fc	***T***	***V***	***C***				
E7(10)	21143	194	193	202				
E8(20)	46069	366	321	202				

Fc	the optimal value of the objective function
T	Solution finding time
V/C	the number of integer variables/constraints
*	the feasible value of the objective function after the time T
**	calculation was stopped after 600s

The analysis of the outcome indicates that the hybrid approach provided better results in terms of the time needed to find the solution in each case, and to obtain the optimal solution in some cases, which was impossible to do within the acceptable time limits using the traditional approaches.

B. Two-Echelon Capacitated Vehicle Routing Problem

For the final validation of the proposed hybrid approach, the benchmark (2E-CVRP) was selected. 2E-CVRP, a well described and widely discussed problem, corresponded to the issues to which our hybrid approach was applied.

The instances for computational examples were built from the existing instances for CVRP [24] denoted as E-n13-k4. All the instance sets can be downloaded from the website [25]. The instance set was composed of 5 small-sized instances with 1 depot, 12 customers and 2 satellites. The full instance consisted of 66 small-sized instances because the two satellites were placed over twelve customers in all 66 possible ways (number of combinations: 2 out of 12).

All the instances had the same position for depot and customers, whose coordinates were the same as those of instance E-n13-k4. Small-sized instances differed in the choice of two customers who were also satellites (En13-k4-2 (1,3), En13-k4-6 (1,6), En13-k4-61 (9,10) etc.).

The analysis of the results for the benchmark instances demonstrates that the hybrid approach may be a superior approach to the classical mathematical programming. For all examples, the solutions were found 2-16 times faster than they are in the classical approach.

As the presented benchmark was formulated as a MILP problem, the HSF was tested for the solution efficiency. Owing to the hybrid approach the 2E-CVRP models can be extended over logical, nonlinear, and other constraints.

TABLE VI
THE RESULTS OF NUMERICAL EXAMPLES FOR BOTH APPROACHES

E-n13-k4	MILP-LINGO				MILP-Hybrid			
	Fc	*T*	*V*	*C*	*Fc*	*T*	*V*	*C*
En13-k4-2	286	40371	368	1262	286	8720	186	1024
En13-k4-6	230	125	368	1262	230	55	186	1024
En13-k4-9	244	153	368	1262	244	44	186	1024
En13-k4-20	276	535	368	1262	276	32	186	1024
En13-k4-61	338	6648	368	1262	338	407	186	1024

Fc	the optimal value of the objective function
T	time of finding solution
V/C	the number of integer variables/constraints
*	the feasible value of the objective function after the time T

VII. CONCLUSION AND DISCUSSION ON POSSIBLE EXTENSION

The efficiency of the proposed approach is based on the reduction of the combinatorial problem and using the best properties of both environments. The hybrid approach (Table V, Table VI) makes it possible to find solutions better solutions in the shorter time.

In addition to solving larger problems faster, the proposed approach provides virtually unlimited modeling options.

Therefore, the proposed solution is recommended for decision-making problems in the supply chain that has a similar structure to the presented model (V). This structure is characterized by the constraints and objective function in which the decision variables are added together. Further work will focus on running the optimization models with non-linear and logical constraints, multi-objective, uncertainty etc. in the hybrid optimization framework.

References

[1] Kanyalkar, A.P., Adil, G.K., *An integrated aggregate and detailed planning in a multi-site production environment using linear programming.* International Journal of Production Research 43, 2005, pp. 4431–4454.

[2] Perea-lopez, E., Ydstie, B.E., Grossmann, I.E., *A model predictive control strategy for supply chain optimization*. Computers and Chemical Engineering 27, 2003, pp. 1201–1218.

[3] Christian Lang J. *Production and Operations Management: Models and Algorithms.* Production and Inventory Management with Substitutions, Lecture Notes in Economics and Mathematical Systems Volume 636, 2010, pp 9–79.

[4] Dang Quang, Nielsen Izabela, Steger-Jensen Kenn, Madsen Ole, Scheduling a Single Mobile Robot for Part-Feeding Tasks of Production Lines, Journal of Intelligent Manufacturing, 2013, DOI 10.1007/s10845-013-0729-y.

[5] Apt K., Wallace M., *Constraint Logic Programming using Eclipse*, Cambridge University Press, 2006.

[6] Sitek P., Wikarek J., *A Declarative Framework for Constrained Search Problems,* New Frontiers in Applied Artificial Intelligence, Lecture Notes in Artificial Intelligence, Nguyen, NT., et al. (Eds.), Vol. 5027, Springer-Verlag, Berlin-Heidelberg, 2008, pp. 728-737.

[7] Bocewicz G., Wójcik R., Banaszak Z., *AGVs distributed control subject to imprecise operation times,* In: Agent and Multi-Agent Systems: Technologies and Applications, Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, Vol. 4953, 2008, pp. 421–430.

[8] Sitek P., Zaborowski M. *Grouping products in a follow-up production control system for parallel partitioned flow production lines,* Intelligent manufacturing systems IMS 2001: 6th IFAC Workshop,: Pergamon, 2001, New York, pp.122-126.

[9] Sitek P. Wikarek J. *The concept of decision support system structures for the distribution center*, MPER (Management and Production Engineering Review), vol.1, no.3, 2010, pp.63-69

[10] Sitek P., Wikarek J., *Cost optimization of supply chain with multimodal transport,* Federated Conference on Computer Science and Information Systems (FedCSIS), 2012, pp. 1111–1118.

[11] Sitek P., Wikarek J., *Supply chain optimization based on a MILP model from the perspective of a logistics provider*, Management and Production Engineering Review, 2012, pp. 49–61.

[12] Sitek P., Wikarek J., *The Declarative Framework Approach to Decision Support for Constrained Search Problems*, INTECH, 2011, pp 163–182.

[13] Jain V., Grossmann I.E., *Algorithms for hybrid MILP/CP models for a class of optimization problems*, INFORMS Journal on Computing 13(4), 2001, pp. 258–276.

[14] Milano M., Wallace M., *Integrating Operations Research in Constraint Programming,* Annals of Operations Research vol. 175 issue 1, 2010, pp. 37 – 76.

[15] Achterberg T., Berthold T., Koch T., Wolter K., *Constraint Integer Programming: A New Approach to Integrate CP and MIP*, Lecture Notes in Computer Science, Volume 5015, 2008, pp. 6–20.

[16] Bockmayr A., Kasper T.,: *A Framework for Combining CP and IP*, Branch-and-Infer, Constraint and Integer Programming Operations Research/Computer Science Interfaces Series, Volume 27, 2004, pp. 59-87.

[17] Sitek P., Wikarek J., *A hybrid approach to supply chain modeling and optimization*, Federated Conference on Computer Science and Information Systems (FedCSIS), 2013, pp. 1223–1230.

[18] Perboli G., Tadei R., Vigo D., *The Two-Echelon Capacitated Vehicle Routing Problem: Models and Math-Based Heuristics*,Transportation Science, 2011,v45, pp.364-380.

[19] Crainic, T., Ricciardi, N., Storchi, G., 2004. *Advanced freight transportation systems for congested urban areas.* Transportation Research part C 12, 119–137.

[20] Schrijver A., *Theory of Linear and Integer Programming,* ISBN 0-471-98232-6, John Wiley & sons, 1998.

[21] www.eclipse.org

[22] www.lindo.com

[23] Ricciardi, N., Tadei, R., Grosso, A., *Optimal facility location with random throughput costs*. Computers and Operations Research 29 (6), 2002,593–607.

[24] Christofides, N., Elion, S.,. *An algorithms for the vehicle dispatching problem.* Operational Research Quarterly 20, 1969, pp.309–318.

[25] http://www.orgroup.polito.it/

[26] Rossi F., Van Beek P., Walsh T., *Handbook of Constraint Programming (Foundations of Artificial Intelligence)*, Elsevier Science Inc. New York, NY, USA © 2006.

APPENDIX A

TABLE I
THE SET OF ORDERS FOR COMPUTATIONAL EXAMPLES E1-E8

Name	k	j	$T_{k,j}$	$Z_{j,k}$	Name	k	j	$T_{k,j}$	$Z_{j,k}$
z_01	p1	m1	8	10	z_11	p1	m3	8	15
z_02	p2	m2	12	10	z_12	p2	m4	12	20
z_03	p3	m3	10	25	z_13	p3	m5	10	25
z_04	p4	m4	8	30	z_14	p4	m1	8	30
z_05	p5	m5	12	10	z_15	p5	m2	12	30
z_06	p6	m1	8	15	z_16	p6	m3	8	15
z_07	p7	m2	12	20	z_17	p7	m4	12	20
z_08	p8	m3	10	25	z_18	p8	m5	10	25
z_09	p9	m4	8	30	z_19	p9	m1	8	30
z_10	p10	m5	12	30	z_20	p10	m2	12	35
z_21	p1	m5	8	2	z_26	p6	m5	8	3
z_22	p2	m1	12	1	z_27	p7	m3	12	2
z_23	p3	m4	10	2	z_28	p8	m4	10	2
z_24	p4	m5	8	1	z_29	p9	m2	8	2
z_25	p5	m3	12	1	z_30	p10	m1	12	2

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 393–400

Biased Random Key Genetic Algorithm with Hybrid Decoding for Multi-objective Optimization

Panwadee Tangpattanakul
CNRS, LAAS
7 avenue du Colonel Roche and
Univ de Toulouse, LAAS
F-31400 Toulouse, France
Email: ptangpat@laas.fr

Nicolas Jozefowiez
CNRS, LAAS
7 avenue du Colonel Roche and
Univ de Toulouse, INSA, LAAS
F-31400 Toulouse, France
Email: njozefow@laas.fr

Pierre Lopez
CNRS, LAAS
7 avenue du Colonel Roche and
Univ de Toulouse, LAAS
F-31400 Toulouse, France
Email: lopez@laas.fr

***Abstract*—A biased random key genetic algorithm (BRKGA) is an efficient method for solving combinatorial optimization problems. It can be applied to solve both single-objective and multi-objective optimization problems. The BRKGA operates on a chromosome encoded as a key vector of real values between $[0, 1]$. Generally, the chromosome has to be decoded by using a single decoding method in order to obtain a feasible solution. This paper presents a hybrid decoding, which combines the operation of two single decoding methods. This hybrid decoding gives two feasible solutions from the decoding of one chromosome. Experiments are conducted on realistic instances, which concern acquisition scheduling of agile Earth observing satellites.**

I. Introduction

THIS PAPER proposes a hybrid decoding to apply with a biased random key genetic algorithm (BRKGA) for solving multi-objective optimization problems. We experiment on instances of multi-user observation scheduling problem for agile Earth observing satellites (EOSs).

The biased random key genetic algorithm (BRKGA) was first presented in [1]. BRKGA combines the concept of random key and the principles of genetic algorithms. The random key vector represents one solution. In the process to apply BRKGA for solving combinatorial problems, there is a step, which depends on the considered problem. It is a decoding step, which is used to decode the random key chromosome to become a feasible solution. The efficient decoding method can obtain a good solution. Hence, the specification of the decoding step is an important issue for BRKGA.

BRKGA was used to solve combinatorial optimization problems in various domains (e.g. communication, transportation, scheduling) [3]. For example, BRKGA was applied to solve the fiber installation in an optical network optimization problem [4]. The objective function was to minimize the cost of the optical components necessary to operate the network. In [5], a resource-constrained project scheduling problem with makespan minimization was solved by BRKGA. Nevertheless, all these works address optimization problems involving a single objective function. This paper considers multi-objective optimization. Several real world problems, e.g., in the area of engineering research and design, can be modeled as multi-objective optimization problems. When many objectives are considered, the search will not give a unique solution but a set of solutions. Hence, our idea for improving the efficiency of BRKGA for solving multi-objective optimization problems, is to combine the importance of its decoding step and the need of a non-unique solution of multi-objective optimization.

A hybrid decoding, which combines two single decoding methods, is proposed in this paper. A hybrid decoding can obtain more than one solution from the decoding of one chromosome. Two separate single-decoding and the hybrid decoding are experimented on the multi-user observation scheduling problem for agile Earth observing satellites.

The mission of Earth observing satellites (EOSs) is to obtain photographs of the Earth surface satisfying users' requirements. When the ground station center receives the requests from users, it has to manage the requirements by selecting and scheduling a subset of photographs and transmit the schedule, which consists of a sequence of selected photographs, to the satellites. We consider an agile satellite, which has only one on-board camera that can move around three axes: roll, pitch, and yaw. The starting time of each photograph is not fixed; it can slide within a given visible time interval. The problem description of agile EOSs scheduling problem is presented in the ROADEF 2003 challenge [10]. This challenge required the scheduling solutions that maximize total profit of the acquired photographs for a single user and have to satisfy all physical constraints of agile EOSs. Algorithms based on simulated annealing [11] and tabu search [12] were particularly proposed for this challenge. In [13], multiple users have been considered. However, a single objective is considered.

The originality of our work also lies in the consideration of multi-user requests, but we need to optimize two objectives. The ground station center should maximize the total profit of the acquired photographs and simultaneously share fairly the satellite resources for all users by minimizing the maximum profit difference between users. In [9], we proposed a biased random-key genetic algorithm (BRKGA) with a single decoding method to solve this multi-objective optimization problem. BRKGA with a single decoding succeeded to obtain quite good solutions. However, the average value of the obtained hypervolumes and the range of the solutions should be improved.

978-83-60810-53-8/$25.00

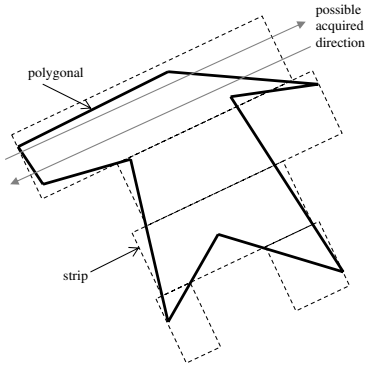


Fig. 1. A polygon is decomposed into several strips; one of two possible directions can be selected for the acquisition of each strip

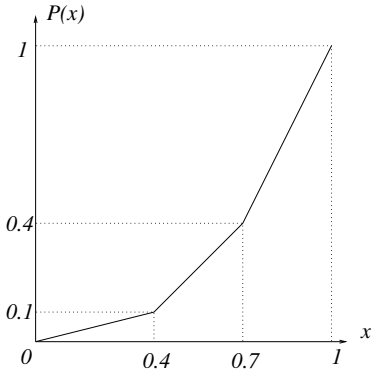


Fig. 2. Piecewise linear function of gain $P(x)$ depending on the effective ratio x of acquired area [10]

For our study, the ROADEF 2003 instances of the observation scheduling problem for agile EOSs are modified in order to consider in case of multi-user requirements. Two possible shapes of area can be required: spot or polygon. The polygon is a big area that the camera cannot take instantaneously. Hence it has to be decomposed into several strips of rectangular shape with fixed width but variable length, as shown in Figure 1. Among two possible directions, one acquisition can be selected for each strip. Two types of photograph can be required: a mono photograph is taken only once, whereas a stereo photograph should be acquired twice in the same direction but from different angles.

The possible starting time interval for taking each acquisition is calculated, depending on the acquired direction, its earliest/latest visible time of the two extremities and the taking duration time of the strip. Moreover, adjacent selected acquisitions must also respect a sufficient transition time. It is a necessary time in order to move the camera from the ending point of the previous acquisition to the beginning point of the next acquisition. These imperative constraints have to be satisfied for finding the feasible solutions, which are the sequences of the selected acquisitions for being transmitted to the satellite. For each solution, the two objective function values can be calculated by using a piecewise linear function of gain. This function is associated with a partial acquisition of the acquired request, as illustrated in Figure 2.

The article is organized as follows. Section II explains the BRKGA for solving multi-objective optimization problems. The proposed hybrid decoding is presented in Section III. Section IV reports the computational results. Finally, conclusions are discussed in Section V.

II. Biased Random Key Genetic Algorithm for Multi-objective Optimization Problems

A genetic algorithm is a metaheuristic method, which operates on several individuals in a population. Individuals should spread through the search space. The genetic algorithm uses the concept of survival of the fittest to find the optimal solutions. Each individual consists of a chromosome, which represents a solution. The process of genetic algorithm is started by generating an initial population with its size equal to p. For generating the next generations, selection, crossover, and mutation operators are applied. The iterations are repeated until a stopping criterion is satisfied.

A biased random key genetic algorithm (BRKGA) was first presented in [1]. The BRKGA has different ways to select two parents for the crossover operation, compared with the original of random key genetic algorithm (RKGA) [2]. For BRKGA, the random key chromosome is formed by several genes, which are encoded by real values in the interval $[0, 1]$. Then, the chromosome is decoded in order to obtain the solution. The decoding strategy is problem dependent. The fitness value of solution is computed in this decoding step. The current population is divided into two groups by using the selection mechanism. Selections are applied to choose p_e preferred chromosomes from the current population to become the elite set. The remaining chromosomes will be stored in the other group of non-elite chromosomes. Then, the process to generate the population in the next generation begins.

The standard procedure for BRKGA can be found in [3]. We will now explain how BRKGA was adapted for multi-objective optimization [9]. We will focus on the selection phase, fitness computation, and population recombination.

A. Population generation for the next iteration

The population of the new generation is generated from three parts, as in Figure 3. The first part is an elite set, which contains p_e preferred chromosomes. The second part is a set of p_m chromosomes, which are generated to avoid the entrapment in a local optimum. These chromosomes are called mutant. They are randomly generated by the same methods, which is used to generate the initial population. The last part is filled by generating offspring from the crossover operation of the elite set and another solution from the current population. Each element in the offspring is obtained from the element of elite parent with the probability ρ_e. Otherwise, the element of offspring is copied from the non-elite parent. Hence, the size of crossover offspring set is equal to $p - p_e - p_m$. The recommended parameter value setting is displayed in Table I.

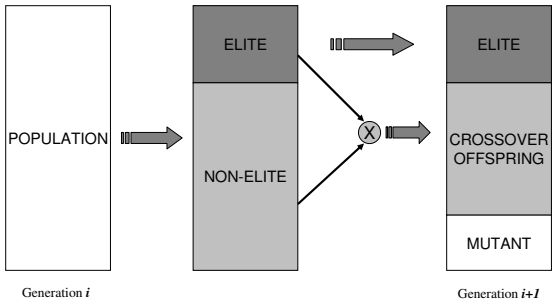


Fig. 3. The population of the new generation by using BRKGA

TABLE I
RECOMMENDED PARAMETER VALUES OF BRKGA [3]

Parameter	Recommended value
p	$p = a.n$, where $1 \leq a \in \mathbb{R}$ is a constant and n is the length of the chromosome
p_e	$0.10p \leq p_e \leq 0.25p$
p_m	$0.10p \leq p_m \leq 0.30p$
ρ_e	$0.5 \leq \rho_e \leq 0.8$

In [3], BRKGA is applied to solve optimization problems arising in several applications. However, all problems consider only one objective. In this work, we study BRKGA for solving a multi-objective optimization problem. The fitness of each chromosome must be taken into account for all objective functions. Algorithms for selecting the preferred chromosomes are needed.

B. Algorithm to select the preferred chromosomes in the context of a multi-objective optimization problem

Three strategies are proposed to select individuals.

1) Fast nondominated sorting and crowding distance assignment: Fast nondominated sorting and crowding distance assignment methods were proposed in the Nondominated Sorting Genetic Algorithm II (NSGA-II) [6]. In our work, the fast nondominated sorting method is used to find the nondominated solutions. If the number of nondominated solutions is more than the parameter setting value of maximum size of elite set, the crowding distance assignment method is applied to select some solutions from the nondominated set to become the elite set. Otherwise all nondominated solutions will become the elite set.

2) $\mathscr{S}$ metric selection evolutionary multi-objective optimization algorithm: $\mathscr{S}$ metric selection evolutionary multi-objective optimization algorithm (SMS-EMOA), which was proposed in [7], is applied to select some solutions in the current population to become the elite set. In our work, we use SMS-EMOA combining with the fast nondominated sorting from NSGA-II. The fast nondominated sorting is applied in order to find the nondominated solutions and SMS-EMOA compute the hypervolume as selection criterion for limiting the size of elite set. The hypervolume selection discards the solution, which obtains the least hypervolume in the set of nondominated solutions and the remaining solutions will become the elite set.

3) Indicator-based evolutionary algorithm based on the hypervolume concept: The use of an indicator based on the hypervolume concept was proposed in the Indicator-Based Evolutionary Algorithm (IBEA) [8]. The indicator based method is used to assign fitness values based on the hypervolume concept to the population members. Then, some solutions in the current population are selected to become elite set for the next population. The indicator based method performs binary tournaments for all solutions in the current population. The selection is implemented environmentally by removing the worst solution from the population and updating the fitness values of the remaining solutions. The worst solution is removed repeatedly until the number of remaining solutions satisfies the recommended size of elite set for BRKGA.

III. DECODING METHODS

In this section, the decoding methods, which are used for obtaining the solutions from the random key chromosomes, are described. A chromosome consists of several genes. Each gene represents one job, which needs to be scheduled. When the processes of genetic algorithm finish, the chromosome is decoded in order to obtain a sequence of jobs, which become the solution of the problem. In this decoding step, the sequence of jobs will be generated. The order to consider each job depends on the priority, which is computed from its associated gene value. The job, which has the highest priority, will be firstly considered to be assigned in a sequence. Then, the next jobs are considered according to the priority order. The considered job can be scheduled in the sequence, only if all constraints are satisfied. Three decoding methods for assigning the priority are studied in this paper. The three methods are:

A. Basic decoding (D1)

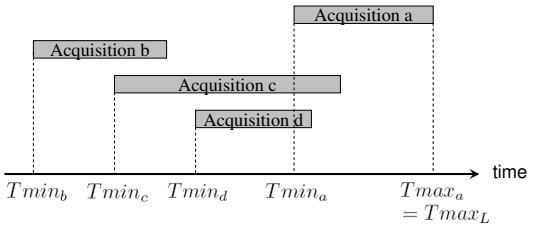
The first decoding method is a basic decoding: the priority is assigned by using directly the gene value:

$$Priority_j = gene_j \tag{1}$$

This decoding method was implemented in the context of multi-objective optimization in [9]. Albeit it gave quite good results, we are convinced that the results regarding average values and standard deviations of hypervolumes can be yet improved. Thus, we searched for an idea to apply some useful data of the problem for assisting the basic decoding.

B. Decoding of gene value and ideal priority combination (D2)

This decoding is presented in [5]. It considers the priority depending on the gene value, and also an ideal priority. For the concept of the ideal priority, the job, which has the earliest possible starting time, should be selected firstly and be scheduled at the beginning of the sequence. Hence the ideal priority gives a higher priority to select and schedule the job



Acquisition	a	b	c	d
Ideal priority	$\frac{Tmax_L - Tmin_a}{Tmax_L}$	$\frac{Tmax_L - Tmin_b}{Tmax_L}$	$\frac{Tmax_L - Tmin_c}{Tmax_L}$	$\frac{Tmax_L - Tmin_d}{Tmax_L}$

Fig. 4. Example of ideal priority calculation

which has the earlier possible starting time. This ideal priority is the real value in the interval $[0, 1]$ which is given by

$$\frac{LLP_j}{LCP}, \tag{2}$$

where LLP_j is the longest length path from the beginning of job j to the end of the project and LCP is the length along the critical path of the project.

The factor that adjusts the priority to account for the gene values of the random key chromosome is given by $(1 + gene_j)/2$. Thus, the second decoding expression of each job j is

$$Priority_j = \frac{LLP_j}{LCP} \times \left[\frac{1 + gene_j}{2}\right] \tag{3}$$

This second decoding method was applied to solve the resource-constrained project scheduling problem with makespan minimization in [5]. In our paper, the second decoding will be implemented to the considered multi-objective optimization problem, which is the multi-user observation scheduling problem for agile Earth observing satellites. Hence, the second decoding expression of each acquisition j becomes:

$$Priority_j = \frac{Tmax_L - Tmin_j}{Tmax_L} \times \left[\frac{1 + gene_j}{2}\right] \tag{4}$$

where $Tmax_L$ is the latest starting time of the last possible acquisition and $Tmin_j$ is the earliest starting time of acquisition j.

The example of the ideal priority calculation of the second decoding method is shown in Figure 4. It is applied to the multi-user observation scheduling problem for agile Earth observing satellites, which needs to select and schedule four acquisitions, which are acquisitions a, b, c, and d. For this example, the sequence of the acquisitions according to the ideal priority, is b, c, d, and a.

C. Hybrid decoding (HD)

Finally, we propose the third decoding method which is the hybrid method. It combines together the first and the second decoding methods. This hybrid method obtains two solutions from one chromosome. When applying the hybrid decoding, the methods to manage the elite set, must be defined. Three methods are tested for selecting the elite set.

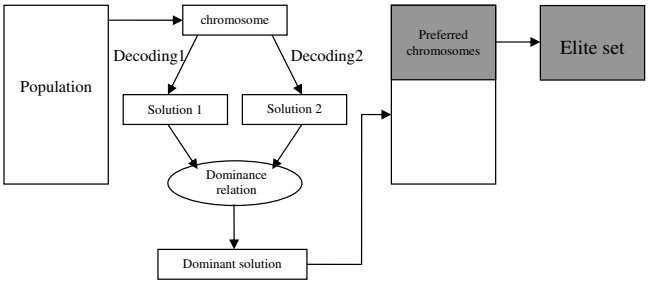


Fig. 5. Elite set management for hybrid decoding - method 1

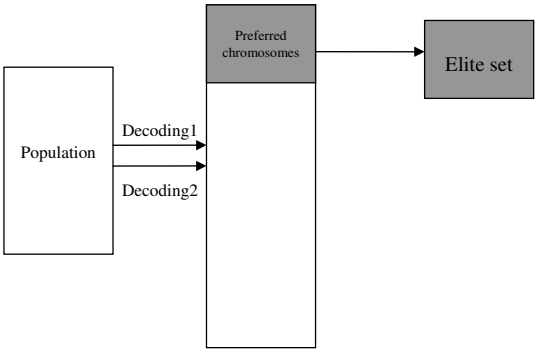


Fig. 6. Elite set management for hybrid decoding - method 2

1) Elite set management - Method 1 (M1): Both solutions, obtained by the two decodings, are compared by using the dominance relation in the Pareto sense. If a solution can dominate the other one, the dominant solution is selected to be stored in the set of solutions. Otherwise, one of the two solutions is selected randomly. The decoding process is repeated until all chromosomes in the population are decoded. When it finishes, the size of the solution set is equal to p. Then, the p_e solutions are selected to become the elite set by using the same methods with only one decoding. The principle of elite set management - method 1 is shown in Figure 5.

2) Elite set management - Method 2 (M2): All chromosomes in the population are decoded by using the two decoding methods. Two solutions are obtained from the decoding of one chromosome. Both of them are stored in the solution set. Hence, the size of the solution set is equal to $2p$, when all chromosomes from the current population are decoded. Then, the p_e solutions are selected from the solution set to become the elite set. The principle of elite set management - method 2 is shown in Figure 6.

3) Elite set management - Method 3 (M3): Each chromosome in the population is firstly decoded by using the priority equation of basic decoding, and the obtained solution is stored in the first solution set. Similarly, the same chromosome is decoded by using the priority equation of the decoding of gene value with ideal priority combination. Then, the obtained solution from this decoding is stored in the second solution set. When all chromosomes in the population are decoded and the solutions are stored into two solution sets, the selection

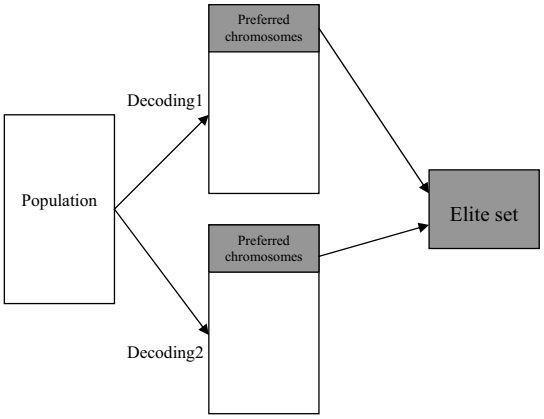


Fig. 7. Elite set management for hybrid decoding

methods are applied to select p_e solutions for becoming the elite set. Hence, the $p_e/2$ preferred solutions must be chosen from each solution set, as shown in Figure 7.

In the decoding step for solving the multi-user observation scheduling problem for agile Earth observing satellites, the priorities for selecting and scheduling each acquisition are computed depending on the decoding methods as previously presented. Then, the solution, which is the sequence of the selected acquisitions, can be generated. The imperative constraints are verified for each acquisition sequentially according to its priority. Each considered acquisition can be assigned in the sequence, only if the obtained sequence can satisfy all constraints. The flowchart of constraint verification and acquisition assignment is depicted in Figure 8. The example of one solution from the smallest size instance is shown in Figure 9. This instance consists of two strips. Hence the number of random key genes, which are associated with the acquisitions, equals to four. This example shows the solution, which is decoded from the basic decoding (the priority to select and schedule of each acquisition equals to its gene value). This decoding step is used to obtain the sequence of the selected acquisitions and the values of the two objective functions.

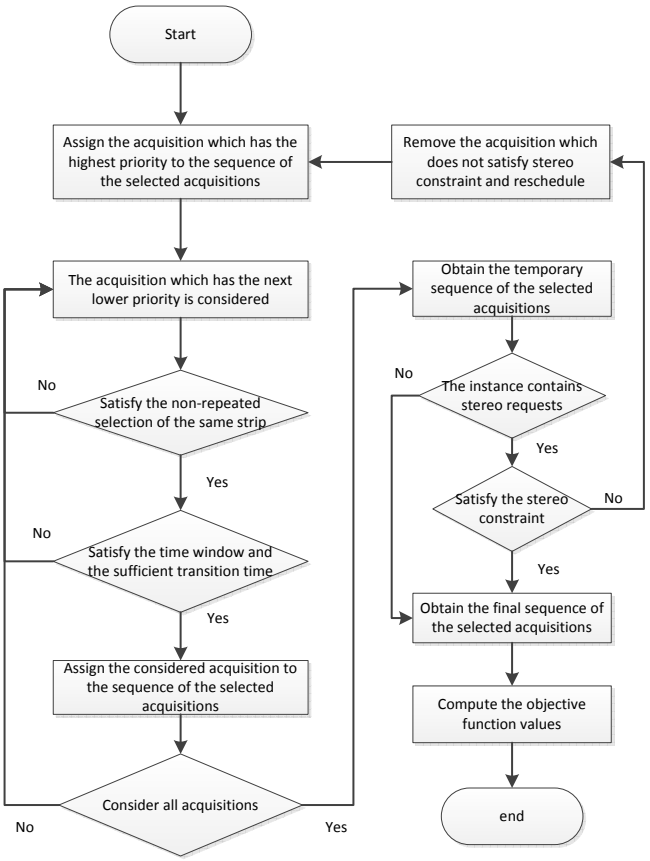


Fig. 8. Flowchart of constraint verification and acquisition assignment

Random-key chromosome	Acquisition0	Acquisition1	Acquisition2	Acquisition3
	0.6984	0.9939	0.6885	0.2509
Sequence of the selected acquisitions	1 | 2			
Total profit	1.04234e+007			
Maximum profit difference	0			

Fig. 9. Solution example from the modified instance, which needs to schedule two strips

IV. Computational Results

The ROADEF 2003 challenge instances (Subset A) are modified for 4-user requirements. The format of instance names are changed to `a_b_c`, where `a` is the number of requests, `b` is the number of stereo requests, and `c` is the number of strips.

For the proposed biased random-key genetic algorithm (BRKGA), parameter values of the algorithm were experimentally tuned for our work. The population size of BRKGA is set equal to the length of the random-key chromosome or twice the number of strips. The sizes of the three parts, which are generated to become the population in the next generation, are set in accordance with the recommended values in Table I. The size of the elite set is equal to the number of non-repeating schedules from the nondominated solutions, but it is not over $0.15p$. The size of mutant set is equal to $0.3p$. The probability of elite element inheritance for crossover operation is set to 0.6. In each iteration, the nondominated solutions are stored in an archive. If there is at least one solution from the current population that can dominate some solutions in the archive, the archive will be updated. Therefore, we use the number of iterations since the last archive improvement to be a stopping criterion. We opt for 50 iterations. Moreover, the computation time is used to be the second stopping criterion. It is adapted to the instance size. The iteration of BRKGA will be stopped, when one of the two stopping criteria is satisfied. The algorithm is implemented in C++ and ten runs per instance are tested. The hypervolumes of the approximate Pareto front are computed by using a reference point of 0 for the first objective (maximizing the total profit) and the maximum of the profit summations of each user for the second one (minimizing the

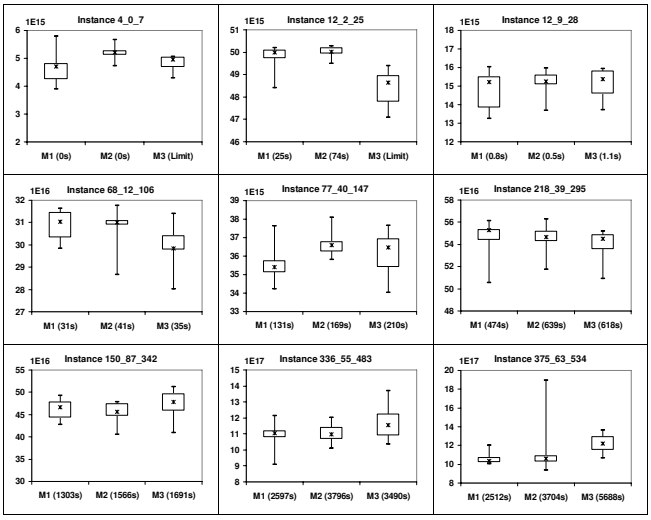


Fig. 10. Comparison of the results of the three methods for management of the elite set for hybrid decoding (M1, M2, and M3) by using the method for elite set selection borrowed from NSGA-II

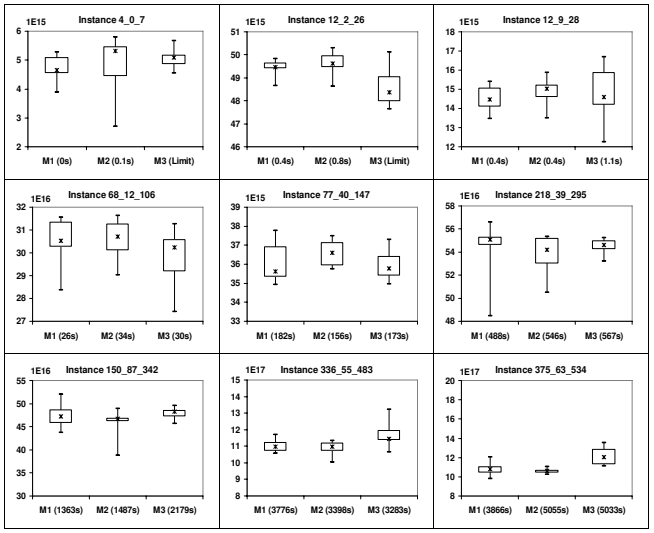


Fig. 11. Comparison of the results of the three methods for management of the elite set for hybrid decoding (M1, M2, and M3) by using the method for elite set selection borrowed from SMS-EMOA

profit difference between users). Three elite selecting methods from three efficient algorithms: NSGA-II, SMS-EMOA, and IBEA, are applied to select some solutions to become the elite set. The set of testing instances consists of ten instances. However, the results of the smallest instance (instance `2_0_2`) cannot be reached, when using the population size equal to the length of the chromosome or twice of number of strips, because the population size is too small for generating the new population from the three sets of chromosomes in BRKGA process. Hence, the results of nine instances will be presented in the experimental results.

Firstly, the three methods of elite set management for hybrid decoding (M1–M3) are tested. The results, which are obtained from each method, are compared. The hypervolume values of the approximate Pareto front are computed. The maximum value, the median value, the minimum value, and the interquartile range are displayed in box plot. The box plots and the average computation times associated with the mechanisms of NSGA-II, SMS-EMOA, and IBEA are reported in Figures 10, 11, and 12, respectively.

The results show that the three methods obtain similar solutions regarding the hypervolume values. Each method has advantages in different instances. However, M2 spends more computation time for the large instances, especially, when using the elite set selection method borrowed from IBEA. Furthermore, M3 spends more computation time for the small instances, particularly when using the elite set selection method borrowed from NSGA-II or SMS-EMOA. Therefore, in the sequel only method M1 will be kept to compare the results between the hybrid decoding (HD-M1) and the two single ones (D1 and D2).

Secondly, the three decoding methods (D1, D2, and HD) are tested and the obtained results are compared. The box plots from the three elite set selection methods, which borrowed

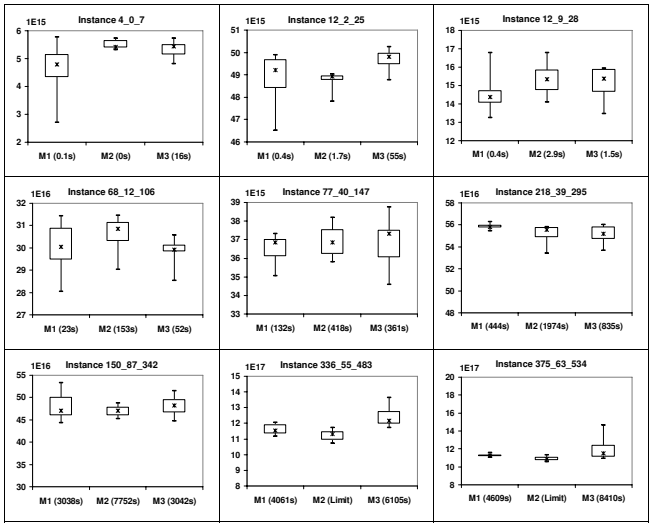


Fig. 12. Comparison of the results of the three methods for management of the elite set for hybrid decoding (M1, M2, and M3) by using the method for elite set selection borrowed from IBEA

from NSGA-II, SMS-EMOA, and IBEA, are reported in Figures 13, 14, and 15, respectively. The graph illustrates the box plots of the hypervolume values, and the average computation times are presented.

Most of the results show that the hybrid decoding obtains the results close to the best ones, when comparing the two single decodings. Indeed, it can preserve the advantages of the two single decodings for all instances. For example, in instance `12_2_26`, the first decoding method obtains better results than the second one, thus the hybrid decoding obtains results similar to the first one. For instance `77_40_147`, the hybrid decoding obtains results similar to the second decoding,

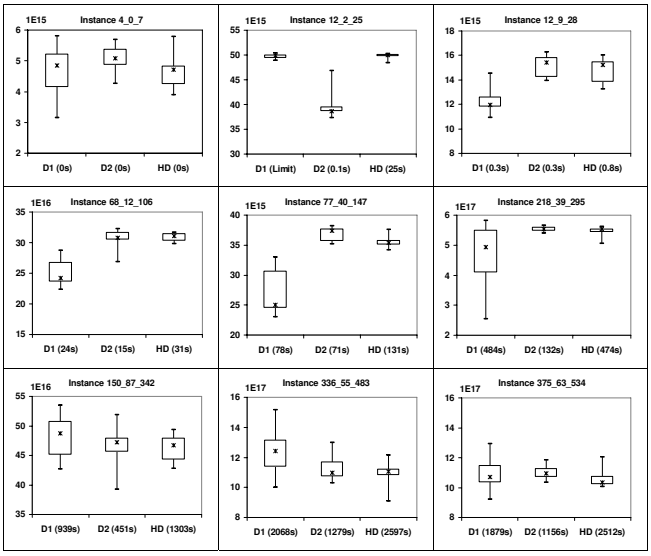


Fig. 13. Comparison of the results of the three decoding methods (D1, D2, and HD-M1) by using the method for elite set selection borrowed from NSGA-II

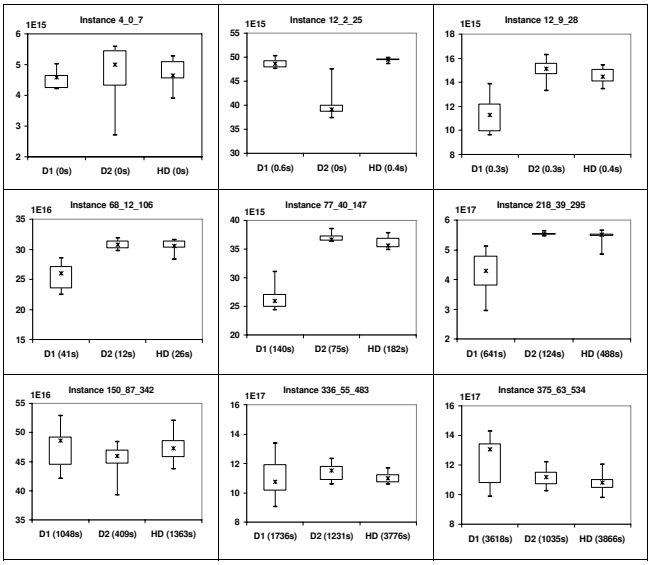


Fig. 14. Comparison of the results of the three decoding methods (D1, D2, and HD-M1) by using the method for elite set selection borrowed from SMS-EMOA

which obtains better results than the first one. Thus, the hybrid decoding method is efficient for solving most of the instances. Compared with D1, it can reduce the range of hypervolume values. This means that the hybrid decoding can provide results with better standard deviations. Moreover, for some instances where the second decoding entraps in local optima, the hybrid decoding is able to reach the global optimum. Regarding the computation time, the hybrid decoding method spends longer time in each iteration, however it can obtain good solutions in a reasonable computation time, which is limited by the second stopping criterion of BRKGA process.

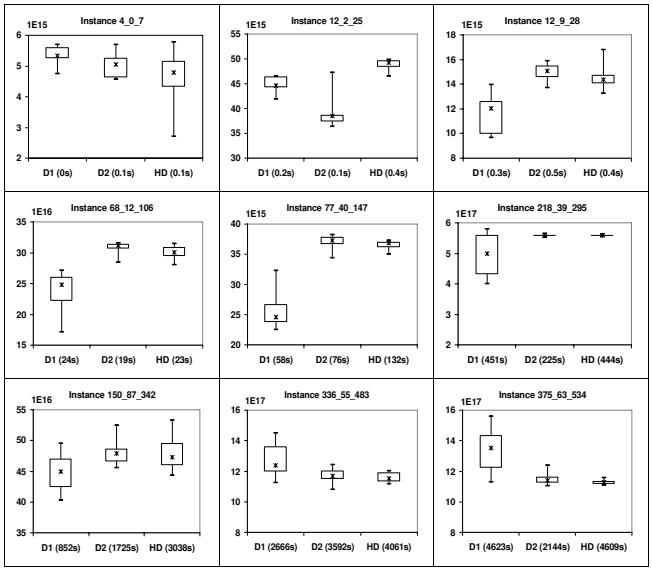


Fig. 15. Comparison of the results of the three decoding methods (D1, D2, and HD-M1) by using the method for elite set selection borrowed from IBEA

V. Conclusions

A biased random-key genetic algorithm or BRKGA is used for solving a multi-objective optimization problem. The BRKGA works on a chromosome encoded as a key vector. The chromosome consists of several genes, which are encoded by the real values in the interval $[0, 1]$. During each iteration of BRKGA, the chromosomes are decoded to obtain the feasible solutions. A hybrid decoding, which combines two single decodings, is proposed in this paper. Two solutions are obtained from the decoding of one chromosome, when using the hybrid decoding. Thus, the methods for elite set management have to be defined and three methods are tested.

The experiments are conducted on the multi-user observation scheduling problem for agile Earth observing satellites. The requests are required from multiple users. The objectives of this problem are to maximize the total profit and simultaneously minimize the maximum profit difference between users for ensure the sharing fairness. Three elite selecting methods, which are borrowed from NSGA-II, SMS-EMOA, and IBEA, are used for selecting a set of preferred solutions to become the elite set of the population. For the three elite selecting methods, the hypervolume values, which are obtained from two single decodings and the hybrid decoding, are compared. The hybrid decoding can preserve the advantages of the two single decodings, since it obtains results close to the best results of the two single decodings in reasonable computation times. Moreover, it can improve the standard deviation of the hypervolume values and avoid to entrap in local optima. Finally, the hybrid decoding is proper to be applied in BRKGA process for solving multi-objective optimization problems, which need several feasible solutions on the Pareto front.

REFERENCES

[1] J. F. Gonçalves and J. Almeida, "A Hybrid Genetic Algorithm for Assembly Line Balancing," *Journal of Heuristics,* vol. 8, 2002, pp. 629–642.

[2] J. C. Bean, "Genetic Algorithms and Random Keys for Sequencing and Optimization," *ORSA Journal on Computing,* vol. 6, 1994, pp. 154–160.

[3] J. F. Gonçalves, M. G. C. Resende, "Biased Random-key Genetic Algorithms for Combinatorial Optimization," *Journal of Heuristics,* vol. 17, no. 5, 2011, pp. 487–525.

[4] N. Goulart, S. R. de Souza, L. G. S. Dias, T. F. Noronha, "Biased Random-key Genetic Algorithm for Fiber Installation in Optical Network Optimization," in *IEEE Congress on Evolutionary Computation (CEC 2011),* pp. 2267–2271, June 2011.

[5] J. J. M. Mendes, J. F. Gonçalves, M. G. C. Resende, "A Random Key Based Genetic Algorithm for the Resource Constrained Project Scheduling Problem," *Computers and Operations Research,* vol. 36, no. 1, 2009, pp. 92–109.

[6] K. Deb, A. Pratep, S. Agarwal, T. Meyarivan, "A Fast and Elite Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation,* vol. 6, no. 2, 2002, pp. 182–197.

[7] N. Beume, B. Naujoks, M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *European Journal of Operational Research,* vol. 181, no. 3, 2007, pp. 1653–1669.

[8] E. Zitzler, S. Künzli, "Indicator-Based Selection in Multiobjective Search," in *Parallel Problem Solving from Nature (PPSN VIII)* X. Yao et al., Ed., LNCS, vol. 3242. Springer, pp. 832–842, September 2004.

[9] P. Tangpattanakul, N. Jozefowiez, P. Lopez, "Multi-objective Optimization for Selecting and Scheduling Observations by Agile Earth Observing Satellites," in *Parallel Problem Solving from Nature (PPSN XII)* C. A. Coello Coello et al., Ed., LNCS, vol. 7492. Springer, pp. 112–121, September 2012.

[10] G. Verfaillie, M. Lemaître, N. Bataille, J. M. Lachiver, "Management of the Mission of Earth Observation Satellites Challenge Description," *Technical report, Centre National d'Etudes Spatiales,* France, 2002.

[11] E. J. Kuipers, "An Algorithm for Selecting and Timetabling Requests for an Earth Observation Satellite," *Bulletin de la Société Française de Recherche Opérationnelle et d'Aide à la Décision,* 2003, pp. 7–10.

[12] J. F. Cordeau, G. Laporte, "Maximizing the Value of an Earth Observation Satellite Orbit," *Journal of the Operational Research Society,* vol. 56. no. 8, 2005, pp. 962–968.

[13] N. Bianchessi, J. F. Cordeau, J. Desrosiers, G. Laporte, V. Raymond, "A Heuristic for the Multi-satellite, Multi-orbit and Multi-user Management of Earth Observation Satellites," *European Journal of Operational Research,* vol. 177, no. 2, 2007, pp. 750–762.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 401–408

Efficient and Scalable Computation of the Energy and Makespan Pareto Front for Heterogeneous Computing Systems

Kyle M. Tarplee*, Ryan Friese*, Anthony A. Maciejewski*, and Howard Jay Siegel*†
*Electrical and Computer Engineering Department
†Computer Science Department
Colorado State University
Fort Collins, CO 80523
Email: {kyle.tarplee, ryan.friese, aam, hj}@colostate.edu

***Abstract*—The rising costs and demand of electricity for high-performance computing systems pose difficult challenges to system administrators that are trying to simultaneously reduce operating costs and offer state-of-the-art performance. However, system performance and energy consumption are often conflicting objectives. Algorithms are necessary to help system administrators gain insight into this energy/performance trade-off. Through the use of intelligent resource allocation techniques, system administrators can examine this tradeoff space to quantify how much a given performance level will cost in electricity, or see what kind of performance can be expected when given an energy budget. A novel algorithm is presented that efficiently computes tight lower bounds and high quality solutions for energy and makespan. These solutions are used to bound the Pareto front to easily trade-off energy and performance. These new algorithms are shown to be highly scalable in terms of solution quality and computation time compared to existing algorithms.**

I. Introduction

THE race for increased performance in high-performance computing (HPC) systems has resulted in a large increase in the power consumption of these systems [1]. This increase in power consumption can cause degradation in the electrical infrastructure that supports these facilities, as well as increase electricity costs for the operators [2]. The goals of HPC users conflict with the HPC operators in that the users' goal is to finish their workload as quickly as possible. That is, the small energy consumption desired by the system operator and the high system performance desired by the users are conflicting objectives that require the sacrifice of one to improve the other. Balancing the performance needs of the users with energy costs proves difficult without tools designed to help a system administrator choose from among a set of solutions.

A set of efficient and scalable algorithms are proposed that can help system administrators quickly gain insight into the energy and performance trade-off of their HPC systems through the use of intelligent resource allocation. The algorithms proposed have very desirable run times and produce schedules that are closer to optimal as the problem size increases. As such, this approach is very well suited to large scale HPC systems.

The focus of our work is on a common scheduling problem where the users submit a set of independent tasks known as a *bag of tasks* [3]. The tasks will run on a dedicated set of interconnected machines. A task runs on only one machine and, likewise, a machine may only process one task at any one time. This class of scheduling problems is often referred to as *static scheduling* because the full bag of tasks is known a priori [4]. Task execution and power consumption are deterministic in this model. The HPC systems of primary interest have highly heterogeneous task run times, machines, and power consumption which are known as *heterogeneous computing* (HC) systems. Some machines in the HC systems are often special purpose machines that can perform specific tasks quickly, while other tasks might not be able to run at all on that hardware. Another cause of heterogeneity is differing computational requirements, input/output bottlenecks, or memory limitations, and therefore cannot take full advantage of the machine. The machines may further differ in the average power consumed for each task type. Machines may have different architectures, leading to vastly different power consumption characteristics. For instance, a task that runs on a GPU might consume less energy to complete, but often more power, than the same task run on a general purpose machine, due to the shorter execution time. We assume one objective is to minimize the maximum finishing time of all tasks, which is known as the *makespan*. The heterogeneity in execution time of the tasks provides the scheduler degrees of freedom to greatly improve the makespan over a naïve scheduling algorithm. Similarly the heterogeneity in the power consumption allows the schedulers to decrease the energy consumption.

The contributions of this paper are:

1) The formulation of an algorithm that efficiently computes tight lower bounds on the energy and makespan and quickly recovers near optimal feasible solutions.
2) Finding a high quality bi-objective Pareto front.
3) An evaluation of the scaling properties of the proposed algorithms.

978-83-60810-53-8/$25.00

4) The addition of idle power consumption to the formulation of the energy/makespan problem in [3].

The rest of this paper is as follows: first the lower bound on the objectives is described in Subsection II-B. Then algorithms are presented in Subsections II-C, II-D, and II-E that reconstruct a feasible schedule from the lower bound. In Subsection II-F, the complexity of the algorithm is analyzed. Algorithm scaling quality and runtime results are shown in Section III. Section IV shows how these bounds can be used with any scalarization technique to form a Pareto front. Section V compares these algorithms to the NSGA-II algorithm.

II. Approximation Algorithms

A. Approach

The fundamental approach of this paper is to apply *divisible load theory* (DLT) [5] to ease the computational requirements of computing a lower bound solution on the energy and makespan. For the lower bound, a single task is allowed to be divided and scheduled onto any number of machines. After the lower bound on the energy and makespan is computed, a two phase algorithm is used to recover a feasible solution from the infeasible lower bound solution. The feasible solution serves as the upper bound on the optimal energy and makespan.

Often HC systems have groups of machines, usually purchased at the same time, that have identical or nearly identical performance and power characteristics. Even when every machine is different, the uncertainty in the system often allows one to model similar machines as groups of machines of the same type. A *machine group* is a collection of machines that have virtually indistinguishable performance and power properties with respect to the workload. Machines within a machine group may differ vastly in feature sets so long as the task performance and power consumption of the tasks under consideration are not affected. Tasks often exhibit natural groupings as well. Tasks of the same *task type* are often submitted many times to perform statistical simulations and other repetitive jobs. In fact, having groupings for tasks and for machines permits less profiling effort to estimate the run time and power consumption for each task on each machine.

Traditionally this static scheduling problem is posed as assigning all tasks to all machines. The classic formulation is not well suited for recovering a high quality feasible solution. The decision variables would be binary valued (assigned or not assigned) and rounding a real value from the lower bound to a binary value can change the objective significantly. Complicated rounding schemes are necessary to iteratively compute a suitable solution. Instead, the problem is posed as determining the number of tasks of *each type* to assign to each *machine group*. With this modification, decision variables will be large integers $\gg 1$, resulting in only a small error to the objective function when rounding to the nearest integer. This approximation holds well when the number of tasks assigned to each machine group is large. For this approximation, machine groups need not be large. In addition to easing the recovery of the integer solution, another benefit of this formulation is that it is much less computationally intensive due to solving the higher level assignment of tasks types to machine groups with DLT, before solving the fine grain assignment of individual tasks to machines. As such, this approach can be thought of as a hierarchical solution to the static scheduling problem.

B. Lower Bound

The lower bound is given by the solution to a linear bi-objective optimization (a.k.a. vector optimization) problem and is constructed as follows. Let there be T task types and M machine types. Let T_i be the number of tasks of type i and M_j be the number of machines of type j. Let x_{ij} be the number of tasks of type i assigned to machine group j, where x_{ij} is the primary decision variable in the optimization problem. Let $\boldsymbol{ETC}$ be a $T \times M$ matrix where ETC_{ij} is the estimated time to compute for task type i on machine type j. Similarly let $\boldsymbol{APC}$ be a $T \times M$ matrix where APC_{ij} is the average power consumption for task type i on machine type j. These matrices are frequently used in scheduling algorithms [4], [6]–[8].

The lower bound of the finishing time of a machine group is found by allowing tasks to be divided among all machines to ensure the minimal finishing time. With this conservative approximation all tasks in machine group j finish at the same time, namely:

$$F_j = \frac{1}{M_j} \sum_i x_{ij} \mathrm{ETC}_{ij} \,. \tag{1}$$

Sums over i always go from 1 to T and sums over j always go from 1 to M, thus the ranges are omitted.

Given that F_j is a lower bound on the finishing time for a machine group, the tightest lower bound on the makespan is:

$$\mathrm{MS}_{\mathrm{LB}} = \max_j F_j \,. \tag{2}$$

Energy consumed by the bag of tasks is $\sum_i \sum_j x_{ij} \mathrm{APC}_{ij} \mathrm{ETC}_{ij}$. To incorporate idle power consumption, one must have a time duration for machines not running any tasks. In this model, the makespan is used for the time the machines' power must be accumulated. Not all machines will finish executing tasks at the same time. All but the last machine(s) to finish will accumulate idle power. When no task is executing on machine j, the power consumption is given by the idle power consumption, $APC_{\emptyset j}$. The equation for the lower bound on the energy consumed, incorporating idle power, is given in:

$$\begin{aligned} E_{\mathrm{LB}} &= \sum_i \sum_j x_{ij} \mathrm{APC}_{ij} \mathrm{ETC}_{ij} \\ &\quad + \sum_j M_j \mathrm{APC}_{\emptyset j} (\mathrm{MS}_{\mathrm{LB}} - F_j) \\ &= \sum_i \sum_j x_{ij} \mathrm{ETC}_{ij} \left(\mathrm{APC}_{ij} - \mathrm{APC}_{\emptyset j} \right) \\ &\quad + \sum_j M_j \mathrm{APC}_{\emptyset j} \mathrm{MS}_{\mathrm{LB}} \end{aligned} \tag{3}$$

where the second term in the first equation accounts for the idle power.

The resulting bi-objective optimization problem for the lower bound is:

$$\begin{array}{lll} \underset{x_{ij},\,\mathrm{MS_{LB}}}{\text{minimize}} & \begin{pmatrix} E_{\mathrm{LB}} \\ \mathrm{MS_{LB}} \end{pmatrix} & \\ \text{subject to:} & \forall i & \sum_j x_{ij} = T_i \\ & \forall j & F_j \leq \mathrm{MS_{LB}} \\ & \forall i, j & x_{ij} \geq 0 \end{array} \quad . \tag{4}$$

The objective of (4) is to minimize E_{LB} and $\mathrm{MS_{LB}}$, where $\boldsymbol{x}$ is the primary decision variable. $\mathrm{MS_{LB}}$ is an auxiliary decision variable necessary to model the objective function in (2). The first constraint ensures that all tasks in the bag are assigned to a machine group. The second constraint is the makespan constraint. Because the objective is to minimize makespan, the $\mathrm{MS_{LB}}$ variable will be equal to the maximum finishing time of all the machine groups. The third constraint ensures that there are no negative assignments in the solutions. This vector optimization problem can be solved to find a collection of optimal solutions. It is often solved by weighting the objective functions to form a *linear programming* (LP) problem. Methods to find a collection of solutions are presented in Section IV.

Ideally this vector optimization problem would be solved optimally with $x_{ij} \in \mathbb{Z}_{\geq 0}$. However, for practical scheduling problems, finding the optimal integral solution is often not possible due to the high computational cost. Fortunately, efficient algorithms to produce high quality sub-optimal solutions exist.

C. Allocation Reconstruction

For infeasible solutions obtained from (4), an algorithm is necessary to recover from each a feasible solution or full allocation. Numerous approaches have been proposed in the literature for solving integer LP problems by first relaxing them to real-valued LP problems [9]. The approach here follows this common technique combined with computationally inexpensive techniques tailored to this particular optimization problem. The problem is broken up into two phases. The first phase rounds the solution while taking care to maintain feasibility of (4). The second phase assigns tasks to actual machines to build the full task allocation.

D. Rounding

Due to the nature of the problem, the optimal solution $\boldsymbol{x}^*$ often has few nonzero elements per row. Usually all the tasks of one type will be assigned to a small number of machine groups. In the original problem, tasks are not divisible so one needs to have an integer number of tasks to assign to a machine group. When all the tasks of a given task type are assigned to one machine group, that row of $\boldsymbol{x}$ has one nonzero value which is equal to T_i, an integer. When tasks are split between machine groups, an algorithm is needed to compute an integer solution from this real-valued solution. The following algorithm finds $\hat{x}_{ij} \in \mathbb{Z}_{\geq 0}$ such that it is near x^*_{ij} while maintaining the task assignment constraint. Algorithm 1 finds $\hat{\boldsymbol{x}}$ that minimizes $\| \hat{x}_{ij} - x^*_{ij} \|_1$ for a given i.

Algorithm 1 Round to the nearest integer solution while maintaining the constraints

for $i = 1$ to T **do**
  $n \leftarrow T_i - \sum_j \lfloor x^*_{ij} \rfloor$
  $f_j \leftarrow x^*_{ij} - \lfloor x^*_{ij} \rfloor$
  Let set K be the indices of the n largest f_j
  if $j \in K$ **then**
    $\hat{x}_{ij} \leftarrow \lceil x^*_{ij} \rceil$
  else
    $\hat{x}_{ij} \leftarrow \lfloor x^*_{ij} \rfloor$
  end if
end for

Algorithm 1 operates on each row of $\boldsymbol{x}^*$ independently. The variable n is the number of assignments in a row that must be rounded up to satisfy the task assignment constraint. Let f_j be the fractional part of the number of tasks that must be assigned to machine j. The algorithm simply rounds up those n assignments that have the largest fractional parts. Everything else is rounded down. The result is an integer solution $\hat{\boldsymbol{x}}$ that still assigns all tasks properly and is near to the original solution from the lower bound.

E. Local Assignment

The last phase in recovering a feasible assignment solution is to schedule the tasks already assigned to each machine group to actual machines within that group. This scheduling problem is much easier than the general case because all machines in a group are the same. This problem is formally known as the multiprocessor scheduling problem [10]. One must schedule a set of heterogeneous tasks on a set of identical machines. The *longest processing time* (LPT) algorithm is a very common algorithm for solving the multiprocessor scheduling problem [10]. Algorithm 2 uses the LPT algorithm to independently schedule each machine group.

Algorithm 2 Assign tasks to machines using LPT per machine group

for $j = 1$ to M **do**
  Let z be an empty list
  for $i = 1$ to T **do**
    $z \leftarrow$ join(z, (task type i replicated $\hat{x}_{ij}$ times))
  end for
  $y \leftarrow \mathrm{sort}_{\text{descending by ETC}}(z)$
  for $k = 1$ to $\| y \|$ **do**
    Assign task y_k to earliest ready time machine in group j
    Update ready time
  end for
end for

Each column of $\hat{\boldsymbol{x}}$ is processed independently. List z contains task type i, $\hat{x}_{ij}$ times. The tasks are then sorted

in descending order by execution time to find y. Next the algorithm loops over y one element (task) at a time and assigns it to the machine that has the earliest ready time. The *ready time* of a machine is the time at which all tasks assigned to it will complete. This heuristic packs the largest tasks first in a greedy manner. Algorithms exist that will produce a more optimal solution, but it will be shown that the effect of the sub-optimality of this algorithm on the overall performance of the systems considered is insignificant.

F. Complexity Analysis

The complexity analysis of this algorithm shows some desirable properties that are now discussed. One must solve a real-valued LP problem to compute the lower bound. Using the simplex algorithm to solve the LP problem yields exponential complexity (i.e. traversing all the vertices of the polytope) in the worst case; however the average case complexity for a very large class of problems is polynomial time. Recall that there are T task types and M machine types. The lower bound LP problem has $T + M$ nontrivial constraints and $TM + 1$ variables. The average case complexity of computing the lower bound is $(T+M)^2(TM+1)$. Next is the rounding algorithm. The outer loop iterates T times, and the rounding is dominated by the sorting of M items. Thus the complexity of Algorithm 1 is $T(M \log M)$. The task assignment algorithm outer loop is run M times. Inside this loop there are two steps. The first step is sorting $n_j = \sum_i x_{ij}$ items which takes $n_j \log n_j$ time. The second step is a loop that iterates n_j times and must find the machine with the earliest ready time each iteration, which is a $\log M_j$ time operation. The worst case complexity of Algorithm 2 is thus $M \max_j (n_j \log n_j + n_j \log M_j)$.

The complexity of the overall algorithm to find both the lower bound and upper bound (full allocation) is driven by either the lower bound algorithm or the local assignment algorithm. Complexity of the lower bound and Algorithm 1 are independent of the number of tasks and machines. Those algorithms depend only on the number of task types and machine types. This is a very important property for large scale systems. Millions of tasks and machines can be handled easily so long as the machines can be reasonably placed in a small number of homogeneous groups and, likewise, tasks can be grouped by type. Only the upper bound's complexity has a dependence on the number of tasks and machines. This phase is only necessary if a full allocation or schedule is required. Furthermore, Algorithm 2 can be trivially parallelized because each machine group is scheduled independently. The lower bound can be used to analyze much of the behavior of the system at less computational cost.

III. Scaling Results

An important property of a scheduling algorithm is its ability to scale well as the size of the problem grows. Simulation experiments were carried out to quantify how the relative error and the computational cost of the algorithm scales. These experiments are used to validate the complexity analysis results from Section II-F. **ETC** and **APC** are needed to test the algorithms. **ETC** and **APC** are based on a set of five benchmarks executed over nine machine types [11]. Then the method found in [7] was used to construct larger **ETC** and **APC** matrices. Nominally there are 1100 tasks made up of 30 task types. There are 36 machines made up of nine machine types. A complete description of the systems and output from the algorithms are available in [12].

The number of tasks, task types, and machine types are swept independently to generate a family of figures. For this size system it is too expensive to solve for the optimal makespan but one can compare bounds on the makespan to gain insight into the algorithm. Each of the three parameter sweeps is computed by taking random subsets with replacement to handle the sweep variable. These results are averaged over 50 Monte Carlo trials. The experiments where performed on a mid-2009 MacBook Pro with a 2.5 GHz Intel Core 2 Duo processor. The code is written in Mathematica 9 and the LP solver uses the simplex method which forwards to the C++ COIN-OR CLP solver [13]. The scaling experiments all optimize makespan while ignoring the energy objective.

Fig. 1 shows the relative change in makespan as the number of tasks increase. The number of task types, machines, and machine types are held constant and are the same as the original nine machine system. The relative increase in makespan is shown from the lower bound ($\mathrm{MS}_{\mathrm{LB}}$) to the makespan after rounding. Also shown is the increase in makespan from the integer solution to the full allocation. The relative increase in makespan from the lower bound to the upper bound or full allocation is also shown. The loss in quality of the makespan from the rounding algorithm is relatively low. Most of the increase in makespan is caused by Algorithm 2. However, Fig. 1 also shows that the relative increase in makespan diminishes as the number of tasks increase. This is because the approximation that tasks are divisible has less of an impact on the solution as the number of tasks increase.

The run time of the scheduler as a function of the number of tasks is shown in Fig. 2 to quantify computational efficiency of the various algorithms. The blue (bottom) portion of the graph is the time taken to compute the lower bound (solve the LP problem). The green (middle) portion is the time it takes to round the solution. Both of the computations required to compute the lower bound and the integer solution do not depend on the number of tasks. This corresponds to the results derived for the complexity of the algorithm. The red (top) portion of the figure shows the full allocation that seems to scale linearly with the number of tasks. Recall that the complexity of Algorithm 2 has a dependency on the number of tasks which is linear or log linear depending on the parameters.

Fig. 3 shows the same three curves as Fig. 1, however this time varying the number of task types. The number of tasks, machines, and machine types are held constant for this experiment. Fig. 3 shows that again the local assignment algorithm is causing most of the degradation in makespan. The relative error in makespan does not tend to zero because increasing the number of task types does not improve the quality of the approximation.

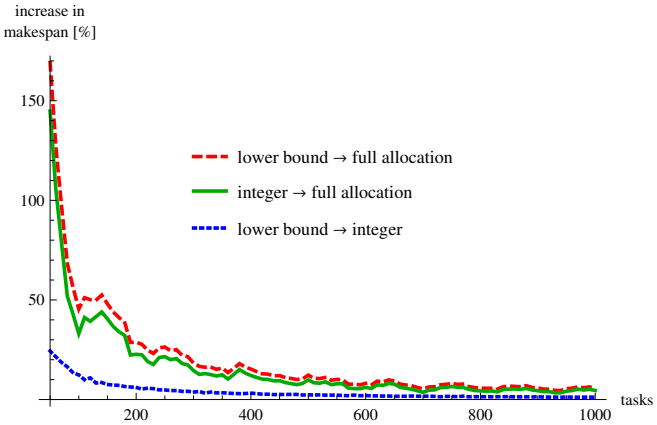


Fig. 1. Relative percent increase in makespan as a function of the *total number of tasks*: The quality of the solution improves as more tasks are used.

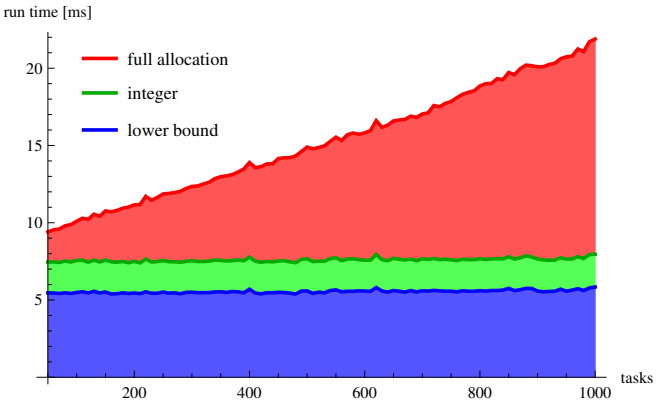


Fig. 2. Algorithm run time versus *total number of tasks*: Both the lower bound and the rounding algorithms are independent of the number of tasks. The local assignment, used to obtain the full allocation, is linearly dependent on the number of tasks.

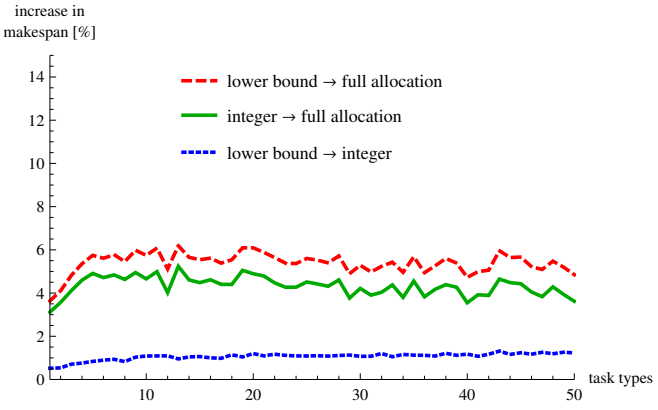


Fig. 3. Relative percent increase in makespan as a function of the *number of task types*: Quality of the solutions are roughly independent of the number of task types.

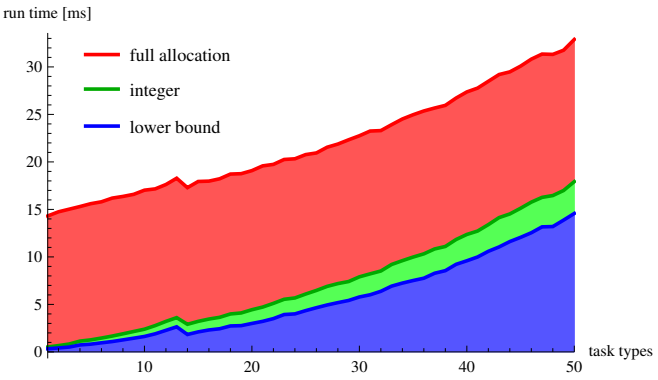


Fig. 4. Algorithm run time as a function of *number of task types*: The complexity of the lower bound and rounding algorithms grows super linearly with the number of task types.

Fig. 4 shows the run time of the three phases. Here the lower bound has super linear dependence on the number of task types. According to the complexity analysis this should be cubic. The rounding algorithm seems to increase linearly, which corresponds to the analysis. The full allocation phase seems to be independent of the number of task types. This agrees with the analysis because the complexity is not a function of the number of task types T, but instead a function of the number of tasks n_j assigned to a group, regardless of the type of task.

Fig. 5 shows the relative increase in makespan as the number of machine types varies. In the previous parameter sweeps, the number of tasks of a particular type may be zero if the random sampling selected that configuration. Allowing the number of machines in a machine group to be zero is more difficult due to (1) because some constraint coefficients will be ∞ in the linear programming problem. Practically, an $M_j = 0$ means that the j^{th} column of **ETC** and **APC** should simply be removed and the solution should never assign a task to that group because it has no machines. To avoid this case altogether each machine group has to have at least one machine (so that there are no degenerate machine groups). Fig. 5 also shows that the quality of the rounding algorithm decreases as the number of machine groups increase. This is expected because there are less tasks to assign to each machine's group making the approximation weaker. At 36 machine types there is one machine per group. There is only one solution to that scheduling problem (assign all tasks to the one machine), resulting in no increase in makespan in that phase.

Fig. 6 shows the run time as the number of machine types is increased. As expected, the lower bound grows roughly cubicly. The rounding algorithm grows roughly linearly also as expected. The time spent scheduling for each group decreases because fewer tasks are scheduled to less machines as the number of machine types increases so it effectively has little dependence on the number of machine types.

Even though the performance of these polynomial time algorithms are desirable, there is some prior work on theoretical bounds that should be noted. In [14] it is proven that there exists no polynomial algorithm that can provably find a schedule that is less than 3/2 the optimal makespan, unless

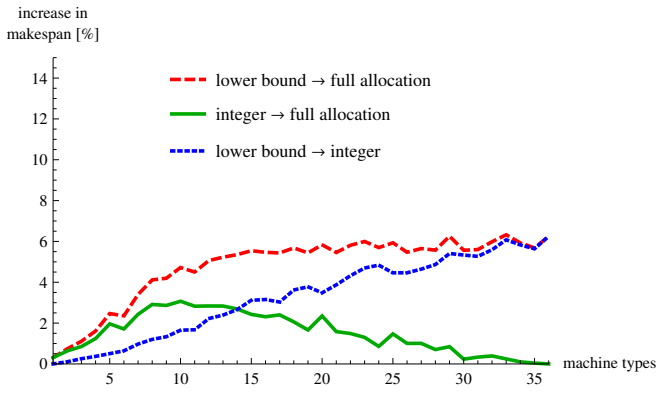


Fig. 5. Relative percent increase in makespan as a function of the *number of machine types*: Overall performance is roughly independent of the number of machine types.

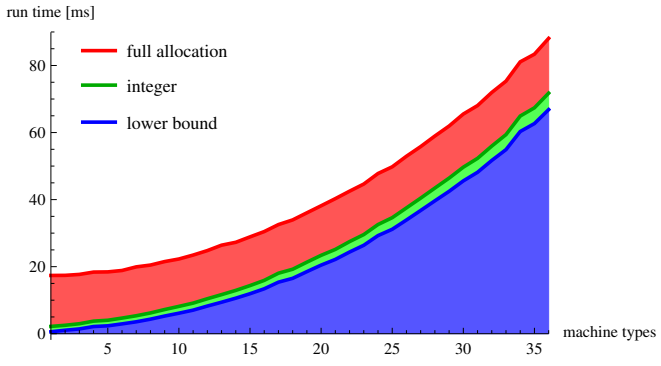


Fig. 6. Algorithm run time versus the *number of machine types*: Lower bound algorithm complexity is super linear in the number of machines types. The rounding and local assignment algorithms are roughly independent.

$P = NP$. Even though Figures 1-6 suggest that one can do better then 3/2, this is only the case on average. In the next section these algorithms are used to generate the Pareto fronts.

IV. Pareto Front Generation

Multi-objective optimization is challenging because there is usually no single solution that is superior to all others. Instead, there is a set of superior feasible solutions that are referred to as the non-dominated solutions [15]. Feasible solutions that are dominated are of little interest because one can always find a better solution in all objectives by picking a solution from the non-dominated set. Formally, a feasible solution x dominates a feasible solution y when:

$$\begin{aligned} \forall i \quad & f_i(x) \leq f_i(y) \\ \exists i \quad & f_i(x) < f_i(y) \end{aligned} \tag{5}$$

where $f_i(\cdot)$ is the i^{th} objective function. The non-dominated solutions, also known as *outcomes*, compose the Pareto front.

Finding the Pareto front can be computationally expensive because it means solving variations of the optimization problem numerous times. Most algorithms use scalarization techniques to convert the multi-objective problem into a set of scalar optimization problems. Major approaches to scalarization include the hybrid method [16], elastic constraint method [16], Benson's algorithm [17] [18], and Pascoletti-Serafini scalarization [19]. Pascoletti-Serafini scalarization is a generalization of many common approaches such as normal boundary intersection, ϵ-constraint, and weighted sum. The focus of our work is on the weighted sum algorithm. The weighted sum algorithm can find all the non-dominated solutions for problems with a convex constraint set and convex objective functions [19]. Weighted sum is used for the linear convex problem in (4) thus all non-dominated solutions can be found. A known issue with the weighted sum algorithm is that it does not uniformly distribute the solutions along the Pareto front. The solutions are often clustered together, but this does not present a problem for our particular use case.

Finding the optimal schedule for makespan alone is NP-Hard in general [4], thus finding the optimal (true) Pareto front is NP-Hard as well. Computing tight upper and lower bounds on the Pareto front is still possible. Specifically, a lower bound on a Pareto front is a set of solutions for which no feasible solution dominates any of the solutions in this set. An upper bound on the Pareto front is a set of feasible solutions which do not dominate any Pareto optimal solutions.

A. Weighted Sum

The weighted sum algorithm simply forms the positive convex combination of the objectives and sweeps the weights to generate the Pareto front. The first phase is to compute the lower bound solution for energy and makespan independently of each other. Next ΔE_{LB}, which is the difference between the maximum and minimum values of energy, is computed. Likewise, $\Delta \text{MS}_{\text{LB}}$ is computed. The scalarized objective is:

$$\min \frac{\alpha}{\Delta E_{\text{LB}}} E_{\text{LB}} + \frac{1-\alpha}{\Delta \text{MS}_{\text{LB}}} \text{MS}_{\text{LB}} \,. \tag{6}$$

A lower bound on the Pareto front can be generated by using several values of $\alpha \in [0, 1]$. Weighted sums will produce duplicate solutions (i.e., $\boldsymbol{x}$ is identical for neighboring values of α). Duplicate solutions are removed to increase the efficiency of the subsequent algorithms. Each solution is rounded by Algorithm 1 to generate an intermediate Pareto front. Rounding often introduces many duplicates that can be safely removed. Each integer solution is converted to a full allocation with Algorithm 2 to create the upper bound on the Pareto front.

B. Non-dominated Sorting Genetic Algorithm II

NSGA-II [20] is an adaptation of the Genetic Algorithm (GA) optimized to find the Pareto front of a multi-objective optimization problem. Similar to all GAs, the NSGA-II uses mutation and crossover operations to evolve a population of chromosomes (solutions). Ideally this population improves from one generation to the next. Chromosomes with a low fitness are removed from the population. The NSGA-II algorithm modifies the fitness function to work well for discovering the Pareto front. In prior work [3], the mutation

and crossover operations were defined for this problem. The NSGA-II algorithm will be seeded in two ways in the following results. The first seeding method is to use the optimal minimum energy solution, sub-optimal minimum makespan solution (from the Min-Min Completion Time [4] algorithm), and a random population as the initial population. This is the original seeding method used in [3]. The second seeding method is to use the full allocations from Algorithm 2 as the initial population for the NSGA-II.

V. Pareto Front Results

The system used for these experiments is the same as in Section III, unless stated otherwise. All 1100 tasks, 30 task types, 36 machines, and nine machines types are used as described in [8]; the complete description of the system and output data files from the new algorithm are available in [12]. The hardware used for running the NSGA-II experiments is a 2011 Sager NP7280 with an Intel Core i7 980 @ 3.33 Ghz. The NSGA-II code is implemented in C++.

Fig. 7 shows the Pareto fronts generated from the various algorithms without idle power consumption. The figure shows the actual solutions as markers that are connected by lines. The lower bound, integer, and full allocation are nearly indistinguishable at the lower portion of the plot. This means that the true Pareto front is tightly bounded even though it is unknown. The curve that is dominated by all other curves is the Pareto front generated by the NSGA-II using the first seeding method. The NSGA-II took hours to find that sub-optimal Pareto front. In contrast, the lower and upper bounds were found in ~ 10 seconds. The last Pareto front is the NSGA-II seeded with the full allocation. Seeding with the full allocation allows the NSGA-II to both converge to an improved Pareto front as well as decreasing the run time necessary to converge. The NSGA-II attempts to evenly distribute the solutions along the Pareto front and tries to find solutions that are in the convex regions as can be seen in Fig. 7. All the algorithms seem to perform well when minimizing energy alone because computing the optimal minimum energy solution is relatively easy. One simply assigns each task to the machine that requires the lowest energy to execute that task. Solving for the optimal makespan is difficult in practice. Fig. 7 shows that all the algorithms agree in the energy dimension, however in the makespan dimension there are significant distinctions in solution quality. Pareto fronts for other representative systems were also computed with similar results. The new algorithms produced better quality Pareto fronts in significantly less time.

Fig. 8 illustrates how the solutions progress through the three phases of the algorithm when there is no idle power consumption. The lowest line represents the lower bound on the Pareto front. Each orange arrow represents a solution as it is rounded. In every case, the makespan increases but the energy may increase or decrease. As $\boldsymbol{x}$ is rounded, machines will finish at different times increasing the makespan. Each blue arrow represents a solution that is being fully allocated. The energy in this case does not change because the local assignment algorithm does not move tasks across

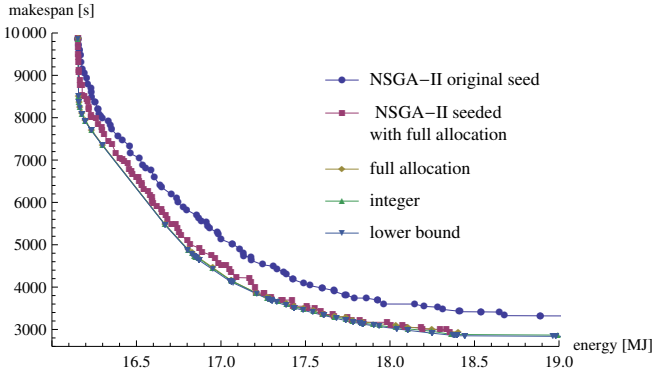


Fig. 7. Pareto front for lower bound, integer, upper bound, and NSGA-II. The lower bound does truly lower bound the other curves. The full allocation or upper bound is very near the lower bound so the optimal Pareto front is tightly bounded. The NSGA-II with the original seed solution quality is rather poor and expensive to compute, however the NSGA-II seeded with the full allocations produces a reasonable result, close to the full allocation, in much less time, but still not as good as the full allocation in places.

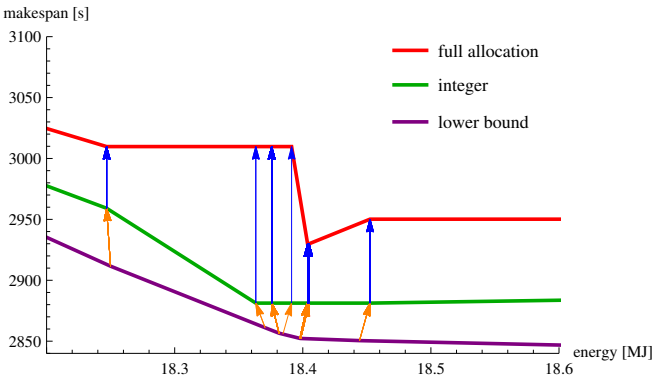


Fig. 8. Progression of solutions from lower bound to integer to upper bound (no idle power)

machine types thus the power consumption cannot change. The makespan increases are highly varying. The full allocation solution second to the right dominates the one on the far right. In this case the solution on the far right is taken out of the estimate of the Pareto front.

Fig. 9 shows the progression of the the solutions with non-zero idle power. The idle power consumption is set to 10% of the mean power for each machine type, specifically $\mathrm{APC}_{\emptyset j} = \frac{0.1}{T} \sum_i \mathrm{APC}_{ij}$. As the makespan increases, more machines will be idle for longer, so the idle energy increases. The local assignment phase now negatively affects the energy consumption because it will often have machines idle for some time.

Fig. 10 shows the effect of idle power on the Pareto front. The curves show the lower bound on the optimal Pareto front with different idle powers. The penalty for having a large makespan increases as the idle power increases. The optimal energy solutions now must have a shorter makespan to reduce energy usage. This causes the Pareto front to contract in the makespan dimension and shift to the right slightly.

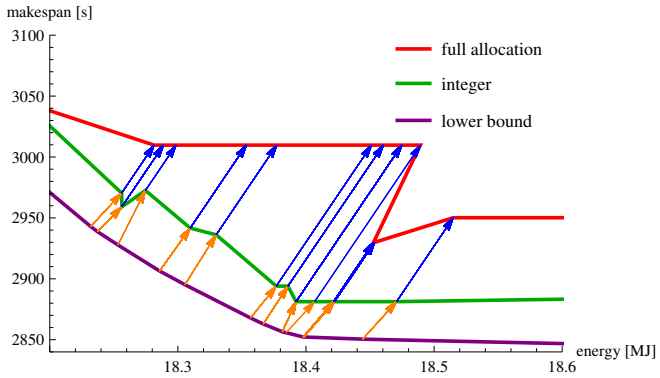


Fig. 9. Progression of solutions from lower bound to integer to upper bound (10% idle power)

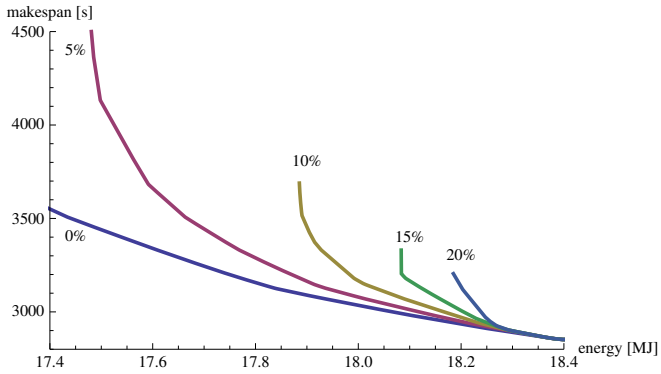


Fig. 10. Pareto front lower bounds when sweeping idle power: Idle power is swept from 5% increments as labelled on the figure. As idle power increases the reward for improving makespan also increases.

VI. Conclusions

A highly scalable scheduling algorithm for the energy and makespan bi-objective optimization problem was presented. The complexity of the algorithm to compute the lower bound on the Pareto front was shown to be independent of the number of tasks. The quality of the solution also improves as the size of the problem increases. These two properties make this algorithm perfectly suited for very large scale scheduling problems. Algorithms were also presented that allow one to efficiently recover feasible solutions. These feasible solutions serve as the upper bound on the Pareto front and can be used to seed other algorithms. This upper bound was compared to the solution found with the NSGA-II algorithm and shown to be superior in solution quality and algorithm run time. These algorithms allow the decision makers to more easily trade-off energy and makespan to reduce operating costs and improve efficiency of HPC systems.

Acknowledgment

This work was supported by the Sjostrom Family Scholarship, Numerica Corporation, the National Science Foundation (NSF) under grants CNS-0905399 and CCF-1302693, the NSF Graduate Research Fellowship, and by the Colorado State University George T. Abell Endowment. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. A special thanks to Brian Stefanović, Mark Oxley, and Tim Hansen for their valuable comments.

References

[1] J. Koomey, "Growth in data center electricity use 2005 to 2010," *Analytics Press.*, vol. 1, 2011.
[2] K. W. Cameron, "Energy oddities, part 2: Why green computing is odd," *Computer*, vol. 46, no. 3, pp. 90–93, 2013.
[3] R. Friese, T. Brinks, C. Oliver, H. J. Siegel, and A. A. Maciejewski, "Analyzing the trade-offs between minimizing makespan and minimizing energy consumption in a heterogeneous resource allocation problem," in *INFOCOMP, The Second International Conference on Advanced Communications and Computation*, 2012, pp. 81–89.
[4] T. D. Braun, H. J. Siegel, N. Beck, L. L. Bölöni, M. Maheswaran, A. I. Reuther, J. P. Robertson, M. D. Theys, B. Yao, D. Hensgen, and R. F. Freund, "A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems," *Journal of Parallel and Distributed Computing*, vol. 61, no. 6, pp. 810 – 837, 2001.
[5] V. Bharadwaj, T. G. Robertazzi, and D. Ghose, *Scheduling Divisible Loads in Parallel and Distributed Systems*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1996.
[6] A. Al-Qawasmeh, A. Maciejewski, H. Wang, J. Smith, H. Siegel, and J. Potter, "Statistical measures for quantifying task and machine heterogeneities," *The Journal of Supercomputing*, vol. 57, no. 1, pp. 34–50, 2011.
[7] R. Friese, B. Khemka, A. A. Maciejewski, H. J. Siegel, G. A. Koenig, S. Powers, M. Hilton, J. Rambharos, G. Okonski, and S. W. Poole, "An analysis framework for investigating the trade-offs between system performance and energy consumption in a heterogeneous computing environment," in *IEEE 22nd Heterogeneity in Computing Workshop (HCW)*, 2013.
[8] R. Friese, T. Brinks, C. Oliver, H. J. Siegel, A. A. Maciejewski, and S. Pasricha, "A machine-by-machine analysis of a bi-objective resource allocation problem," in *International Conference on Parallel and Distributed Processing Technologies and Applications (PDPTA)*, 2013, accepted.
[9] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, ser. Optimization and Neural Computation. Athena Scientific, 1997.
[10] R. Graham, "Bounds on multiprocessing timing anomalies," *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 416–429, 1969.
[11] (2013, May) Intel core i7 3770k power consumption, thermal. [Online]. Available: http://openbenchmarking.org/result/1204229-SU-CPUMONITO81#system_table
[12] (2013, April) Energy and makespan bi-objective optimization data. [Online]. Available: http://goo.gl/Z2Utv
[13] (2013, March) Coin-or clp. [Online]. Available: https://projects.coin-or.org/Clp
[14] J. Lenstra, D. Shmoys, and É. Tardos, "Approximation algorithms for scheduling unrelated parallel machines," *Mathematical Programming*, vol. 46, no. 1-3, pp. 259–271, 1990.
[15] V. Pareto, *Cours d'economie Politique*. Lausanne: F. Rouge, 1896.
[16] M. Ehrgott, *Multicriteria Optimization*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
[17] H. Benson, "An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem," *Journal of Global Optimization*, vol. 13, no. 1, pp. 1–24, 1998.
[18] A. Löhne, *Vector Optimization with Infimum and Supremum*, ser. Vector Optimization. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
[19] G. Eichfelder, *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer, 2008.
[20] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 409–416

Efficient Models for Special Types of Non-Linear Maximum Flow Problems

Marina Tvorogova
Department of Computer Science
TU Braunschweig, Germany
Email: m.tvorogova@tu-bs.de

***Abstract*—In this paper, we consider the maximum flow problem on networks with non-linear transfer functions. We consider special types of transfer functions, which are particularly relevant for applications. For concave transfer functions, we reduce the NL-flow problem to the generalized flow problem and solve it using a polynomial-time approximation scheme. For convex, s-shaped and monotonically growing piecewise linear (PWL) transfer functions (the latter can always be divided into s-shaped fragments), we present an equivalent network representation that allows us to build a MILP model with a better performance than if we were using standard MILP formulations of PWL functions. The latter requires additional variables and constraints to force the correct (depending on the amount of flow) linear segment of PWL functions to be taken. In our model, the correct segment in an s-shaped fragment is chosen automatically due to the network's structure. For the case when transfer functions are non-linear, we provide an error estimation for the approximated solution.**

I. Introduction

IN THIS paper, we consider flows in networks with non-linear losses (NL-flow problem). We have a directed graph $D = (V, A)$, where V is a set of vertices and A is a set of edges. The maximum flow that we can send through edge $a \in A$ is bounded by the edge's capacity $u_a \in R_+$. Each edge a of graph D has an associated non-linear function $F_a(f_a)$ that defines how outflow depends on inflow: we assume that if we send f_a units of flow into edge $a = (v, w)$, then $F_a(f_a)$ units of flow arrive at the head of edge a (that is node w).

In the classical maximum flow problem, the goal is to send as much flow as possible from the source node(s) to the target node(s), taking capacity and flow conservation constraints into account. *Transfer function* $F_a(f_a)$ defines the outgoing flow from edge a depending on the incoming flow f_a to edge a, $a \in A$. For the classical case, flow does not change while going through an edge, i.e. $F_a(f_a) = f_a$. This problem is well-studied, see e.g. Korte and Vygen [4].

The generalized flow problem (GFP) is a step closer to the NL-flow problem. The goal of the generalized maximum flow problem is to maximize flow at the target node. In a generalized graph, flow changes its value while going through edges. Outgoing flow $F_a(f_a)$ from edge a linearly depends on the incoming flow f_a to edge a, $a \in A$. Transfer functions corresponding to this type of flow are linear, i.e. $F_a(f_a) = \gamma_a \cdot f_a$, where γ_a is the proportionality coefficient corresponding to edge a. The GFP has already been studied by Onaga [6] and Truemper [9]. There are fully polynomial-time approximation schemes for GFP, see e.g. Fleischer and Wayne [2] and Oldham [5], and polynomial algorithms for GFP with assumptions about transfer coefficients, see Radzik [7] and Tardos and Wayne [8].

In this paper, we introduce flows with affine-linear transfer functions, i.e. $F_a(f_a) = \gamma_a \cdot f_a + b_a$, where k_a and b_a are constants corresponding to edge a, $a \in A$. We call optimization problems that correspond to this type of flow, *affine generalized flow problems* (AGFP).

For non-linear PWL functions, there exist standard ways to present them in mixed-integer formulations (see Vielma et al. [10]). Our representation is more efficient than representations applied to the general type of PWL functions. Standard ways use extra variables to force the use of the right segment of the PWL functions. We modify the network in such a way that by flow maximizing, the right segment will be chosen automatically. This allows us to get a problem formulation of significantly reduced size. The main contribution of this paper is that for a wide range of applications, we propose a solution, which deals with large MILP formulations arising by modeling problems with non-linearities. We propose efficient formulations of the maximum flow problem for networks with transfer functions of special types.

The remainder of this paper is organized as follows. In Section II-C, we introduce the required definitions and terms. We describe how to build a residual network for flows with NL functions and provide flow decomposition theorem for flows with NL losses. In Section III, motivated by the applications areas, we distinguish three special types of transfer functions: *concave, convex* and *s-shaped*. In Section IV, we design an equivalent problem representation of the original instance for the considered types of transfer functions. We show that the optimal solution of the maximization problem for this problem representation and the optimal solution of the maximization problem for the original instance are the same. For the problem with concave transfer functions, we propose a fully polynomial-time approximation scheme. For convex and s-shaped transfer functions we replace edges with NL transfer functions by network structures, where transfer functions of the edges are affine-linear. In Section V, we introduce MILP models for AGFP. In Section VI, we evaluate our MILP model. In section VII we consider the case of non PWL transfer functions and estimate the solution error,

978-83-60810-53-8/$25.00

which arises by approximating the original transfer functions by PWL functions. Section VIII completes our paper with the conclusions.

II. Preliminaries

A. Problem Formulation

We can formulate the maximization problem for flows with NL as follows:

Problem 1:

Given $D = (V, A)$, $u_a\ \forall a \in A$, $F_a\ \forall a \in A$.
Find an $s-t-$flow, that
maximizes
$$\sum_{a\in\delta^-(t)} F_a(f_a)$$
subject to
$$\sum_{a\in\delta^+(v)} f_a - \sum_{a\in\delta^-(v)} F_a(f_a) \leq 0, \forall v \in V\backslash\{s,t\} \quad (1)$$
$$0 \leq f_a \leq u_a, \forall a \in A. \quad (2)$$

Equation (1) describes the flow conservation law, inequality (2) the edge capacity constraints.
Here, we allow positive excess at nodes. This does not contradict flow maximization at the target node.

B. Assumptions

Inspired by the applications of the NL-flow models (see Section III for details), we make the following assumptions on the type of transfer functions. The considered transfer functions are loss functions ($\gamma_a(f_a) \leq 1$). Increasing flow f_a incoming to edge a increases the outgoing flow $F_a(f_a) = \gamma_a(f_a) \cdot f_a$. Thus, the transfer functions are strictly monotonically growing functions. The next natural assumption is $F_a(0) = 0$.

The considered networks have no flow generating-cycles. In the context of our applications, flow-generating cycles would refer to perpetual flow sources, which is not possible in practice.

C. Definitions

Definition 1: The transfer multiplier $\gamma_a(f_a)$: $R_+ \to R_+$ is a quotient of the edge's transfer function $F_a(f_a)$ and flow f_a entering the edge:

$$\gamma_a(f_a) = \frac{F_a(f_a)}{f_a}, \text{ for } f_a \neq 0.$$

If $f_a = 0$, $\gamma_a(f_a)$ may be assigned an arbitrary number. Let's assume for this paper that if $f_a = 0$, then $\gamma_a(f_a) = 0$.
$\gamma_a(f_a)$ can be interpreted as the efficiency of sending flow f_a through edge a.

Definition 2: A transfer function is called *loss function* (flow decreases) if the transfer multiplier corresponding to this function is less than or equal to one, i.e.

$$\gamma_a(f_a) \leq 1\ \forall f_a \in R_+.$$

Proposition 1: The optimization problem on a graph with multiple source nodes and multiple target nodes can always be transferred to the optimization problem on a graph with one source node and one target node.

Proof: The solution of the flow maximization problem (also for flows with NL-losses) is not influenced by the following two operations.

T is a set of the target nodes and S is a set of the source nodes.
g_i is the flow at node i. $g_i = 0$ for $g_i \in V\backslash\{S,T\}$. $g_i < 0$ for $g_i \in S$. $g_i > 0$ for $g_i \in T$.

- We can add an integrated source node s and an extra edge $a = (s, w)$ with $\gamma_a(\cdot) = 1$, $u_a = g_w$ to all nodes $w \in S$; node w becomes a transfer node, $g_s = \sum_{w\in S} g_w$ and g_w is assigned to zero.
- We can add an integrated target node t and an extra edge $a = (v, t)$ with $\gamma_a(\cdot) = 1$, $u_a = \infty$ to all nodes $v \in T$; node v becomes a transfer node, $g_t = \sum_{v\in T} g_v$ and g_v is assigned to zero.

■

Using Proposition 1 without loss of generality, we assume that the graph D has only one target node and only one source node.

Definition 3: An $s-t-$flow is a flow from the supply nodes s to target nodes t.

Definition 4: Let us call *flow at node* v the difference between flow outgoing from node v ($\sum_{a\in\delta^+(v)} f_a$) and flow incoming to node v ($\sum_{a\in\delta^-(v)} F_a(f_a)$), where $\delta^-(v) := \{(u, v) \in A\}$ is a set of edges entering node v and $\delta^+(v) := \{(v, w) \in A\}$ is a set of edges leaving node v.

Definition 5: The *affine-generalized network* is the network, in which edges have transfer functions of type $F_a(f_a) = f_a \cdot k_a + b_a$.

Definition 6: A network flow distribution $X : R_+^A \to R_+$ defines for all edges $a \in A$ *partition coefficients* $x_a : A \to [0, 1]$, which denote the ratio of the flow leaving node v and the flow sent through edge $a = (v, w)$, i.e.

$$x_a = \begin{cases} 0, \text{ if } \sum_{a\in\delta^+(v)} f_a = 0, \\ \frac{f_a}{\sum_{a\in\delta^+(v)} f_a}, \text{ otherwise.} \end{cases}$$

A feasible flow through the network f determines a flow distribution X.

Definition 7: A flow distribution X (for the given supply) is *feasible* if the corresponding flow is feasible, i.e. the flow does not violate flow feasibility constraints (1) and (2).

Lemma 1: The flow at the source node together with the feasible flow distribution X uniquely determine the flow at each edge of the given graph.

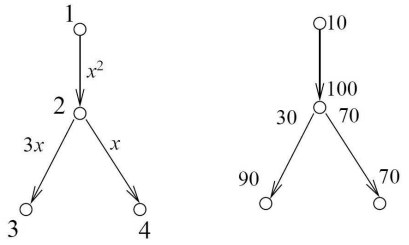


Fig. 1. Graph G and flow on it.

D. Decomposition theorem for NL flows

Here, we introduce the decomposition theorem for flows with NL transfer functions. This theorem will be used in the proof of Theorem 3.

Definition 8: We define the *residual capacity* for flow function $f_a : V \longrightarrow R_+$ as

$$u_a^R = u_a - f_a, a \in A, \tag{3}$$

$$u_{\overline{a}}^R = F_a(f_a), \overline{a} \in \overline{A}, \tag{4}$$

where $\overline{A}$ is a set of backward edges.

Definition 9: The residual function for backward edges can be found as:

$$F_{\overline{a}}^R(l) = f_a^* - F_a^{-1}(F_a(f_a^*) - l), \tag{5}$$

where f_a^* is the current flow on edge a, $l = f_{\overline{a}}$, we use notation l for better observability.

Definition 10: The residual function for forward edges can be found as:

$$F_a^R(g) = F_a(g + f_a^*) - F_a(f_a^*), \tag{6}$$

where f_a^* is the current flow on edge a, $g = f_a$, we use notation g for better observability.

Theorem 1: For every feasible flow f on a graph $D = (A, V)$, there exists a collection of $k \leq m = |A(D)|$ elementary residual flows $\mathcal{F} = \mathcal{F}(1), \ldots, \mathcal{F}(k)$ such that $f_a = \sum_{i=1..k} \mathcal{F}_a(i)$ for all $a \in A$. Elementary residual flows are residual flows on a path, on a unit-gain cycle, on a cycle-path, on a path-cycle or on a bicycle.

We omit the prove of the theorem, but give an example of decomposition.

Given graph G and flow on it, see Figure 1. We want to decompose the current flow to flow $\mathcal{F}(1)$ (through path $1-2-3$) and to flow $\mathcal{F}(2)$ (through path $1-2-4$) so that $f_a = \sum_{i=1,2} \mathcal{F}(i)$. Consider the graph and flow $f = (10, 30, 70)$ on it. Calculate the residual transfer function and the residual capacities for backward edges using formulas 4 and 5. $F_{21}^R(f) = f_{12} - \sqrt{(f_{12})^2 - f}$, where $f_{12} = 10$, $F_{32}^R(f) = \frac{1}{3}$, $F_{42}^R(f) = x$. $u_{21}^R(f) = 100$, $u_{32}^R(f) = 90$, $u_{42}^R(f) = 70$.

a is a part of the flow at the source node, that forms $\mathcal{F}(1)$, b is a part of the flow at the source node, that forms $\mathcal{F}(2)$, $a, b \in [0, 1]$, $a + b = 1$.

There are two ways to define $\mathcal{F}(1)$ and $\mathcal{F}(2)$ depending on the order of augmentation. Let us call the flow after the first augmentation $f(1)$, after the second augmentation $f(2)$.

1: We first augment $a \cdot f_{12}$ units of flow along path $1-2-3$ and then $b \cdot f_{12}$ units of flow along path $1-2-4$. Augment the flow along path $4-2-1$ by $F_{24}(f_{24})$ units. $F_{24}(f_{24}) = u_{1-2-4} = 70$. After augmenting, $f_{24}(1) = 0$, $f_{12}(1) = 10 - F_{21}^R(F_{42}^R(70)) = 10 - 10 + \sqrt{10^2 - 70} = \sqrt{30}$. $\mathcal{F}_{24}(1) = f_{24} = 70$, $\mathcal{F}_{12}(1) = f_{12} - f_{12}(1) = 10 - \sqrt{30}$.

We augment the current flow along path $3-2-1$ by $F_{23}(f_{23})$ units of flow. $F_{23}(f_{23}) = u_{1-2-3} = 90$. After augmenting, $f_{23}(2) = 0$, $f_{12}(2) = \sqrt{30} - F_{21}^R(F_{32}^R(90) = \sqrt{30} - \sqrt{30} + \sqrt{\sqrt{30}^2 - \frac{1}{3} \cdot 90} = 0$.

$\mathcal{F}_{23}(2) = f_{23} = 30$, $\mathcal{F}_{12}(2) = f_{12}(1) - f_{12}(2) = \sqrt{30}$.

Thus, we obtain $\mathcal{F}(1) = (10 - \sqrt{30}, 30, 0)$ and $\mathcal{F}(2) = (\sqrt{30}, 0, 70)$.

2: We first augment $a \cdot f_{12}$ units of flow along path $1-2-3$ and then $b \cdot f_{12}$ units of flow along path $1-2-4$. Augment the flow along path $3-2-1$ by $F_{23}(f_{23})$ units of flow. $F_{23}(f_{23}) = u_{1-2-4} = 90$. After augmenting, $f_{23}(1) = 0$, $f_{12}(1) = 10 - F_{21}^R(F_{42}^R(70)) = 10 - 10 + \sqrt{10^2 - \frac{1}{3} \cdot 90} = \sqrt{70}$. $\mathcal{F}_{24}(1) = f_{23} = 30$, $\mathcal{F}_{12}(1) = f_{12} - f_{12}(1) = 10 - \sqrt{70}$.

We augment the current flow along path $4-2-1$ by $F_{24}(f_{24})$ units of flow. $F_{24}(f_{24}) = u_{1-2-4} = 70$. After augmenting, $f_{24}(2) = 0$, $f_{12}(2) = \sqrt{70} - F_{21}^R(F_{32}^R(70) = \sqrt{70} - \sqrt{70} + \sqrt{\sqrt{70}^2 - 70} = 0$.

$\mathcal{F}_{24}(2) = f_{24} = 70$, $\mathcal{F}_{12}(2) = f_{12}(1) - f_{12}(2) = 10 - \sqrt{70}$.

Thus, we obtain $\mathcal{F}(1) = (\sqrt{70}, 0, 70)$ and $\mathcal{F}(2) = (10 - \sqrt{70}, 30, 0)$.

III. Special types of transfer functions and their applications

In this section, we consider three types of PWL transfer functions, which are especially interesting from the application point of view. We list them together with the application examples of the models, which use those types of transfer functions. Further, we use the special properties of these functions to find the solution of the maximum flow problem.

A. Convex transfer functions

In this case, the slope of the transfer function (and, thus, the transfer efficiency) grows as the amount of flow we send increases.

One application of convex transfer functions is modeling of information flows with "learning" effects. The flow passing along the graph represents information. Transfer functions model information transmission processes with "learning" effect, e.g. handwriting recognition, face recognition, speech

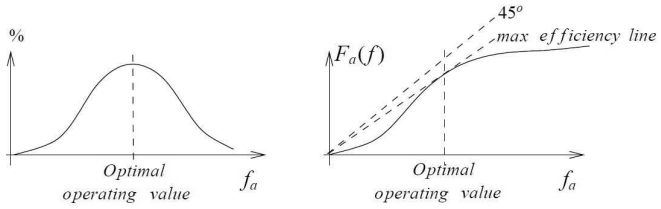


Fig. 2. Efficiency around optimal operating value and the corresponding S-shaped transfer function.

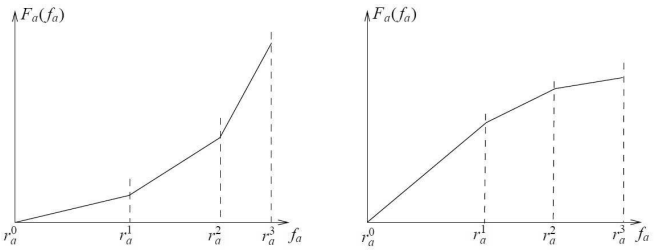


Fig. 3. PWL convex and concave functions.

recognition. Efficiency of the information recognition/transmission process increases together with the amount of information.

B. Concave transfer functions

For this case, the slope of the transfer function (and thus, the transfer efficiency) shrinks as the amount of flow we send increases.

We can use concave transfer functions, e.g., in traffic flow modeling. It is a maximum NL-flow problem to find a flow routing with minimal overall flow-losses (deceleration). Edges represent road segment. Flow corresponds to the number of vehicles passing a reference point per unit of time. Transfer functions describe the flow-deceleration effect, which increases as flow approaches the capacity of a road segment.

C. S-shaped transfer functions

S-shaped transfer functions are used to model processes with optimal operating value. The slope of the transfer function grows until it reaches the optimal operating value. As we increase the amount of flow beyond the optimal operating value, the slope decreases.

S-shaped transfer functions are interesting for modeling energy flows. Nodes represent different types of energy (e.g. raw materials, electricity or heat energy), flow is energy (in different forms), edges represent transformation of one type of energy into another. Technical equipment that enables energy transformations usually has an optimal operating value, at which this equipment reaches its maximum efficiency. Thus, transfer functions show how the efficiency of the energy transformation first grows until the optimal operating value and then shrinks (see Fig. 2).

IV. Design of equivalent problem representations

In this section, we modify the underlying network in such a way that *Problem 2* from subsection IV-C has the same optimal solution for the original (*Problem 1*) and for the modified network. Further, in the next section, we use this modified network for establishing an efficient MILP formulation.

In the procedure described below, we replace a single edge $a = (v, w)$ by a set of p parallel edges $(a_1, a_2, \dots a_p)$.

Proposition 2: Edges $((v, w)_1, (v, w)_2, \dots (v, w)_p)$ are parallel edges between nodes v and w. Flow f is a flow that optimally solves the maximum flow problem (non-linear, generalized or classical). If flow f between nodes v and w is positive, then the capacity of the edges $(v, w)_i$, $i \in 1..p$, with higher efficiency is exhausted first.

A. Convex transfer functions

Let function F_a be a PWL convex function consisting of p segments defined by breakpoints: $0 = r_a^0 < r_a^1 < r_a^2 < r_a^3 < \dots < r_a^p$, for all $a \in A$ (see Fig. 3, left). Let us denote the function in the interval $[r_a^k, r_a^{k+1}]$ as F_a^k. F_a^k is a linear function of type $F_a^k = \gamma_a^k \cdot f + b$, where γ_a^k is the slope of function F_a^k. Since function F_a is convex, the slopes of the segments are related as follows:

$$\gamma^1 < \gamma^2 < \dots < \gamma^p. \tag{7}$$

We replace edge a by p parallel edges. The transfer function of the k-th edge is F_a^k, $a \in [1..n]$. The capacity of the k-th edge is equal to u_a.

Let us show that this replacement does not influence the optimal solution.

Function F_a^* is convex. Thus, the efficiency of sending flow through edge a grows as the amount of flow through edge a increases.

Since we maximize flow at the target node and, according to Proposition 2, prefer higher efficiency, only one edge of p parallel edges will be used.

$$F_a(f_a) = \begin{cases} \gamma_a^i \cdot f_a^i + b_a^i, & \text{if } f_a \in [r_a^{i-1}, r_a^i], \\ 0, & \text{if } f_a = 0 \end{cases} \tag{8}$$

If we send an amount of flow f_a through edge a and $f_a \in [r_a^k, r_a^{k+1}]$, the k-th edge will be used, because transfer functions F_a^i, $i > k$ or $i < k$ provide lower efficiency for f_a.

Thus, the replacement of edge a by p parallel edges in the way described above does not influence the solution of the maximization problem.

Now, the transfer functions of the edges are of type $\gamma f + b$ and the problem can be transformed to MILP-form (see Section V).

B. Concave transfer functions

Let function F_a be a PWL concave function consisting of p segments defined by breakpoints: $0 = r_a^0 < r_a^1 < r_a^2 < r_a^3 < \dots < r_a^p$, for all $a \in A$. Let us denote the slope of a PWL function in the interval $[r_a^k, r_a^{k+1}]$ as γ_a^k. Since function F_a is concave, the slopes of the segments are related as follows:

$$\gamma^1 > \gamma^2 > \dots > \gamma^p. \tag{9}$$

Flow f_a along edge a is the sum of flows along its segments.

$$f_a = \gamma^1 \cdot f_a^1 + \gamma^2 \cdot f_a^2 + \cdots + \gamma^p \cdot f_a^p = \sum_{k \in \{1..p\}} \gamma_a^k \cdot f_a^k \quad (10)$$

The capacity of the edge representing the k-th segment of edge a is equal to $r_a^k - r_a^{k-1}$. The flow along the k-th segment can be found as follows:

$$f_a^k = \begin{cases} 0, & \text{if } f_a \le r_a^{k-1} \\ f_a - r_a^{k-1}, & \text{if } r_a^{k-1} \le f_a \le r_a^k \\ r_a^k - r_a^{k-1}, & \text{if } f_a \ge r_a^k \end{cases}$$

Every segment a^k can be represented as an edge with transfer function $F_a^k = f_a \cdot \gamma_a^k$ and assigned capacity $(r_a^k - r_a^{k-1})$. An edge a can be replaced by p parallel edges corresponding to the segments of F_a, for all $a \in A$. This replacement is an equivalent replacement, because the edges with the higher γ will be used (exhausted) first when flow is maximized.

The transfer function corresponding to the edges of the modified network are linear. Now we can transform the problem to LP-form and use standard LP-solvers or apply algorithms for the GFP-case.

For example, we can use the algorithm from [2], which takes $O(\epsilon^{-2} m^* (m^* + n \log m^*) \log n)$ time to compute the ϵ-optimal flow on a network with no flow-generating cycles. If we model energy flows, flow generating cycles refer to perpetual energy sources, which is not possible in practice. Flow is ϵ-optimal if $SOL(F) \ge (1-\epsilon) OPT(F)$, $m = |A|$, $n = |V|$ and m^* is the amount of edges in the modified network.

There are two ways to reduce the algorithm's complexity:

1: The generalized shortest-path problem is a subroutine of the GFP. On a graph with no flow-generating paths, it takes $O(m + n \log m)$ time to find a shortest path. We search for the shortest path on the residual graph. We can speed up this algorithm by a special way to set a residual graph. We use the idea described in [1] for convex cost functions. In the residual network, we do not need to consider all $2p$ (p forward and p backward) copies of edges between nodes v and w; it is sufficient to maintain only two edges: the first is for increasing flow through edge (v, w), the second edge is for decreasing flow on it.

Thus, it takes only $O(\epsilon^{-2} m^* (m + n \log m) \log n)$ time to compute the ϵ-optimal flow.

2: The second way to improve the computational performance of the algorithms is to use scaling of the transfer functions. The polynomial-time algorithm for convex cost (with integer costs) flow problems based on this method is presented in [1]. In the first scaling phase, we linearize a concave function $F_{(v,w)}$ by a PWL function consisting of 2 segments of length $\frac{u_{(v,w)}}{2}$. In the second phase, we linearize a concave function $F_{(v,w)}$ by a PWL function consisting of 4 segments of length $\frac{u_{(v,w)}}{4}$, and so on until we reach a segment length of $\frac{u_{(v,w)}}{2^n} \le \epsilon$. Thus, we conduct n scaling phases, $n \ge \lceil \log \frac{U}{\epsilon} \rceil$, where $U = \max_{(v,w) \in A} u_{(v,w)}$. After each scaling phase, we make sure that the flow on the linearized residual network is maximum. It can be shown (analogously to the proof from [1]) that to keep the flow on the network maximum in k-th scaling phases, it is enough to increase or decrease the flow on every edge by $\frac{u_{(v,w)}}{2^k}$ units. There can be at most $O(m)$ augmentations in each scaling phase. The overall algorithm takes $O(m \lceil \log \frac{U}{\epsilon} \rceil)$ augmentations and requires $O(m + n \log m)$ time to find a shortest path, thus, runs in $O(m \lceil \log \frac{U}{\epsilon} \rceil (m + n \log m))$ time.

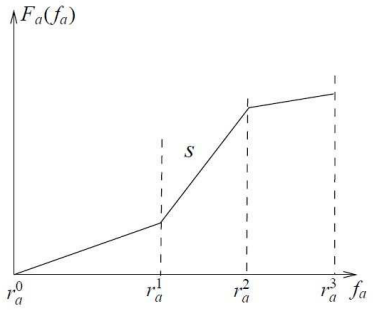


Fig. 4. S-shaped transfer function of edge a.

Algorithm 1 How to find inflection segment

Require: S-shaped PWL function
Ensure: Number of the inflection segment, s
lower bound of s s_b:=1;
upper bound of s s_e:=p;
while $s_b \neq s_e$ **do**
  s_m:=$\lfloor \frac{s_e - s_b}{2} \rfloor$;
  if $\gamma_a^{s_b+s_m} > \gamma_a^{s_b+s_m+1}$ **then**
    slope decreases, continue search on the left part
    s_e:=$s_b + s_m$;
  else
    slope grows, continue search on the right part
    s_b:=$s_b + s_m + 1$;
  end if
end while
s:=s_e; (s_b=s_e, thus s=s_b).

C. S-shaped transfer functions

Let function F_a be an s-shaped PWL function consisting of p segments defined by breakpoints: $0 = r_a^0 < r_a^1 < r_a^2 < r_a^3 < \cdots < r_a^p$, for all $a \in A$. Let us denote the slope of a PWL function in the interval $[r_a^k, r_a^{k+1}]$ as γ_a^k.

First, we divide the s-shaped function in two parts: a concave part and a convex part. With the help of Algorithm 1, which uses the fact, that $\gamma_a^j \neq \gamma_a^{j+1}$, $a \in A$, $j \in [1..p-1]$, we can easily find the number of the *inflection segment* s for edge a in $O(\lceil \log p \rceil)$ time. s is the number of the first segment, where the slope of the following segment is lower than the slope of the current segment. The inflection segment corresponds to the highest slope.

We use Algorithm 2 to build an equivalent representation of an s-shaped PWL transform function, see Fig 5. The new representation fits into MILP framework. The solution of the

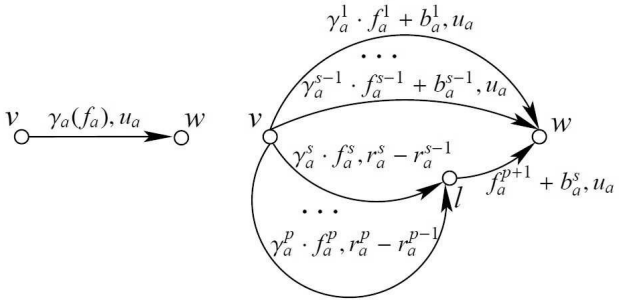


Fig. 5. An equivalent representation of edge a with an s-shaped transfer function.

Algorithm 2 How to find the equivalent representation for s-shaped PWL transfer functions

Require: Initial underlying graph G=(V,A)
Ensure: Equivalent representation of graph G
m:=| V |
for $a = 1 \to m$ **do**
  apply Algorithm 1 for function F_a to find the number of inflection segment for transfer function of edge $a = (v, w)$
  for $i = 1 \to s-1$ **do** {the convex part of the function}
    add edge (v, w) with $F_a^i = \begin{cases} \gamma_a^i \cdot f_a^i + b_a^i, \text{if} f_a^i > 0, \\ 0, \text{otherwise} \end{cases}$ and $u_a^i = u_a$;
  end for
  for $i = s \to p$ **do** {the concave part of the function}
    add edge (v, l) with $F_a^i = \gamma_a^i \cdot f_a^i$ and $u_a^i = r_a^i - r_a^{i-1}$;
  end for
  add edge (l, w) with $F_a^{p+1} = f_a^{p+1} + b_s$ and $u = \infty$.
end for

flow maximization problem on the new representation (by summing over edges a^i, $i \in [1..p]$) and on the initial underling network will be equivalent.

The transfer function of edge a can be found as

$$F_a(f_a) = \begin{cases} 0, \text{if } f_a = 0, \\ \gamma_a^i \cdot f_a^i + b_a^i, \text{if } f_a \in [r_a^{i-1}, r_a^i] \text{ and } i < s, \\ \gamma_a^s \cdot f_a^s + b_a^s, \text{if } f_a \in [r_a^{i-1}, r_a^i] \text{ and } i = s, \\ \sum_{j \in [s..i-1]} \gamma_a^j \cdot u_a^j + \gamma_a^i \cdot f_a^i + b_a^s, \\ \qquad \text{if } f_a \in [r_a^{i-1}, r_a^i] \text{ and } i > s. \end{cases}$$

Capacity limitation edge. To maintain the capacity limit on edge a, $a \in A$, we introduce a total capacity limitation on edges a_i, $i \in [1..p_a]$.

$$\sum_{i \in [1..p]} u_{a_i} \le u_a.$$

We can either integrate this constraints into the MILP formulation for all $a \in A$ or add a bottleneck edge with capacity u_a to node v so that all flow incoming to node v must first pass through this bottleneck edge. For convex and concave cases, the number of edges on the expanded graph per edge on the original graph p_a^+ becomes $p_a + 1$. For s-shaped transfer functions, p_a^+ becomes $p_a + 2$.

We can formulate the maximization problem on the modified network as follows:

Problem 2:

Given $D = (V, A)$, $u_a \ \forall a \in A$, $F_a \ \forall a \in A$.
Find an $s - t-$flow, that
maximizes
$$\sum_{a \in \delta^-(t)} \sum_{i \in 1..p_a^+} F_a^i(f_a)$$
subject to
$$\sum_{a \in \delta^+(v)} \sum_{i \in 1..p_a^+} f_a^i - \sum_{a \in \delta^-(v)} \sum_{i \in 1..p_a^+} F_a^i(f_a^i) \le 0, \forall v \in V \backslash \{s, t\}$$
$$0 \le f_a^i \le u_a^i, \sum_{i \in [1..p_a]} u_a^i \le u_a, \forall a \in A, i \in 1..p_a^+.$$

Theorem 2: The optimal solution of the flow maximization problem for *Problem 1* with s-shaped transfer functions can be generated from the optimal solution of the flow maximization problem for *Problem 2*.

Proof: If we send an amount of flow f_a through edge a and $f_a \in [r_a^{k-1}, r_a^k]$ and $k < s$, then (like for the standard convex case) the k-th edge will be used, because transfer functions F_a^i, $i > k$ or $i < k$, provide lower efficiency for f_a. And if we send flow $f_a < r_a^{s-1}$, maximization of the flow leads us to choose one of the edges in new representation, and send the whole flow f_a through this edge.

If we send $f_a \in [r_a^k, r_a^{k+1}]$ and $k \ge s$, then we are on the concave part of the function. Segment s is less efficient as any segment of the concave part. The function, which describes the s-shaped PWL function of segment s, can be written as $F_a^s = \gamma_a^s + b_a^s$. Per definition of concavity, for $i > s$ any following segment is less efficient than the previous and, thus, $\gamma_a^i > \gamma_a^{i+1}$ and $\gamma_a^i + b_a^s > \gamma_a^{i+1} + b_a^s$ for $i > s$. In the other words, if we send flow $f_a \ge r_a^{s-1}$, we first fulfill edge s, and only then, if $u_s < f_a$, the following edges. Then, $f_a = \sum_{i=s \ldots p_a} f_a^i$. ■

Parallel edges. By allowing parallel edges, we meet some notational difficulties, because an edge cannot be uniquely specified by its tail and its head. We can use the following to build the equivalent network without using parallel edges. If there are parallel edges between node v and w, we replace all but one of these edges by a pair of series-connected edges. The first edge in the pair stays as original, the second one gets a transfer function equal to one and capacity equal to ∞. For implementation, instead of ∞, we use a big number, which guarantees no capacity limitation on this edge. Flow at any edge of our network is less than $\sum_{a \in \delta_-(s)} f_a$. Thus, in order not to create capacity limitations, we can use any number greater than $\sum_{a \in \delta_-(s)} f_a$. Thus, we can easily replace a graph with parallel edges by an equivalent graph without parallel edges. A negative effect of this replacement is that it expands the underlying network. For convex and concave cases, the number of edges on the expanded graph per edge on the original graph p_a^+ becomes $p_a + 1 + p_a - 1 = 2 \cdot p_a$.

For s-shaped transfer functions, in the subgraph replacing edge a, there are two sets of parallel edges, thus, p_a^+ becomes $p_a + 2 + p_a - 2 = 2 \cdot p_a$.

Graph expansion. If the PWL function of any edge $a \in A$ consists of p segments, the number of edges in the expanded graph becomes $m \cdot 2p$, where m is the amount of edges in the initial graph (with NL PWL transfer functions). If the number of segments in the approximation function p_a is different for every edge a, the number of edges in the modified network becomes $\sum_{a \in A} 2 \cdot p_a$ Let us denote the number of edges in the modified network as m^*.

D. Monotonically growing transfer functions

Monotonically growing PWL transfer functions can be divided into s-shaped fragments. For each s-shaped fragment, we establish an equivalent network representation as described in section IV-C. Thus, the correct segment within each s-shaped fragment is chosen automatically due to the network's structure. We need to use the extra variables only to force the right (depending on the amount of flow) s-shaped fragment to be taken.

V. MILP-model for max-flow problems on AGFP network

In this section, we present the formulation of the maximum flow problem on affine-linear generalized network in MILP form.

In the model, we have to integrate the following properties of F_a:

$$F_a = \begin{cases} f_a \cdot \gamma_a + b_a, & \text{if } f_a > 0 \\ 0, & \text{if } f_a = 0 \end{cases}$$

To do this, we use binary variables $f_a^* \in \{0,1\}$, $a \in A$. If $f_a^* = 1$, then flow f_a is greater then zero. If $f_a^* = 0$, then flow f_a is zero. The capacity of flow f_a through edge a is set to zero unless $f_a^* = 1$.

Given $D^* = (V, A)$ (D^* is an expanded/modified graph), u_a, k_a, b_a, r_a $\forall a \in A$.
Find an $s-t-$flow, that
maximizes

$$\sum_{a \in \delta^-(t)} (f_a \cdot \gamma_a + b_a \cdot f_a^*)$$

subject to

$$\sum_{a \in \delta^+(v)} f_a - \sum_{a \in \delta^-(v)} (f_a \cdot \gamma_a + b_a \cdot f_a^*) = 0, \forall v \in V \backslash \{s,t\}$$

$$f_a^* \in \{0,1\}, \forall a \in A$$

$$0 \le f_a \le u_a \cdot f_a^*, \forall a \in A$$

VI. Model Evaluation

Vielma et al. [10] study different ways to model PWL functions in MILP form. They consider the disaggregated convex combinations model, the logarithmic model, the logarithmic disaggregated model and other models. These models are characterized by size of the formulation. Since the quantitative measurement of constraints, continuous and binary variables in formulations is ambivalent (e.g. some models require less constraints, but need more binaries), computational experiments were conducted and presented in [10]. The performance of the logarithmic model was considered the best. According to [10], the crucial parameter that defines time performance is the number of additional continuous variables. Our model, which uses the special properties of the considered PWL functions, does not need additional continuous variables to choose the right edge. Remember, the right edge is the edge that corresponds to the segment of a PWL function that would be chosen if we sent $\sum_{i \in [1..p_a]} f_a^i$ through edge a. Thus, we reduce the number of continuous variables, and thus, improve the computational performance.

VII. Error Estimation

Let us consider the situation in which the original transfer functions $F(\cdot)$ are not PWL functions. Non-linear monotonically increasing functions can be approximated by monotonically increasing PWL functions ($F^A(\cdot)$), and the approach described in this paper can be applied. Obviously, the optimal solution for the original transfer functions $OPT(F)$ does not have to be equivalent to the optimal solution for the approximating transfer function $OPT(F^A)$. It is important to be able to estimate how the approximation quality influence dependence between $OPT(F)$ and $OPT(F^A)$.

The approximation error from above, $\epsilon^\uparrow$, can be defined as follows:

$$\epsilon^\uparrow = \max_{a \in A} \max_{0 \le f_a \le u_a} \left(\frac{\gamma_a^{A\uparrow}(f_a) - \gamma_a(f_a)}{\gamma_a(f_a)} \right) =$$

$$= \max_{a \in A} \max_{0 \le f_a \le u_a} \left(\frac{\gamma_a^{A\uparrow}(f_a)}{\gamma_a(f_a)} - 1 \right),$$

i.e. $\epsilon^\uparrow \ge \frac{\gamma_a^{A\uparrow}(f_a)}{\gamma_a(f_a)} - 1, \forall a \in A, 0 \le f_a \le u_a$.
The latter can be reformulated as follows:
$(\epsilon^\uparrow + 1)\gamma_a(f_a) \ge \gamma_a^{A\uparrow}(f_a), \forall a \in A, 0 \le f_a \le u_a$.

The approximation error from below, $\epsilon^\downarrow$, can be defined as follows:

$$\epsilon^\downarrow = \max_{a \in A} \max_{0 \le f_a \le u_a} \left(\gamma_a(f_a) - \frac{\gamma_a^{A\downarrow}(f_a)}{\gamma_a(f_a)} \right) =$$

$$= \max_{a \in A} \max_{0 \le f_a \le u_a} \left(\frac{1 - \gamma_a^{A\downarrow}(f_a)}{\gamma_a(f_a)} \right),$$

i.e. $\epsilon^\downarrow \ge 1 - \frac{\gamma_a^{A\downarrow}(f_a)}{\gamma_a(f_a)}, \forall a \in A, 0 \le f_a \le u_a$.
The latter can be reformulated as follows:
$(1 - \epsilon^\downarrow)\gamma_a(f_a) \le \gamma_a^{A\downarrow}(f_a), \forall a \in A, 0 \le f_a \le u_a$.

Theorem 3:

$$\frac{OPT(\gamma^{A\uparrow})}{(1+\epsilon^\uparrow)^z} \le OPT(\gamma) \le \frac{OPT(\gamma^{A\downarrow})}{(1-\epsilon^\downarrow)^z},$$

where z is the amount of edges in the longest $s-t$-path, at most $|A|$.

Proof: We split the proof into two parts.

Part 1. We prove that $OPT(\gamma) \geq \frac{OPT(\gamma^{A\uparrow})}{(1+\epsilon^{\uparrow})^z}$.

Suppose we know the optimal solution[1] $f^{A\uparrow}_{OPT}$ for $\gamma^{A\uparrow}(\cdot)$. According to the flow decomposition theorem (Theorem 1), we can decompose $f^{A\uparrow}_{OPT}$ into l paths, $l \leq |A|$. The amount of flow we send through path $P_i = a_1, a_2, \ldots, a_{ki}$, $i \in [1, l]$ is f_i. The amount of flow that reaches the target node through path P_i is $\gamma^{A\uparrow}_{P_i}(f_i)$[2].

Let us take the same amount of flow at the source node as we use to obtain $OPT^{A\uparrow}$ and send it through the network with transfer functions $\gamma(\cdot)$ according to distribution $X(f^{A\uparrow}_{OPT})$ (Def. 6). This yields flow f. Now, let us apply the same path representation (which was obtained by flow decomposition) in its distribution form X to flow f. Thereby, we do a valid decomposition of flow f into l paths. The flow we send through path P_i stays f_i, flow arriving at the target node through path P_i is $\gamma_{P_i}(f_i) = \gamma_{ki}(\gamma_{ki-1} \ldots \gamma_1(f_i))$.

Assume that we have a limit on the rate of transfer function growth, $\gamma'_a(f_a)$, $0 \leq f_a \leq u_a$, $a \in A$. Then, $\gamma_k(\gamma_j(x)(1 + \epsilon^{\uparrow})) \leq \gamma_k(\gamma_j(x))(1 + \epsilon^{\uparrow})$, $\epsilon^{\uparrow} \geq 0$, $j, k \in A$.

Then, $\gamma^{A\uparrow}_{P_i}(f_i) \leq \gamma_{ki}(\gamma_{ki-1}(\ldots\gamma_1(f_i)(1+\epsilon^{\uparrow}))(1+\epsilon^{\uparrow}))(1+\epsilon^{\uparrow}) \leq \gamma_{ki}(\gamma_{ki-1}(\ldots\gamma_1(f_i)))(1+\epsilon^{\uparrow})^{ki} = \gamma_{P_i}(f_i)(1+\epsilon^{\uparrow})^{ki}$. Taking into account that $OPT(\gamma) \geq \sum_{i\in[1..l]} \gamma_{P_i}(f_i)$, we can state the following:

$OPT(\gamma^{A\uparrow}) = \sum_{i\in[1..l]} \gamma^{A\uparrow}_{P_i}(f_i) \leq \sum_{i\in[1..l]} \gamma_{P_i}(f_i)(1+\epsilon^{\uparrow})^{ki} \leq OPT(\gamma) \cdot (1+\epsilon^{\uparrow})^z$. q.e.d.

Part 2. We prove that $\frac{OPT(\gamma^{A\downarrow})}{(1-\epsilon^{\downarrow})^z} \geq OPT(\gamma)$.

This part of the proof is analogous to the first part.

First, we suppose to know the optimal solution for $\gamma(\cdot)$. Let us take a flow decomposition for this solution and apply it to $\gamma^{A\downarrow}(\cdot)$. By this, no capacity constraint is broken. The amount of flow at the target node for the second solution is greater than for the first. The optimal solution for $\gamma^{A\downarrow}(\cdot)$ is the solution that provides the greatest amount of flow at the target node for the graph with $\gamma^{A\downarrow}(\cdot)$. Thus, we can compare $OPT(\gamma^{A\downarrow})$ and $OPT(\gamma)$.

$OPT(\gamma) = \sum_{i\in[1..l]} \gamma_{P_i}(f_i) \leq \sum_{i\in[1..l]} \frac{\gamma^{A\downarrow}_{P_i}(f_i)}{(1+\epsilon^{\downarrow})^{ki}} \leq \frac{OPT(\gamma^{A\downarrow})}{(1+\epsilon^{\downarrow})^z}$. q.e.d.

■

[1]Remember, that the solution is the amount of flow at all edges a, $a \in A$, and $OPT(\cdot)$ is the amount of flow at the target node.

[2]Here, we do not consider transfer functions, but residual transfer functions. We omit writing next to each transfer function that it is a residual function to avoid overloading notations. Remember, if flow on an edge is zero, then the residual transfer function of this edge is equal to the transfer function of this edge, i.e. $F^R(0) = F$.

VIII. Conclusions

Better performance of the model introduced in this paper is based on the usage of the characteristics of special types of PWL functions.

For the maximum flow problem on networks with concave transfer functions, we propose a polynomial-time approximation scheme.

For the maximum flow problem on networks with monotonically growing transfer functions (this case can be reduced to convex and s-shaped transfer functions), we propose to split the transfer functions into s-shaped segments and then, with the help of extra variables, force only the right s-shaped fragment to be taken. We establish such a network representation that the correct segment within each s-shaped fragment is chosen automatically due to the network's structure. This yields to MILP model with better performance than if we were using standard MILP formulations of PWL functions, e.g. the logarithmic model. Moreover, it is worth to mention that our model is simple and transparent, which makes implementation easy.

References

[1] Ahuja, R.K., Magnanti, T.L. and Orlin, J.B. "Network flows: Theory, Algorithms and Applications." Prentice Hall, NJ, 1993.

[2] Fleischer, L. and Wayne, K.D. "Fast and simple approximation schemes for generalized flow." *Mathematical Programming*, Vol. 91, pp. 215-238, 2002.

[3] Keha, A.B., de Farias, I.R. and Nemhauser, G.L. "A branch-and-cut algorithm without binary variables for non-convex piecewise linear optimization." *Operations Research*, Vol. 54, pp. 847-857, 2005.

[4] Korte, B. and Vygen, J. *Combinatorial Optimization: Theory and Algorithms.* Algorithms and Combinatorics, 21 Springer, 2006.

[5] Oldham, J. "Combinatorial Approximation Algorithms for Generalized Flow Problems", In *Proceedings of ACM/SIAM*, 1999, pp. 135-169.

[6] Onaga, K. "Dynamic programming of optimum flows in lossy communication nets." *IEEE Transactions. Circuit Theory*, Vol. 13, pp. 308-327, 1966.

[7] Radzik, T. "Faster algorithms for the generalized network flow problem." *Mathematics of Operations Research*, Vol. 23, pp. 69-100, 1998.

[8] Tardos, E. and Wayne, K. "Simple Generalized Maximum Flow Algorithms." In *Integer Programming and Combinatorial Optimization,* Lecture Notes in Computer Science, Vol. 1412, pp. 310-324. Springer, 1998.

[9] Truemper, K. "On max flows with gains and pure min-cost flows." *SIAM Journal on Applied Mathematics*, Vol. 32, pp. 450-456, 1977.

[10] Vielma, J.P., Ahmed, S. and Nemhauser, G. "Mixed-Integer Models for Nonseparable Piecewise Linear Optimization." *Discrete Optimization*, Vol. 5, pp. 467-488, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 417–420

A Hybrid Algorithm based on Differential Evolution, Particle Swarm Optimization and Harmony Search Algorithms

Ezgi Deniz Ülker
Computer Engineering Department,
Girne American University, University Drive, Karaoglanoglu, Girne, Mersin 10, Turkey
Email: ezgideniz@gau.edu.tr

Ali Haydar
Computer Engineering Department,
Girne American University, University Drive, Karaoglanoglu, Girne, Mersin 10, Turkey
Email: ahaydar@gau.edu.tr

***Abstract*—Evolutionary optimization algorithms and their hybrid forms have become popular for solving multimodal complex problems which are very difficult to solve by traditional methods in the recent years. In the literature, many hybrid algorithms are proposed in order to achieve a better performance than the well-known evolutionary optimization methods being used alone by combining their features for balancing the exploration and exploitation goals of the optimization algorithms. This paper proposes a novel hybrid algorithm composed of Differential Evolution algorithm, Particle Swarm Optimization algorithm and Harmony Search algorithm which is called HDPH. The proposed algorithm is compared with these three algorithms on the basis of solution quality and robustness. Numerical results based on several well-studied benchmark functions have shown that HDPH has a good solution quality with high robustness. Also, in HDPH all parameters are randomized which prevents the disadvantage of selecting all possible combination of parameter values in the selected ranges and of finding the best value set by parameter tuning.**

I. Introduction

IN RECENT years, many different optimization techniques have been proposed for solving the complex, multimodal functions in several fields [1-4]. Some of the well-known optimization algorithms are the Genetic Algorithm (GA), Particle Swarm Optimization (PSO) algorithm, Ant Colony Optimization (ACO) algorithm, Differential Evolution (DE) algorithm, and Harmony Search (HS) algorithm. These algorithms are used in various fields by many researchers to obtain the optimum value of the problems [5-10]. Each optimization algorithm uses different properties to keep a balance between the exploration and exploitation goals which can be a key for the success of an algorithm. Exploration attribute of an algorithm enables the algorithm to test several areas in the search space. On the other hand, exploitation attribute makes the algorithm focus the search around the possible candidates. Although the optimization algorithms have positive characteristics, it is shown that these algorithms do not always perform as well as it is desired [11]. Because of this, hybrid algorithms are growing area of interest since their solution quality can be made better than the algorithms that form them by combining their desirable features. Hybridization is simply the combination of two or more techniques in order to outperform their performances by the use of their good properties together. Hybridization has been done in several different ways in the literature and it is observed that the new hybridization techniques are very efficient and effective for optimization [11-16].

A novel hybrid algorithm proposed in this paper is called HDPH and it is a combination of three well known evolutionary algorithms, namely Differential Evolution (DE) algorithm, Particle Swarm Optimization (PSO) algorithm, and Harmony Search (HS) algorithm. It merges the general operators of each algorithm recursively. This achieves both good exploration and exploitation in HDPH without altering their individual properties.

HDPH is compared with the three algorithms that form it on the basis of the solution quality and the robustness on random initialization of a solution set. The set of well studied benchmark functions which are Multimodal (M)/Separable (S) or Multimodal (M)/Non-separable(N) are used for the evaluation.

The rest of the paper is organized as follows; Section II describes the HDPH algorithm that is proposed in detail. Section III presents the performances of the hybrid algorithm and the algorithms that generate it together and also our discussions. In the last section, Section IV, the concluding remarks of the paper are given.

HDPH Algorithm

In the literature, many different ways of combining the well-known algorithms are performed to obtain more powerful optimization algorithms [11-16]. The main aim of the hybridization is to use different properties of different algorithms to improve the solution quality.

Among the well-known algorithms, DE, PSO and HS algorithms are the three algorithms that are used in many fields by researchers and these algorithms are proven to be very powerful optimization tools [5-8]. Each algorithm has different strong features. As an example, DE usually requires less computational time and also has better approxi-

978-83-60810-53-8/$25.00

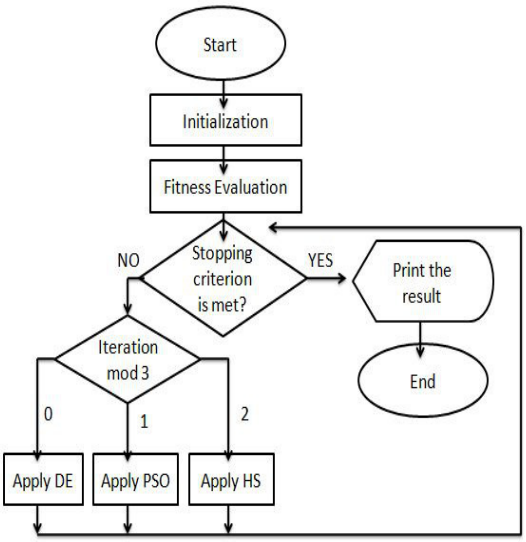


Fig. 1.Flowchart of HDPH

mation of solutions for most of the problems. PSO generally avoids the solution from trapping into local minima by using its diversity. HS on the other hand, is an efficient algorithm that has a very good performance on different applications.

HDPH uses the operators of these three algorithms with randomly selected parameters consecutively and by not altering their properties. The new candidate set, obtained by each algorithm, is used as a new solution set for the other algorithm. Fig.1 shows the HDPH algorithm in the form of a flowchart which demonstrates the main steps of the process.

The summarized steps of HDPH can be given as follows:

Step 1. *Generation of the candidate population with given dimensions:* Initialize the candidate population $X_{i,j}$ in a given range.

Step 2. *Crossover and mutation operators of DE:* The mutation and crossover operators are applied to find the better approximation to a solution by using (1), (2), and (3).

The mutant vector V_{ij} is calculated as corresponding to each member in population using (1) where *a, b,* and *c* are distinct numbers. Mutant vector V_{ij} is crossoverred with X_{ij} and trial vector U_{ij} is generated by using (2) where r_j is a uniformly distributed number for each j^{th} parameter of X_i. Also, *F* and *CR* are the main control parameters of DE.

$$V_i = X_a + F(X_b - X_c) \tag{1}$$

$$U_{ij} = \begin{cases} V_{ij} & if \;\; r_j \le CR \\ X_{ij} & otherwise \end{cases} \tag{2}$$

$$X_i = \begin{cases} U_i & if \;\; f(U_i) < f(X_i) \\ X_i & otherwise \end{cases} \tag{3}$$

Selection process determines U_{ij} to survive to the next generation by using (3).

TABLE I
MULTIMODAL-SEPARABLE AND MULTIMODAL-NON-SEPARABLE BENCHMARK FUNCTIONS

D	Function	Formula	f_{min}
2	Booth	$\left[(x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2 \right]$	0
30	Rastrigin	$\sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) + 10 \right]$	0
30	Schwefel	$\left[\sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert}\right) \right]$	-418.9*D
10	Michalewicz 10	$\left[-\sum_{i=1}^{n} \sin(x_i) * \left(\sin\left(\frac{ix_i^2}{\pi} \right) \right)^{2m} \right]$ $m = 10$	-9.6602
2	Schaffer	$\left[0.5 + \frac{\sin^2\left(\sqrt{x_1^2 + x_2^2}\right) - 0.5}{\left(1 + 0.001\left(x_1^2 + x_2^2\right)\right)^2} \right]$	0
2	Six Hump Camel Back	$\left[4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4 \right]$	-1.03163
2	Shubert	$\left[\left(\sum_{i=1}^{5} i\cos((i+1)x_1 + i) \right) * \left(\sum_{i=1}^{5} i\cos((i+1)x_2 + i) \right) \right]$	-186.73
30	Griewank	$\left[\frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}} \right) + 1 \right]$	0
30	Ackley	$\left[-20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n} \cos(2\pi x_i) \right) + 20 + e \right]$	0
30	Penalized	$\left[\frac{\pi}{n} \left\{ 10\sin^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 \left[1 + 10\sin^2(\pi y_{i+1}) \right] + (y_n - 1)^2 \right\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4) \right]$ $y_i = 1 + \frac{1}{4}(x_i + 1)$ $u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, x_i > a \\ 0, -a \le x_i \le a \\ k(-x_i - a)^m, x_i < a \end{cases}$	0

Step 3. *Particle movement by PSO:* The randomly selected parameters are applied on the velocities by using (4). When a better solution is being discovered, all particles improve their positions by using (5). This movement avoids the particles to be trapped to the local minima by increasing the diversity of solution. V_{ij} refers to the velocity values and for each row is calculated according to the control parameters c_1, c_2, and w by using (4). $global_{best}$ is the best position obtained by any particle and P_{best} is the personal best of a particle. X_{ij} refers to current positions of a particle and can be updated by using (5) for each row.

$$V_i = w * V_i + c_1 * (P_{best} - X_i) + c_2 * (global_{best} - X_i) \quad (4)$$

$$X_i = X_i + V_i \quad (5)$$

Step 4. *Choosing a neighboring value by HS:* HS can search in different zones of the search space by using the control parameters that are *hmcr, par* and *fw*. With a given probability of *hmcr*, a value is selected from the candidate population. With a given probability of 1-*hmcr*, a random candidate is generated in the given range. The population can have non-updated candidates to keep the diversity in the population with a given probability of 1-*par*. With a given probability of *par,* the candidates are updated by applying (6) where *rand()* is a random number $\in$ (-1,1).

$$X_i = X_i + rand() * fw \quad (6)$$

Step 5. Consecutively Step 2, Step 3, and Step 4 are applied.

The algorithm is performed until the termination criterion is not satisfied. Elitism is included in HDPH by keeping the best solution at the end of each iteration.

III. Numerical Results and Discussions

The proposed hybrid algorithm HDPH is tested using 10 well known benchmark functions with different characteristics and is compared for the solution quality and robustness for random initialization of the population with the three algorithms used to form it. The benchmark functions are selected as Multimodal (M)/Separable (S) and Multimodal (M)/Non-separable (N). These benchmark functions are presented in Table I. The population size for the functions is fixed to 100 for all algorithms.

The two control parameters of DE algorithm which are *F* and *CR* are selected from the sets given as follows; *F* $\in$ {0.3, 0.5, 0.7, 0.8, 0.9, 1.2, 1.4} and *CR* $\in$ {0.1, 0.2, 0.4, 0.6, 0.8, 0.9}. The three control parameters that are used in PSO algorithm are selected from the sets as given; *c1* and *c2* $\in$ {0.3, 0.6, 0.9, 1.2, 1.5, 1.8}, and *w* $\in$ {0.4, 0.5, 0.6, 0.7, 0.8, 0.9}. For the HS algorithm, the control parameters called *hmcr* and *par* are selected from the sets as follows; *hmcr* $\in$ {0.7, 0.8, 0.9, 0.93, 0.96, 0.98} and *par* $\in$ {0.01, 0.02, 0.05, 0.1, 0.2}. The control parameter *fw* is adjusted as 0.01, 0.05, 0.1, and 0.2 times the upper bound of each function in HS. Each possible combination of control parameters is selected and each selection is run for 20 times for each algorithm. The selected parameter ranges are chosen similar to the commonly used ranges in the literature. For each function, the control parameter values that are closest to the optimum solution are selected and the function is further evaluated around these selected control parameters. By doing this, we try to achieve a good parameter tuning.

For the HDPH model, instead of selecting the eight control parameters discretely from the sets used for three algorithms, they are selected randomly from the parameter ranges that are formed by selecting the minimum and maximum elements of each parameter set as the lower and upper bounds of the ranges for these parameters respectively. This is done because it would have been very difficult to test all possible combination of parameter values otherwise. The results are obtained only by running the hybrid model for 20 times.

In Tables II and III, the performances of these algorithms over 10000 function evaluations are shown. For each algorithm, the best value (*BestVal*), the average (*Avg*) of the 20 runs for the selected best value parameters and the standard deviations (*Stdev*) are shown. In case that there are more than one control parameter values that give the best value, the one that has a closer average to the optimal value and smaller standard deviation is chosen.

The results that are obtained for the selected MS functions are shown in Table II. The best values obtained using HDPH, except Rastrigin function, are either better or similar to the best values obtained by the other three algorithms. The standard deviations and the averages of HDPH for Schwefel and Michalewicz10 functions are substantially better than the other three algorithms. However, for the Rastrigin function, the standard deviation and the average of HS algorithm are better than HDPH and for the Booth function, all three algorithms have a better standard deviation values compared to HDPH.

In Table III, the results for MN functions are tabulated. For Griewank, Ackley, and Penalized functions, the best values obtained using HDPH outperform the other three algorithms. For these three functions, when both the average and standard deviation values are taken into consideration, the HDPH gives better results than DE and PSO algorithms. When it is compared by the HS algorithm, except Ackley function which gives similar results, HDPH is again better than HS algorithm. For the Schaffer, Six Hump Camel Back and Shubert functions, both the best values and standard deviations are comparable for all four algorithms.

It can be seen from the results that HDPH generally worked as good as or sometimes better than other three algorithms in terms of solution quality and robustness. This is achieved by running the HDPH algorithm only 20 times. For the other three algorithms, the tabulated results are obtained by running the programs 20 times for all possible combinations of parameters, finding the parameter set that gives the best performance, making a parameter tuning around those values and using those parameters that has achieved the best performance. This point is a verification of the good performance of HDPH algorithm.

TABLE II
RESULTS FOR MULTIMODAL-SEPARABLE FUNCTIONS

Function MS	Values	HDPH	DE	PSO	HS
Booth	Avg	0.0001	1.37E-25	0	3.36E-07
	Stdev	0.0007	1.80E-25	0	5.02E-07
	BestVal	0	2.41E-27	0	1.27E-08
Rastrigin	Avg	36.18	137.899	46.3655	18.1256
	Stdev	14.203	6.68121	17.416	3.41769
	BestVal	21.82	126.013	19.0816	12.7443
Schwefel	Avg	-12567.6	-7485.74	-8531.08	-12554.6
	Stdev	2.5759	270.62	949.247	28.8299
	BestVal	-12569.5	-8128.58	-10353.9	-12566.1
Micha10	Avg	-9.65918	-9.13592	-8.00576	-9.6111
	Stdev	0.0025	0.11902	0.93836	0.05591
	BestVal	-9.66015	-9.31606	-9.65524	-9.66004

TABLE III
RESULTS FOR MULTIMODAL-NON-SEPARABLE FUNCTIONS

Function MN	Values	HDPH	DE	PSO	HS
Schaffer	Avg	0.007923	0.00107	0.00250	0.00923
	Stdev	0.003701	0.00088	0.00428	0.00217
	BestVal	0	7.80E-05	0	1.85E-06
Six Hump Camel Back	Avg	-1.03163	-1.03163	-1.03163	-1.03163
	Stdev	0	0	0	3.66E-06
	BestVal	-1.03163	-1.03163	-1.03163	-1.03163
Shubert	Avg	-186.722	-185.624	-186.729	-186.727
	Stdev	0.026154	1.40940	0.00959	0.00438
	BestVal	-186.731	-186.703	-186.731	-186.731
Griewank	Avg	0.045248	1.53190	0.39352	1.04977
	Stdev	0.071979	0.19544	0.31861	0.0222
	BestVal	0.000208	1.29684	0.05316	1.00414
Ackley	Avg	0.885152	16.7758	2.86698	1.09805
	Stdev	0.594669	0.75719	1.01934	0.29879
	BestVal	0.007775	15.1746	0.65150	0.56317
Penalized	Avg	0.031359	5.08107	4.36306	0.29210
	Stdev	0.056563	2.13136	2.94708	0.24432
	BestVal	3.59E-06	2.79334	0.37862	0.04269

IV. CONCLUSION

In this work, the new hybrid algorithm, called HDPH, is proposed to achieve a robust algorithm with a good solution quality by combining the three well-known algorithms, DE, PSO and HS. The performances of chosen algorithms are based on the parameter selection. Therefore, all combination of parameter values are tested for each function for all three algorithms and the results that are tabulated are selected as the best values obtained through all possible trials. However, in the HDPH algorithm the parameters are chosen randomly in the given ranges which make the algorithm easier to implement. Even with this kind of simplification in HDPH algorithm, the good performance is verified. Also, the experimental results have shown that, when both solution quality and robustness of an algorithm are taken into consideration, in most of the test functions, HDPH is more robust than the other three algorithms. At the same time, HDPH, for many functions analyzed, has similar or even better solution quality than the three algorithms that composes it. Hence, the proposed hybrid algorithm, HDPH, makes use of the features of the three algorithms and has similar or better solution quality with high robustness to random initialization of the population.

REFERENCES

[1] Z.W. Geem, J.H. Kim, and G.V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search", *Simul., Trans. of the Soc. for Model. and Simul. Int*., pp. 60-68, 2001.

[2] J. Kennedy, R. Eberhart, "Particle Swarm Optimization. Piscataway", *Proc. of IEEE Int. Conf. on Neural Netw. IV, NJ: IEEE Press*, pp. 1942-1948, 1995.

[3] R. Storn, K. Price, "Differential Evolution; A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces", *J. of Glob. Optim*., vol.11, pp. 341-359, 1997.

[4] M. Dorigo, G. Dicaro, "Ant colony optimization: A new meta-heuristic", *Evol. Comput. CEC 99. Proc. of the 1999 Congr. on.*, vol. 2, pp. 1470-1477, 1999.

[5] F. Wenlong, M. Johnston, and M.Zhang, "Soft edge maps from edge detectors evolved by genetic programming", *Evol. Comput. Proc. IEEE,* pp.1-8, 2012.

[6] R. Storn, "Differential Evolution Design of an IIR-filter", *Evol. Comput. Proc. IEEE,* pp. 268-273, 1996.

[7] Z.W. Geem, J.H. Kim, and G.V. Loganathan, "Harmony Search optimization, Application to pipe network design", *Int. J. of Model.&Simul.,* vol. 22, pp. 125-133, 2002.

[8] Z.W. Geem, C. Tseng, and Y. Park, "Harmony Search for Generalized Orienteering Problem: Best touring in China", *Springer Lect. Notes in Comput. Sci.,* vol. 3412, pp. 741-750, 2005.

[9] I. Hitoshi, "Using genetic programming to predict financial data", *Evol. Comput. CEC 99. Proc. of the 1999 Congr. on.*, vol.1, pp.244-251, 1999.

[10] R.S. Parpinelli, H.S. Lopes, and A.A. Freitas, "Data mining with an Ant Colony Optimization Algorithm", *IEEE Trans. on Evol. Comput.,* vol.6, pp. 312-332, 2002.

[11] R. Thangaraj, M. Pant, A. Abraham, and P. Bouvry, "Particle Swarm Optimization: Hybridization perspectives and experimental illustrations", *Appl. Math. and Comput.,* vol. 217, pp. 5208-5226, 2011.

[12] X.H. Shi, Y.C. Liang, and L.M. Wang, "An improved GA and novel PSO-GA-based hybrid algorithm", *Inf. Process. Lett.,* vol. 93, pp. 255-261, 2005.

[13] N. Holden, A.A. Freitas, "A hybrid particle swarm/ant colony algorithm for the classification of hierarchical biological data." *Swarm Intell. Symp. SIS 2005,* pp.100-107, 2005.

[14] A.A.A. Esmin, G.T. Torres, and G.B. Alvarenga, "Hybrid Evolutionary Algorithm Based on PSO and GA mutation", *In proc. of the Sixth Int. Conf. on Hybrid Intell. Syst*., pp.57, 2006.

[15] H. Li, and H. Li, "A Novel Hybrid Particle Swarm Optimization Algorithm Combined with Harmony Search for High Dimensional Optimization Problems", *The Int. Conf. on Intell. Pervasive Comput*., pp. 94-97, 2007.

[16] I. Cionei, E. Kyriakides, "Hybrid Ant Colony-Genetic Algorithm (GAAPI) for Global Continuous Optimization", *IEEE Trans. on Syst., Man, and Cybern. - Part B: Cybern.,* vol.42, pp. 234-245, 2012.

Computer Science & Network Systems

CSNS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to more technical aspects of computer science and related disciplines. The CSNS area spans themes ranging from hardware issues close to the discipline of computer engineering via software issues tackled by the theory and applications of computer science and to communications issues of interest to distributed and network systems. Events that constitute CSNS are:

- CANA'2013 - Computer Aspects of Numerical Algorithms
- MMAP'2013 - International Symposium on Multimedia Applications and Processing

Computer Aspects of Numerical Algorithms

Numerical algorithms are widely used by scientists engaged in various areas. There is a special need of highly efficient and easy-to-use scalable tools for solving large scale problems. The workshop is devoted to numerical algorithms with the particular attention to the latest scientific trends in this area and to problems related to implementation of libraries of efficient numerical algorithms. The goal of the workshop is meeting of researchers from various institutes and exchanging of their experience, and integrations of scientific centers.

Topics

- Parallel numerical algorithms
- Novel data formats for dense and sparse matrices
- Libraries for numerical computations
- Numerical algorithms testing and benchmarking
- Analysis of rounding errors of numerical algorithms
- Languages, tools and environments for programming numerical algorithms
- Numerical algorithms on GPUs
- Paradigms of programming numerical algorithms
- Contemporary computer architectures
- Heterogeneous numerical algorithms
- Applications of numerical algorithms in science and technology

Event Chairs

Bylina, Beata, Maria Curie-Sklodowska University, Poland

Bylina, Jarosław, Maria Curie-Sklodowska University, Poland

Stpiczyński, Przemysław, Maria Curie-Sklodowska University, Poland

Program Committee

Amodio, Pierluigi, Università di Bari, Italy

Anastassi, Zacharias, Qatar University, Qatar

Brugnano, Luigi, Universita' di Firenze, Italy

Czachorski, Tadeusz, Poland

Filippone, Salvatore, University Rome Tor Vergata, Italy

Gansterer, Wilfried, University of Vienna, Austria

Georgiev, Krassimir, IICT - BAS, Bulgaria

Gimenez, Domingo, University of Murcia, Spain

Gravvanis, George, Democritus University of Thrace, Greece

Kierzenka, Jacek, United States

Knottenbelt, William, Imperial College London, United Kingdom

Lirkov, Ivan, Institute of Information and Communication Technologies , Bulgarian Academy of Sciences, Bulgaria

Maksimov, Vyacheslav, Institute of Mathematics and Mechanics, Russia

Marowka, Ami, Bar-Ilan University, Israel

Mauro, Francaviglia, University of Torino, Italy

Petcu, Dana, West University of Timisoara, Romania

Pultarova, Ivana, Czech Technical University in Prague, Czech Republic

Rybka, Piotr, The University of Warsaw, Poland

Satco, Bianca-Renata, Stefan cel Mare University of Suceava, Romania

Sedukhin, Stanislav, The University of Aizu, Japan

Sergeichuk, Vladimir, Institute of Mathematics of NAS of Ukraine, Ukraine

Srinivasan, Natesan, Indian Institute of Technology, India

Szajowski, Krzysztof, Institute of Mathematics and Computer Science, Poland

Szyld, Daniel, Temple University, United States

Tokarzewski, Jerzy, Warsaw University of Technology, Poland

Tudruj, Marek, Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland

Ustimenko, Vasyl, Marie Curie-Sklodowska University, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 425–431

Mixed precision iterative refinement techniques for the WZ factorization

Beata Bylina Jarosław Bylina
Institute of Mathematics
Marie Curie-Sklodowska University
Pl. M. Curie-Skłodowskiej 5, 20-031 Lublin, Poland
beatas@hektor.umcs.lublin.pl
jmbylina@hektor.umcs.lublin.pl

***Abstract*—The aim of the paper is to analyze the potential of the mixed precision iterative refinement technique for the WZ factorization. We design and implement a mixed precision iterative refinement algorithm for the WZ factorization with the use of the single (a.k.a. float), double and long double precision. For random dense square matrices with the dominant diagonal we report the performance and the speedup of the solvers using different machines and we investigate the accuracy of obtained solutions. Additionally, the results (performance, speedup and accuracy) for our mixed precision implementation based on the WZ factorization were compared to the similar ones based on the LU factorization.**

I. Introduction

SOLUTION of linear systems of equations of the form:

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \text{where} \quad \mathbf{A} \in \mathbb{R}^{n\times n}, \quad \mathbf{b} \in \mathbb{R}^{n}, \qquad (1)$$

is an important and common problem in engineering and scientific computations. One of the direct methods of solving a dense linear system (1) is to factorize the matrix $\mathbf{A}$ into some simpler matrices — and then solving simpler linear systems. The most known factorization is the LU factorization. In this work we study another form of the factorization, namely the WZ factorization. In [5], [6], [7] we showed that there are matrices for which applying the incomplete WZ preconditioning gives better results than the incomplete LU factorization and we also showed the use of the WZ factorization for Markovian models.

One of quite known techniques to accelerate computations is the mixed precision iterative refinement. The iterative refinement is a well-known concept and it was analyzed by [11], [13], [12]. The mixed precision iterative refinement technique is used for the high performance computing [2], for example, for the solution of dense linear systems [3] — the LU factorization was considered among others, for accelerated block-asynchronous iteration methods [1].

The idea of such a refinement is that we perform the most time-consuming computations with the use of the low precision and then we improve the accuracy of the solution with the use of the high precision — by the iterative refinement. The mixed precision method uses properties of modern computer architectures where the single precision computations are about twice faster than the double precision ones — and the same can be observed for the memory access for both precisions.

Here we will modify the WZ solver to use the mixed precision approach. The aim of the paper is to analyze the potential of the mixed precision iterative refinement technique for the WZ factorization. We design and implement a mixed precision iterative refinement algorithm for the WZ factorization and compare its performance, speedup and accuracy with the pure implementation of the WZ factorization with the use of the float, double and long double precisions. We also compare it to an analogous LU solvers (pure ones and with the mixed precision).

The content of the paper is following. In Section II we describe the idea of the WZ factorization [8], [14] and the way the matrix $\mathbf{A}$ is factorized to a product of matrices $\mathbf{W}$ and $\mathbf{Z}$ — such a factorization exists for every nonsingular matrix (with pivoting) what was shown in [8].

Section III provides some mathematical background by outlining the idea of the iterative refinement algorithm.

In Section IV we describe the mixed precision iterative refinement technique for the WZ factorization. We present the algorithm for matrices which can be factorized without pivoting, for example, strictly diagonally dominant ones (as it was proved in [8]) and we give details about the implementation of the mixed precision iterative refinement for the WZ factorization.

In Section V we present the results of our experiments. We analyze the performance of our algorithm and its speedup. We study the influence of the size of the matrix on the achieved numerical accuracy.

Section VI is a summary of our experiments.

II. WZ Factorization

Here we describe shortly the WZ factorization usage to solve (1). The WZ factorization is described in [8], [10]. Assume that the $\mathbf{A}$ is a square nonsingular matrix. We are to find matrices $\mathbf{W}$ and $\mathbf{Z}$ that fulfill $\mathbf{WZ} = \mathbf{A}$ and the matrices $\mathbf{W}$ and $\mathbf{Z}$ consist of the following columns $\mathbf{w}_i$ and rows $\mathbf{z}_i^T$

This work was partially supported within the project N N516 479640 of the Ministry of Science and Higher Education of the Polish Republic (MNiSW) "Modele dynamiki transmisji, sterowania zatłoczeniem i jakością usług w Internecie".

978-83-60810-53-8/$25.00

respectively:

$$\begin{aligned}
\mathbf{w}_i &= (\underbrace{0,\ldots,0,1}_{i}, w_{i+1,i},\ldots,w_{n-i,i},0,\ldots,0)^T \\
&\quad \text{for} \quad i=1,\ldots,m,\\
\mathbf{w}_i &= (\underbrace{0,\ldots,0,1}_{i},0,\ldots,0)^T \\
&\quad \text{for} \quad i=p,q,\\
\mathbf{w}_i &= (\underbrace{0,\ldots,0}_{n-i+1}, w_{n-i+2,i},\ldots,w_{i-1,i},1,0,\ldots,0)^T \\
&\quad \text{for} \quad i=q+1,\ldots,n,\\
\mathbf{z}_i^T &= (\underbrace{0,\ldots,0}_{i-1}, z_{ii},\ldots,z_{i,n-i+1},0,\ldots,0) \\
&\quad \text{for} \quad i=1,\ldots,p,\\
\mathbf{z}_i^T &= (\underbrace{0,\ldots,0}_{i-1}, z_{i,n-i+1},\ldots,z_{ii},0,\ldots,0) \\
&\quad \text{for} \quad i=p+1,\ldots,n,
\end{aligned}$$

where

$$\begin{aligned}
m &= \lfloor (n-1)/2 \rfloor,\\
p &= \lfloor (n+1)/2 \rfloor,\\
q &= \lceil (n+1)/2 \rceil.
\end{aligned}$$

(see also Figure 1).

After the factorization we can solve two linear systems:

$$\begin{aligned}
\mathbf{W}\mathbf{y} &= \mathbf{b},\\
\mathbf{Z}\mathbf{x} &= \mathbf{y}
\end{aligned}$$

(where $\mathbf{c}$ is an auxiliary intermediate vector) instead of one (1).

III. Refinement

Let $\mathbf{x}_{cr}$ be the exact solution of the system (1):

$$\mathbf{A}\mathbf{x}_{cr} = \mathbf{b} \tag{2}$$

and $\mathbf{x}_{cm}$ be the machine-computed solution, thus with some rounding error — which we denote by $\mathbf{e}$ (all $\mathbf{x}_{cr}$, $\mathbf{x}_{cm}$, $\mathbf{e}$ are vectors of the size n).

Then, we can write:

$$\mathbf{x}_{cm} = \mathbf{x}_{cr} - \mathbf{e} \tag{3}$$

Let

$$\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}_{cm} \tag{4}$$

be a residual vector for the not exact solution $\mathbf{x}_{cm}$. Using (3) for $\mathbf{x}_{cm}$ in (4) we get:

$$\mathbf{r} = \mathbf{b} - \mathbf{A}(\mathbf{x}_{cr} - \mathbf{e})$$

and then (from (2)):

$$\mathbf{r} = \mathbf{A}\mathbf{e}. \tag{5}$$

Now, we can compute the error vector $\mathbf{e}$ from a linear system (5) and find a new, better solution of (1):

$$\mathbf{x}'_{cm} = \mathbf{x}_{cm} + \mathbf{e}.$$

However, the vector $\mathbf{e}$ as a solution of (5) is also prone to rounding errors, so the new $\mathbf{x}'_{cm}$ is also not exact — although better — and it can be further improved iteratively with the same process. This routine is known as the iterative refinement.

Algorithm 1 describes steps of such an iterative refinement for the solution of the linear system (1), with the use of the WZ factorization. The computational complexity is also given for every step. The stop condition is given by the infinity norm of the residual vector:

$$||\mathbf{r}||_\infty = \max_{1\le i\le n} |r_i|,$$

and the refinement stops when there is no further improvement (that is why we return $\mathbf{x}$, not $\mathbf{x}'$).

Algorithm 1 The iterative refinement technique for the WZ factorization

Require: $\mathbf{A}$, $\mathbf{b}$
Ensure: $\mathbf{x} \leftarrow \mathbf{A}^{-1}\mathbf{b}$

1:	$\mathbf{WZ} \leftarrow \mathbf{A}$	$\{O(n^3)\}$
2:	Solve the equation $\mathbf{W}\mathbf{y} = \mathbf{b}$	$\{O(n^2)\}$
3:	Solve the equation $\mathbf{Z}\mathbf{x} = \mathbf{y}$	$\{O(n^2)\}$
4:	$\mathbf{r} \leftarrow \mathbf{A}\mathbf{x} - \mathbf{b}$	$\{O(n^2)\}$
5:	$\varepsilon \leftarrow ||\mathbf{r}||_\infty$	$\{O(n)\}$
6:	**loop**	
7:	Solve the equation $\mathbf{W}\mathbf{y} = \mathbf{r}$	$\{O(n^2)\}$
8:	Solve the equation $\mathbf{Z}\mathbf{p} = \mathbf{y}$	$\{O(n^2)\}$
9:	$\mathbf{x}' \leftarrow \mathbf{x} + \mathbf{p}$	$\{O(n)\}$
10:	$\mathbf{r} \leftarrow \mathbf{A}\mathbf{x}' - \mathbf{b}$	$\{O(n^2)\}$
11:	$\varepsilon' \leftarrow ||\mathbf{r}||_\infty$	$\{O(n)\}$
12:	**if** $\varepsilon' \ge \varepsilon$: **return** $\mathbf{x}$	$\{O(1)\}$
13:	$\mathbf{x} \longleftrightarrow \mathbf{x}'$	$\{O(1)$, only references swapped$\}$
14:	$\varepsilon \leftarrow \varepsilon'$	$\{O(1)\}$
15:	**end loop**	

IV. Mixed Precision

The next algorithm, Algorithm 2, describes the use of the mixed precision technique with the iterative refinement of the solution. Every step is also labeled with its precision.

The only operations performed with the use of the double (or long double) precision are:

- matrix-vector operations with the complexity $O(n^2)$ — Steps 7 and 15 (computing the residual vector);
- vector operations with the complexity $O(n)$ — Steps 8, 16 (computing the norm) and 14 (applying the correction);
- scalar operations with the complexity $O(1)$ — Steps 17 and 19;
- conversions — almost all with the complexity $O(n)$ — the only exception ($O(n^2)$) is Step 1, but it is done only once, before the loop.

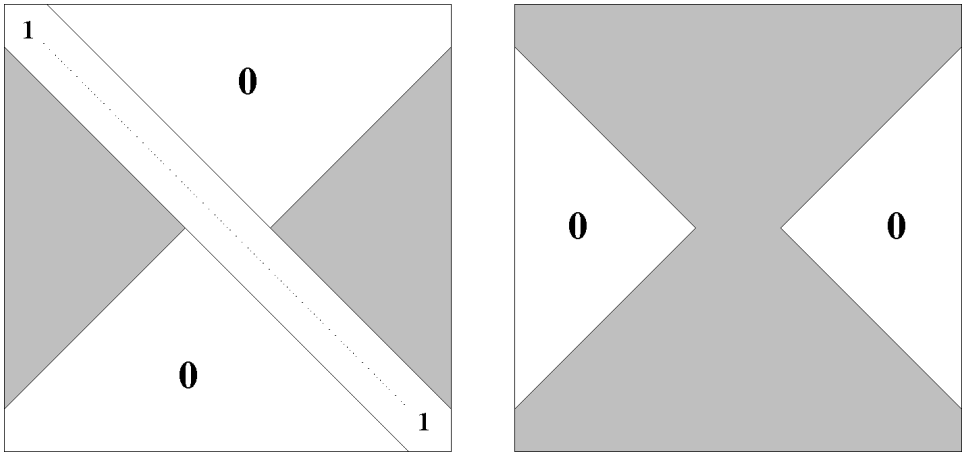


Fig. 1. Structures of the matrices $\mathbf{W}$ (left) and $\mathbf{Z}$ (right)

Algorithm 2 The mixed precision iterative refinement technique for the WZ factorization

Require: $\mathbf{A}_d$, $\mathbf{b}_d$
Ensure: $\mathbf{x}_d \leftarrow {\mathbf{A}_d}^{-1}\mathbf{b}_d$

1: $\mathbf{A}_s \leftarrow \mathbf{A}_d$ {conversion}
2: $\mathbf{b}_s \leftarrow \mathbf{b}_d$ {conversion}
3: $\mathbf{W}_s\mathbf{Z}_s \leftarrow \mathbf{A}_s$ {$O(n^3)$, single}
4: Solve the equation $\mathbf{W}_s\mathbf{y}_s = \mathbf{b}_s$ {$O(n^2)$, single}
5: Solve the equation $\mathbf{Z}_s\mathbf{x}_s = \mathbf{y}_s$ {$O(n^2)$, single}
6: $\mathbf{x}_d \leftarrow \mathbf{x}_s$ {conversion}
7: $\mathbf{r}_d \leftarrow \mathbf{A}_d\mathbf{x}_d - \mathbf{b}_d$ {$O(n^2)$, (long) double}
8: $\varepsilon \leftarrow ||\mathbf{r}_d||_\infty$ {$O(n)$, (long) double}
9: **loop**
10: $\mathbf{r}_s \leftarrow \mathbf{r}_d$ {conversion}
11: Solve the equation $\mathbf{W}_s\mathbf{y}_s = \mathbf{r}_s$ {$O(n^2)$, single}
12: Solve the equation $\mathbf{Z}_s\mathbf{p}_s = \mathbf{y}_s$ {$O(n^2)$, single}
13: $\mathbf{p}_d \leftarrow \mathbf{p}_s$ {conversion}
14: $\mathbf{x}'_d \leftarrow \mathbf{x}_d + \mathbf{p}_d$ {$O(n)$, (long) double}
15: $\mathbf{r}_d \leftarrow \mathbf{A}_d\mathbf{x}'_d - \mathbf{b}_d$ {$O(n^2)$, (long) double}
16: $\varepsilon' \leftarrow ||\mathbf{r}_d||_\infty$ {$O(n)$, (long) double}
17: **if** $\varepsilon' \geq \varepsilon$: **return** $\mathbf{x}_d$ {$O(1)$, (long) double}
18: $\mathbf{x}_d \longleftrightarrow \mathbf{x}'_d$ {$O(1)$, only references swapped}
19: $\varepsilon \leftarrow \varepsilon'$ {$O(1)$, (long) double}
20: **end loop**

All the other computations are single precision ones.

We denote the matrices and vectors stored in the (long) double precision with the $_d$ subscript and the matrices and vectors stored in the single precision with the $_s$ subscript. So, the input data are the matrix $\mathbf{A}_d$ and the vector $\mathbf{b}_d$ ((long) double precision). The output vector $\mathbf{x}_d$ is also stored in the (long) double precision.

The coefficient matrix $\mathbf{A}_d$ is converted to the single precision for the WZ factorization and denoted as $\mathbf{A}_s$. Some vectors in the algorithm are also converted between single and (long) double precision (that is why we have: $\mathbf{b}_d$ and $\mathbf{b}_s$; $\mathbf{x}_d$ and $\mathbf{x}_s$; $\mathbf{r}_d$ and $\mathbf{r}_s$; $\mathbf{p}_d$ and $\mathbf{p}_s$).

The method used in Algorithm 2 can give a significant improvement for the solution of a linear system, because the cost of every iteration is very small comparing to the cost of the factorization.

The disadvantage of this approach is a lot larger memory requirement — a great deal of data are to be duplicated in both precisions. It consumes up to 50% more memory than it is used in usual double precision solution.

V. Numerical Experiments

In the experiment, we analyze how the use of the mixed precision iterative refinement techniques for the WZ factorization influences the performance, the speedup and the accuracy of the WZ solver for the linear equation system. In this section we test the performance, the speedup and the accuracy on three devices of different architectures. Additionally, we compare properties of the WZ solver with the LU solver. For both kinds of factorization we consider five implementations:

- a traditional single precision implementation;
- a traditional double precision implementation;
- a traditional long double precision implementation;
- a mixed precision implementation with double precision refinement (denoted as mix(double));
- a mixed precision implementation with long double precision refinement (denoted as mix(long double)).

The former three are made according to Algorithm 1 and the latter two are made according to Algorithm 2. All the implementations were sequential (single-threaded).

All the computations were carried out for dense random matrices with a dominant diagonal. The sizes of the matrices were from 500 up to 9000.

A. Environment

The architectures used for tests are shown in Table I. All the machines worked under the Debian GNU/Linux 6.0 operating system and the programs were compiled with the use of the GCC compiler (ver. 4.7.2, compiler command: `g++ -O3`).

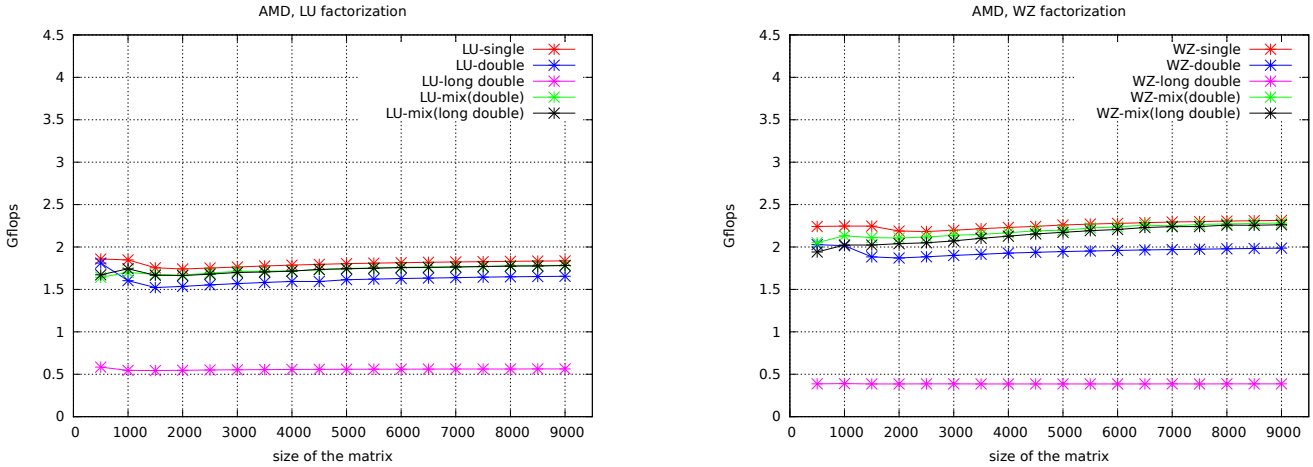


Fig. 2. The performance of the LU (left) and WZ (right) solver on the AMD architecture

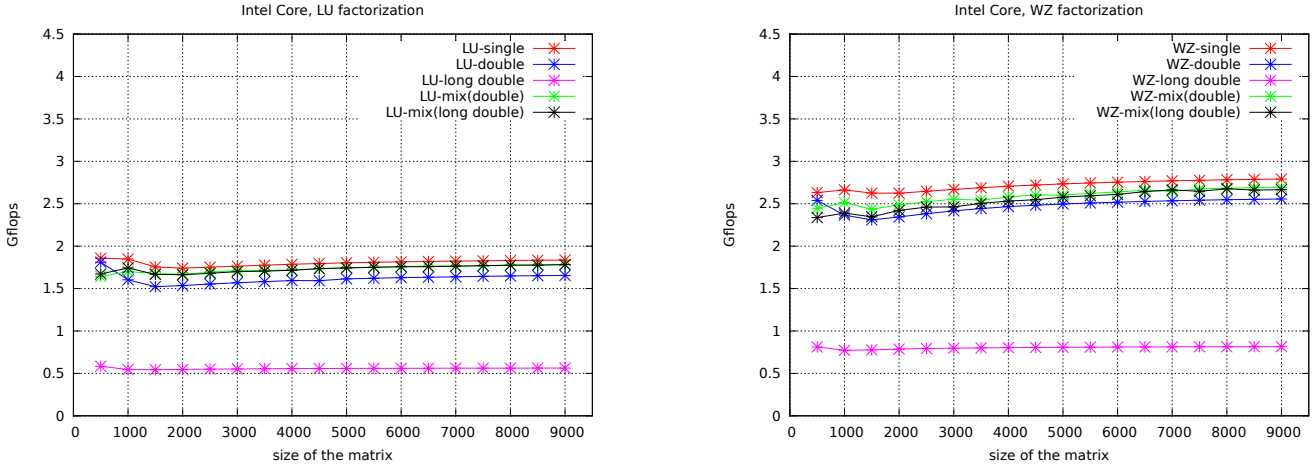


Fig. 3. The performance of the LU (left) and WZ (right) solver on the Intel Core architecture

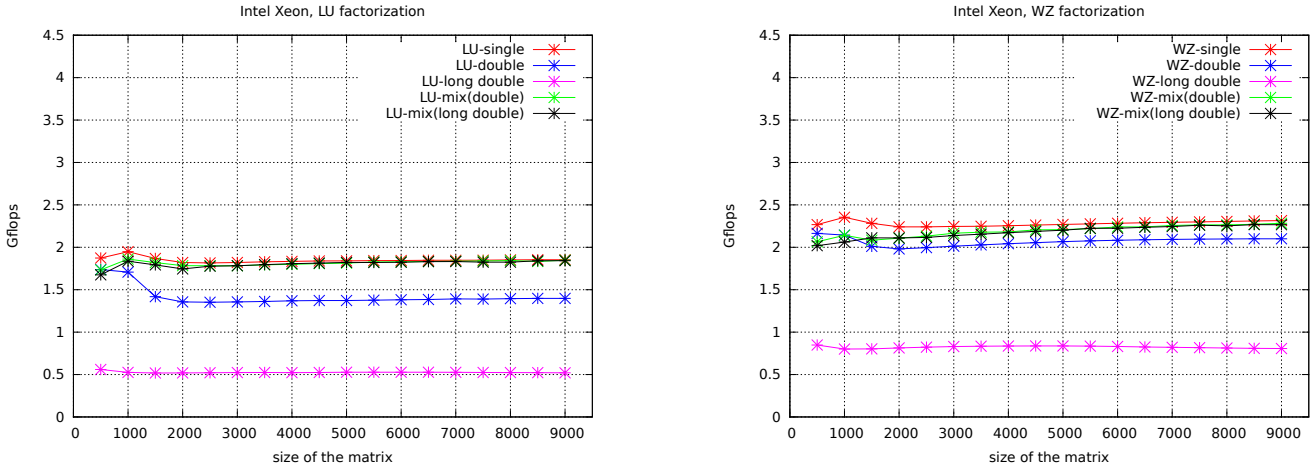


Fig. 4. The performance of the LU (left) and WZ (right) solver on the Intel Xeon architecture

B. Performance

Figures 2, 3, 4 show the performance of the single-core implementations of the LU and WZ solvers on the given architectures. The performance is based on the number of floating-point operations in the LU solver $\left(\frac{2}{3}n^3 + \frac{1}{2}n^2 - \frac{7}{6}n\right)$.

We see that:

- the architecture has almost no impact on the performance — however, not mixed implementations on AMD are somewhat slower;
- the size of the matrix has no impact on the performance, either;
- the WZ solver performs better than the LU solver;
- the long double implementation is always the slowest, the others perform quite similarly — even (what is very important) the mix(long double) implementation.

C. Speedup

Figures 5, 6, 7 show the speedup of the same implementations. We labeled our speedups by S(F)-P where:

- F $\in \{$LU, WZ$\}$ is the kind of the factorization used;
- P $\in \{$double, long double$\}$ is the precision of the result.

Thus, S(F)-P denotes the speedup achieved by the mix(P) implementation over the P implementation — while both are conducted with the use of the F factorization.

We see that:

- the architecture has some impact on the speedup: the best is for the AMD architecture — because that architecture gives somewhat slower performance for double and long double implementations;
- for very small problem sizes, the cost of even a few iterative refinement iterations is high compared to the cost of the factorization and thus, the mix(double) implementations are less efficient than the double ones;
- the LU solvers have higher speedups than the WZ solvers for all architectures — because the original LU solver performs slightly worse than the WZ one;
- if the problem size is big enough, the mix(double) implementation can provide a speedup of up to 1.3 and the mix(long double) — even up to 7.

D. Accuracy

Figures 8, 9, 10 show the accuracy of the implementations. We define the measure of the accuracy as

$$\text{accu} = -\log_{10} ||\mathbf{Ax} - \mathbf{b}||_\infty.$$

We see that:

TABLE I
HARDWARE PROPERTIES OF THE TEST MACHINES

AMD	CPU	AMD FX-8120 3.1 GHz
	Host memory	16 GB
Intel Core	CPU	Intel Core I7 2670QM 2.2 GHz
	Host memory	8 GB
Intel XEON	CPU	Intel Xeon X5650 2.67GHz
	Host memory	48 GB

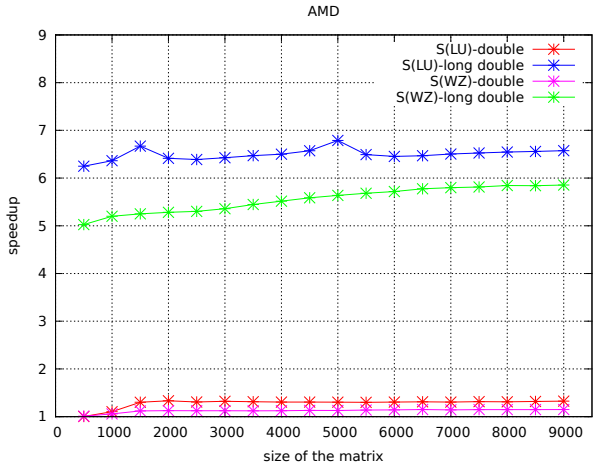


Fig. 5. The speedup of the LU and WZ solvers on the AMD

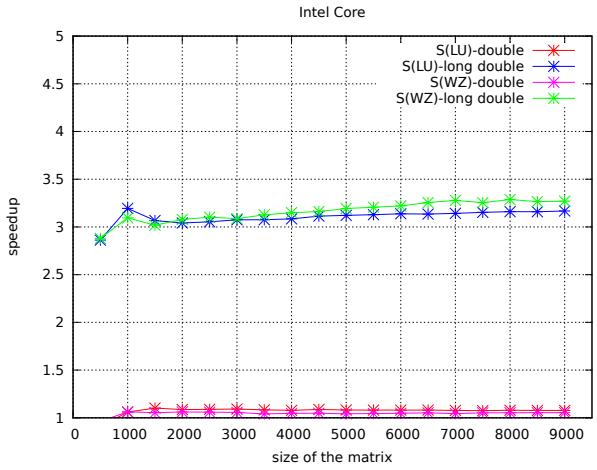


Fig. 6. The speedup of the LU and WZ solvers on the Intel Core

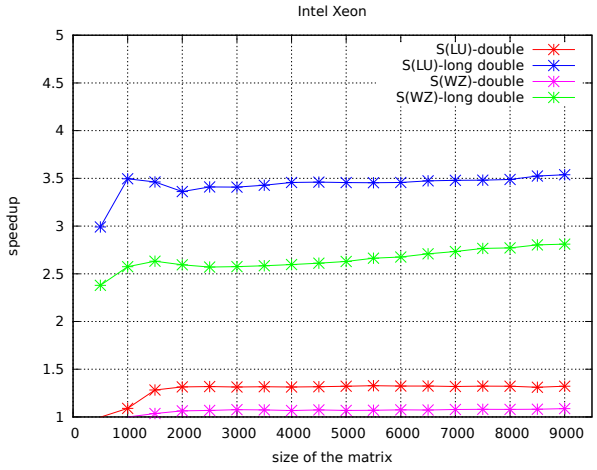


Fig. 7. The speedup of the LU and WZ solvers on the Intel Xeon

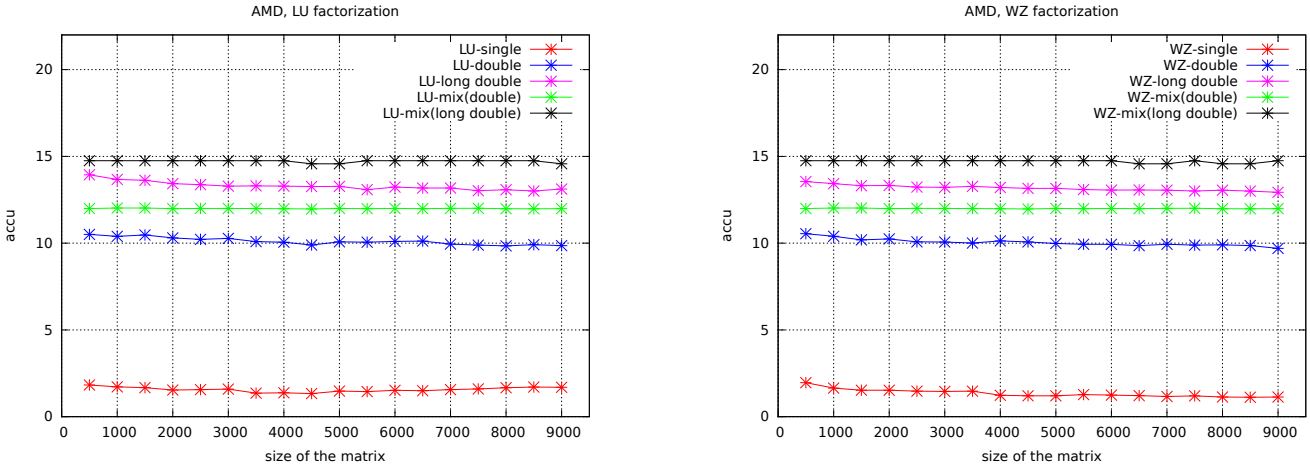


Fig. 8. The accuracy of the LU (left) and WZ (right) solver on the AMD architecture

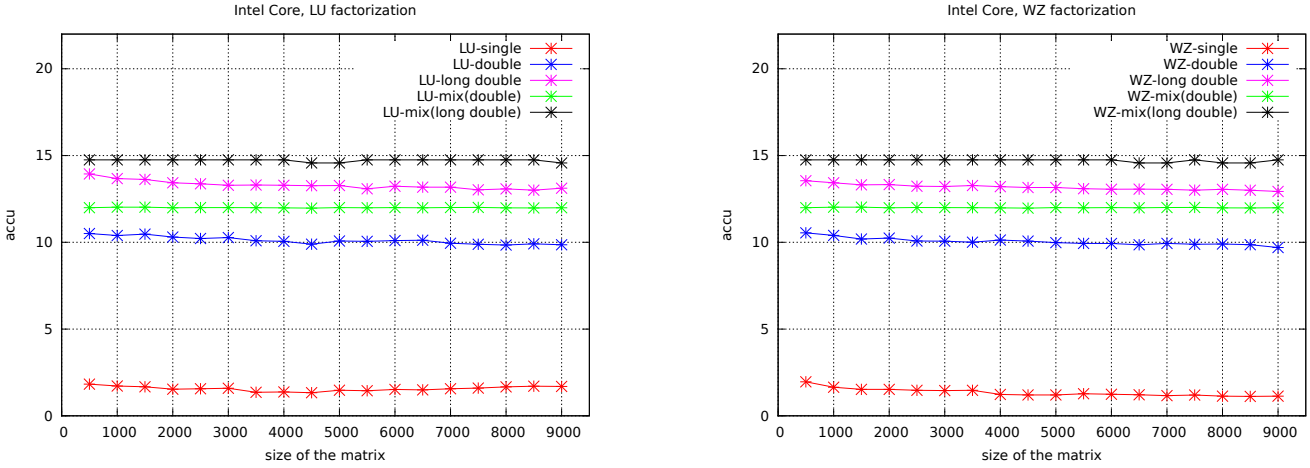


Fig. 9. The accuracy of the LU (left) and WZ (right) solver on the Intel Core architecture

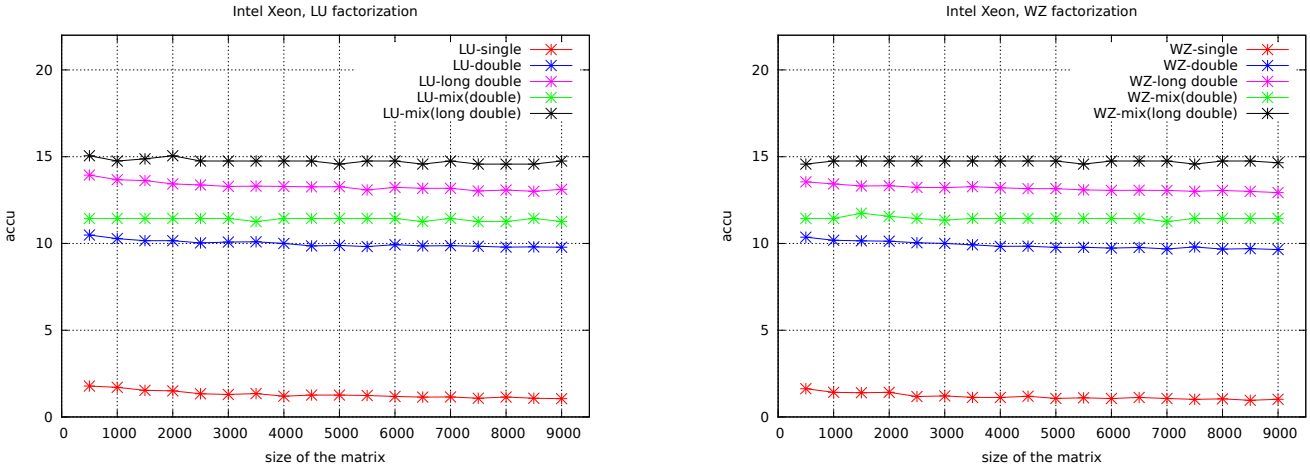


Fig. 10. The accuracy of the LU (left) and WZ (right) solver on the Intel Xeon architecture

- the architecture has no impact on the accuracy;
- the size of the matrix and the type of the factorization has almost no impact on the accuracy, either;
- the worst accuracy we get is (of course) for the single implementation; the best (about 10^{-15}) — for the long double one;
- the mixed precision significantly improves the accuracy.

The number of iterations needed for our mixed precision method to outdo the accuracy of the (long) double precision solver is not too high and is about 5–6 iterations, somewhat less on the AMD architecture (about 3–4 iterations).

VI. Conclusion

In this article we described an iterative refinement algorithm for the WZ solver with the use of the mixed precision technique and investigated properties of this new algorithm. We compared the mixed precision iterative refinement for the WZ factorization with a similar algorithm for the LU factorization.

Both the types of algorithms gave similar accuracy. However, the performance was better for the WZ solvers but the higher speedup was achieved for the LU solver.

These experiments show that the mixed precision iterative refinement method can run faster than the (long) double precision solver — delivering the same (or even better) accuracy as the (long) double precision one for both the factorizations. Moreover, the experiments also show that the mixed precision iterative refinement method for the long double precision solver is much faster than the traditional one (up to 7 times) — with the same or better accuracy.

The results do not depend significantly on the size of the matrix.

The approach presented here causes a significant acceleration of solving the linear systems with the use of direct methods and we think that the similar problems on different architectures (as GPU, for example) could be also improved.

References

[1] H. Anzt, P. Luszczek, J. Dongarra, V. Heuveline: GPU-Accelerated Asynchronous Error Correction for Mixed Precision Iterative Refinement, *Euro-Par 2012*, pp. 908–919.

[2] M. Baboulin, A. Buttari, J. J. Dongarra, J. Langou, J. Langou, P. Luszczek, J. Kurzak, S. Tomov: Accelerating scientific computations with mixed precision algorithms, *Computer Physics Communications* 180(12) (2009), pp. 2526–2533.

[3] A. Buttari, J. J. Dongarra, J. Langou, J. Langou, P. Luszczek, J. Kurza: Mixed precision iterative refinement techniques for the solution of dense linear systems. *Int. J. of High Performance Computing and Applications* 21(4) 2007, pp. 457–466.

[4] B. Bylina, J. Bylina: Analysis and Comparison of Reordering for Two Factorization Methods (LU and WZ) for Sparse Matrices, *Lecture Notes in Computer Science* **5101**, Springer-Verlag Berlin Heidelberg 2008, pp. 983–992.

[5] B. Bylina, J. Bylina: Incomplete WZ Factorization as an Alternative Method of Preconditioning for Solving Markov Chains, *Lecture Notes in Computer Science* **4967**, Springer-Verlag Berlin Heidelberg 2008, 99–107.

[6] B. Bylina, J. Bylina: Influence of preconditioning and blocking on accuracy in solving Markovian models, *International Journal of Applied Mathematics and Computer Science* 19 (2) (2009), pp. 207–217.

[7] B. Bylina, J. Bylina: The Vectorized and Parallelized Solving of Markovian Models for Optical Networks, *Lecture Notes in Computer Science* **3037**, Springer-Verlag Berlin Heidelberg 2004, 578–581.

[8] S. Chandra Sekhara Rao: Existence and uniqueness of WZ factorization, *Parallel Computing* **23** (1997), pp. 1129–1139.

[9] D. J. Evans, M. Barulli: BSP linear solver for dense matrices, *Parallel Computing* **24** (1998), pp. 777–795.

[10] D. J. Evans, M. Hatzopoulos: The parallel solution of linear system, *Int. J. Comp. Math.* **7** (1979), pp. 227–238.

[11] C. B. Moler: Iterative refinement in floating point. *J. ACM* **14(2)** (1967), pp. 316–21.

[12] G. W. Stewart: Introduction to Matrix Computations, Academic Press, 1973.

[13] J. H. Wilkinson: Rounding Errors in Algebraic Processes, Prentice-Hall, 1963.

[14] P. Yalamov, D. J. Evans: The WZ matrix factorization method, *Parallel Computing* **21** (1995), pp. 1111–1120.

[15] `http://software.intel.com/en-us/articles/intel-mkl/`

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 433–440

Surface Reconstruction from Scattered Point via RBF Interpolation on GPU

Salvatore Cuomo*, Ardelio Galletti†, Giulio Giunta†, Alfredo Starace†
*Department of Mathematics and Applications "R. Caccioppoli" University of Naples Federico II
c/o Universitario M.S. Angelo 80126 Naples Italy
email:salvatore.cuomo@unina.it
†Department of Applied Science. University of Naples "Parthenope".
Centro Direzionale, Isola C4 80143 Naples Italy
emails:{ardelio.galletti,giulio.giunta}@uniparthenope.it, alfredo.starace@gmail.com

***Abstract*—In this paper we describe a parallel implicit method based on radial basis functions (RBF) for surface reconstruction. Practical applicability of RBF methods is hindered by their high computational demand, that requires the solution of linear systems of size equal to the number of data points. The implementation of our method relies on parallel scientific libraries and is designed for exploiting Graphic Processor Units (GPUs) acceleration. The performance of the proposed method in terms of accuracy of the reconstruction and computing time shows that RBF interpolation can be very effective for large scale surface reconstruction problems.**

I. Introduction

MANY applications in engineering and science need to build accurate digital models of real-world objects defined in terms of *point cloud data*, *i.e.* a set of scattered points in 3D. Typical examples include the digitalization of manufactured parts for quality control, statues and artifacts in archeology and arts [12], human bodies for movies or video games, organs and anatomical parts for medical diagnostic [4] and terrain elevation models for simulations and modeling [16]. Modern 3D scanners are able to acquire point clouds containing millions of points sampled from an object. The process of building a geometric model from such point clouds is usually referred to as *surface reconstruction*.

There are several approaches to reconstruct surfaces from 3D scattered datasets. Generally, such methods fall into two categories [17]: Delaunay-based methods and implicit surface methods. Delaunay triangulation, and other related approaches like Voronoi diagrams, are widely used in digital elevation modeling, and their core numerical problem is a nearest neighbor interpolation [2]. Implicit surface modeling is mostly popular in describing complex shapes and interactive graphical operations. Level set methods [24], moving least square methods [11], variational implicit surfaces [21] and adaptively sampled distance field [9] are recent developments in that field. In this paper, we present an implicit surface method based on radial basis functions (RBFs). In the 1980's, Franke [8] firstly used radial basis functions to interpolate scattered point cloud and proved accuracy and stability of such methods. In this approach, an implicit surface is constructed by calculating the weights of a linear combination of a set of radial basis functions that interpolates the given data points.

Practical applicability of RBF methods is hindered by their computational demand, since they require the solution of a linear system of size equal to the number of data points and current 3D data scanners allow the acquisition of tens of millions points of an object surface.

High Performance Computing is a natural solution to provide the computational power required in such large scale problems [15]. Here, we propose an RBF surface reconstruction method designed for a massively multi-core architecture, namely Graphics Processing Units (GPUs) [18]. Recently, GPUs have been effectively exploited to improve the performance of software tools in several scientific applications, such as computational fluid dynamics, molecular dynamics, climate modeling [6], [7], [10]. The core of our method is the construction of an RBF interpolant. We are not aware of RBF interpolation algorithms for GPU. To our knowledge, the most efficient parallel algorithm for RBF interpolation on multiprocessor clusters is PetRBF [22]. PetRBF exhibits $\mathcal{O}(N)$ complexity, requires $\mathcal{O}(N)$ storage, and scales excellently up to a thousand processes. Our proposed method relies on PetRBF and one of our main contributions consist in developing an improved version of PetRBF for surface reconstruction and GPU acceleration. Moreover, we show how to make a suitable choice of the algorithm parameters for accurate reconstruction from synthetic, real or incomplete datasets. The paper is organized as follows. In Section II we deal with some mathematical preliminaries about implicit surfaces and RBF interpolation. In Section III, first we briefly recall main features of PetRBF, then we describe our method and illustrate our GPU implementation strategy. Section IV contains a discussion on the results of numerical experiments for assessing accuracy and performance of the method. Final conclusions are reported in Section V.

II. Preliminaries

In this section we recall basic ideas underlying implicit surface reconstruction and define the related RBF interpolation problem.

978-83-60810-53-8/$25.00

A. Implicit Surface Reconstruction

Given a point cloud

$$\mathcal{X} := \{(x_i, y_i, z_i) \in \mathbb{R}^3, i = 1, \ldots, N\}$$

belonging to an unknown surface $\mathcal{M}$, *i.e.* $\mathcal{X} \subset \mathcal{M}$, the goal is to find another surface $\mathcal{M}^*$ which is a reconstruction of $\mathcal{M}$. In the implicit surface approach, $\mathcal{M}$ is defined as the surface of all points $(x, y, z) \in \mathbb{R}^3$ that satisfy the implicit equation

$$f(x, y, z) = 0 \tag{1}$$

for an unknown function f. A way to approximate f is to impose the interpolation conditions (1) on the point cloud $\mathcal{X}$. However, the use of those interpolation conditions only leads to the trivial solution given by the identically zero function, whose zero surface is $\mathbb{R}^3$. Therefore, the key for finding an approximation of the function f is to use additional significant interpolation conditions, *i.e.* involving from off-surface points (where $f \neq 0$). This ensures the existence of a non trivial interpolant $\mathcal{P}_f$, whose zero surface contains a meaningful surface $\mathcal{M}^*$. This approach leads to a surface reconstruction method which consists of three main steps:

1) off-surface points generation;
2) interpolant model identification on the extended dataset;
3) computation of the zero iso-surface of the interpolant.

1) Off-surface points generation:

A common practice [20] is to consider the set of surface normals $\mathbf{n}_i = (n_i^x, n_i^y, n_i^z)$ to the surface $\mathcal{M}$ at points $\mathbf{x}_i = (x_i, y_i, z_i)$. If these normals are not explicitly known, there are techniques and tools[1] that allow to estimate them. Given the oriented surface normals ($\mathbf{n}_i$ and $-\mathbf{n}_i$), the extra off-surface points can be generated by marching a small distance δ along the normals. So for each data point, $\mathbf{x}_i = (x_i, y_i, z_i)$ two additional off-surface points are obtained. The first lies "outside" the surface $\mathcal{M}$ and is given by

$$\begin{aligned}(x_{N+i}, y_{N+i}, z_{N+i}) &= \mathbf{x}_i + \delta \mathbf{n}_i = \\ &= (x_i + \delta n_i^x, y_i + \delta n_i^y, z_i + \delta n_i^z);\end{aligned}$$

the second point lies "inside" and is given by

$$\begin{aligned}(x_{2N+i}, y_{2N+i}, z_{2N+i}) &= \mathbf{x}_i - \delta \mathbf{n}_i = \\ &= (x_i - \delta n_i^x, y_i - \delta n_i^y, z_i - \delta n_i^z).\end{aligned}$$

The union of the sets $\mathcal{X}_\delta^+ = \{\mathbf{x}_{N+1}, \ldots, \mathbf{x}_{2N}\}$, $\mathcal{X}_\delta^- = \{\mathbf{x}_{2N+1}, \ldots, \mathbf{x}_{3N}\}$, and $\mathcal{X}$ gives the overall set of points on which the interpolation conditions are assigned (see Fig. 1). The set $\mathcal{X}_\delta^+$ implicitly defines a surface $\mathcal{M}_\delta^+$ which passes through its points. Analogously, $\mathcal{X}_\delta^-$ defines the surface $\mathcal{M}_\delta^-$. Those two surfaces can be considered as external and internal to $\mathcal{M}$, respectively. The value of δ represents a small step size, whose specific magnitude may be rather critical for a good surface reconstruction [3]. In particular, if δ is chosen too large, this results in self intersecting $\mathcal{M}_\delta^+$ or $\mathcal{M}_\delta^-$ auxiliary surfaces. In our implementation we fix δ to 1% of the side of the bounding box of the data, as suggested in [23].

[1]package ply.tar.gz provided by Greg Turk, available at http://www.cc.gatech.edu/projects/large_models/ply.html

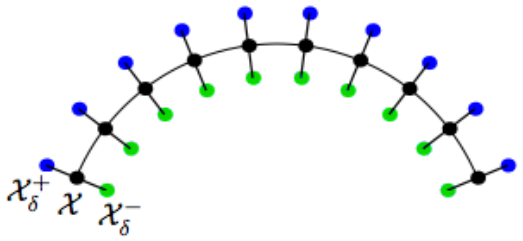


Fig. 1. Extended data set. In black points from $\mathcal{X}$, in blue points from $\mathcal{X}_\delta^+$ and in green points from $\mathcal{X}_\delta^-$

2) Interpolant model identification on the extended dataset:

This step consists in determining a function $\mathcal{P}_f$ whose zero contour (iso-surface $\mathcal{P}_f = 0$) interpolates the given point cloud data $\mathcal{X}$, and whose iso-surface $\mathcal{P}_f = 1$ and $\mathcal{P}_f = -1$ interpolate $\mathcal{X}_\delta^+$ and $\mathcal{X}_\delta^-$, respectively, *i.e.*

$$\mathcal{P}_f(x_i) = \begin{cases} 0 & i = 1, \ldots, N \\ 1 & i = N+1, \ldots, 2N \\ -1 & i = 2N+1, \ldots, 3N \end{cases} \tag{2}$$

The values of ± 1 for the auxiliary data are assigned in an arbitrary way. Such choice does not affect the quality of the results. Here we are interested to the zero iso-surface of $\mathcal{P}_f$.

3) Computation of the zero iso-surface of the interpolant:

In order to evaluate the $\mathcal{P}_f$ zero iso-surface and visualize it, we simply evaluate the interpolant $\mathcal{P}_f$ on a dense grid in a bounding box containing the point cloud. This approach leads to some undesired artifacts, since there are points in the bounding box which do not belong to $\mathcal{M}^*$. A way to overcome this drawback and display only $\mathcal{M}^*$ consists in evaluating the interpolant in a small *surrounding* volume of the surface $\mathcal{M}$. This set is denoted as $\mathcal{M}_{ext}^\varepsilon = \{\mathbf{x} \in \mathbb{R}^3 : \ d(\mathbf{x}, \mathcal{M}) \leq \varepsilon\}$, where $d(\mathbf{x}, \mathcal{M}) = \inf_{\mathbf{y} \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}\|$. For a small enough value of ε, it holds that

$$\mathcal{M}^* \approx \mathcal{M}_{ext}^\varepsilon \cap \mathcal{S}_0,$$

where $\mathcal{S}_0$ is the zero iso-surface of $\mathcal{P}_f$.

B. RBF interpolation

Given a set of N distinct points $(x_j,\ y_j)$, $j = 1, \ldots, N$, where $x_j \in \mathbb{R}^s$ and $y_j \in \mathbb{R}$, the scattered data interpolation problem consists in finding an interpolant function $\mathcal{P}_f$ such that:

$$\mathcal{P}_f(x_j) = y_j,\ j = 1, \ldots, N. \tag{3}$$

In the univariate setting ($s = 1$), the interpolant $\mathcal{P}_f$ is usually chosen in a suitable function space. A common approach assumes the function $\mathcal{P}_f$ as a linear combination of certain basis functions B_j

$$\mathcal{P}_f(x) = \sum_{j=1}^{N} c_j B_j(x). \tag{4}$$

In a multivariate setting ($x_j \in \mathbb{R}^s,\ s > 1$), the problem is more complex. As stated by the *Mairhuber-Curtis* theorem [5], [13], in order to have a well-posed multivariate scattered data interpolation problem, it is not possible to fix in advance the basis $\{B_1, \ldots, B_N\}$, since the basis functions must depend on the data sites x_j.

The data dependent space for RBF interpolation can be easily generated by means of the radial functions:

$$B_j \equiv \Phi_j = \varphi(\|x - x_j\|).$$

The points x_j to which the basic function φ is shifted are usually referred to as *centers*. While there may be circumstances that suggest to choose these centers different from the data sites one generally picks the centers to coincide with the data sites.

In fact, a practical interpolation problem consists of two subproblems: finding the interpolant $\mathcal{P}_f$ and evaluating it on an assigned set of points. The coefficients c_j in (4) are obtained by imposing the interpolation conditions (3)

$$\mathcal{P}_f(x_i) = \sum_{j=1}^{N} c_j \varphi(\|x_i - x_j\|) = y_i,\ i = 1, \ldots, N.$$

This leads to solve the linear system of equations $Ax = b$ in (5).

Given a set of M points $\xi = \{\xi_1, \xi_2, \ldots, \xi_M\}$, the evaluation of the interpolant $\mathcal{P}_f$ on ξ can be computed as a matrix-vector product (6).

It is well known that in order to have a well-posed problem (5), the matrix A must be non-singular. Unfortunately, a complete characterization of the class of all basic functions φ that generate a non-singular matrix for an arbitrary set $\mathcal{X} = \{x_1, \ldots, x_N\}$ of distinct data sites is still lacking. The situation gets better in case of *positive definite matrices*, that are always non-singular. Popular radial basis functions Φ_j, that give rise to positive definite interpolation matrices, are summarized in Table II-B. We focused our work on the gaussian function, as in [22].

III. GPU Parallel Surface Reconstruction

A brief description of the surface reconstruction algorithm is reported below:

Algorithm 1 Surface Reconstruction

Requirements:
point cloud $\mathcal{X}$, surface normals $\mathbf{n}_i$,
evaluation grid ξ

1: compute extended data set:
$\mathcal{X}_{ext} = \mathcal{X} \bigcup \mathcal{X}_\delta^+ \bigcup \mathcal{X}_\delta^-$ by using $\mathbf{n}_i$;
2: find the interpolant $\mathcal{P}_f$ on $\mathcal{X}_{ext}$;
3: evaluate $\mathcal{P}_f$ on ξ;
4: render the surface;

Steps 1 and 2 have been already discussed in §II.A. Step 3 requires a matrix-vector multiplication as stated in §II.B and the rendering step can be simply accomplished using either the MATLAB `isosurface` feature or other specific tools. The most computationally expensive step is the second one, that requires the solution of a system of $3N$ linear equation, where N is the initial point cloud size. In the following, we describe a parallel scheme for solving this RBF interpolation linear system.

A. Parallelization scheme

In handling problems with a large number of data points, as in surface reconstruction from clouds of millions of points, the large amount of memory storage can become a critical point. As the problem size grows, parallelization on distributed memory architectures becomes necessary. Domain Decomposition is a useful parallelization strategy for large scale interpolation, since the solution of the original system is built up by solving a set of smaller subproblems that interact through their interfaces. Below, we briefly overview the parallelization strategy of PetRBF, in a 3D setting. Let Ω be a 3D domain containing the point cloud and let partition Ω in overlapping sub-domains $\Omega_k,\ k = 1, \ldots, S$. Moreover, let $\tilde{\Omega}_k$ denote the empty intersection portions of subdomains Ω_k (see Fig. 2). The solution of the RBF interpolation linear system

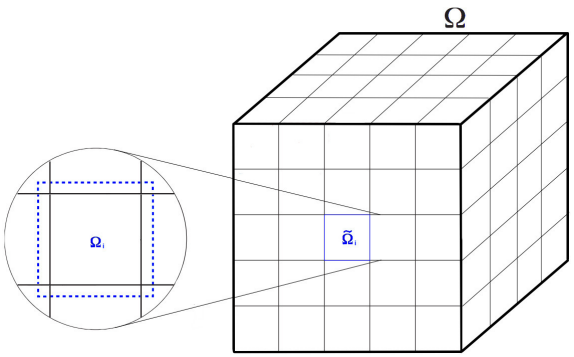


Fig. 2. Illustration of the Domain Decompostion Method.

$Ax = b$ on the whole domain Ω can be obtained through an iterative method consisting in sequentially solving, for any Ω_k, the linear (sub)system $A_k x_{\Omega_k} = b_{\Omega_k}$, where A_k, x_{Ω_k} and b_{Ω_k} are the entries in A, x, and b belonging to the sub-domain Ω_k, respectively. If at each iteration the solution on the entire domain is updated simultaneously after the solutions on every sub-domain are computed, we have an *additive* Schwarz method. Moreover, if the entries of x_{Ω_k} which are outside of the subdomain $\tilde{\Omega}_k$ are discarded after the calculation on each subdomain Ω_k, a *restricted additive* Schwarz method (RASM) in defined. RASM is known to converge faster than additive Schwarz, and requires less communication in a parallel setting. Notice that solving smaller systems of equations has the same effect of a preconditioning technique, and then RASM can be used in combination with any iterative method. PetRBF uses a Krylov subspace methods, namely the Generalized Minimum Residual (GMRES).

If the basis functions exhibit negligible global effects then A can be considered a band matrix, so that the matrix-vector

$$\underbrace{\begin{bmatrix} \varphi(\|x_1-x_1\|) & \varphi(\|x_1-x_2\|) & \cdots & \varphi(\|x_1-x_N\|) \\ \varphi(\|x_2-x_1\|) & \varphi(\|x_2-x_2\|) & \cdots & \varphi(\|x_2-x_N\|) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(\|x_N-x_1\|) & \varphi(\|x_N-x_2\|) & \cdots & \varphi(\|x_N-x_N\|) \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{b} \tag{5}$$

$$\begin{bmatrix} \mathcal{P}_f(\xi_1) \\ \mathcal{P}_f(\xi_2) \\ \vdots \\ \mathcal{P}_f(\xi_M) \end{bmatrix} = \begin{bmatrix} \varphi(\|\xi_1-x_1\|) & \varphi(\|\xi_1-x_2\|) & \cdots & \varphi(\|\xi_1-x_N\|) \\ \varphi(\|\xi_2-x_1\|) & \varphi(\|\xi_2-x_2\|) & \cdots & \varphi(\|\xi_2-x_N\|) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(\|\xi_M-x_1\|) & \varphi(\|\xi_M-x_2\|) & \cdots & \varphi(\|\xi_M-x_N\|) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} \tag{6}$$

TABLE I
EXAMPLES OF RADIAL BASIS FUNCTIONS.

RBF	Φ	
Poisson radial function	$\frac{J_{s/2-1}(\|x-x_j\|)}{\|x-x_j\|^{s/2-1}}$	$s \geq 2$
Inverse Multiquadric	$(1+\|x-x_j\|^2)^{-\beta}$	$\beta > \frac{s}{2}$
Matèrn function	$\frac{K_\alpha(\|x-x_j\|)\|x-x_j\|^\alpha}{2^{\beta-1}\Gamma(\beta)}$	$\alpha = \beta - \frac{s}{2} > 0$
Whittaker function	$\int_0^{+\infty}(1-\|x-x_j\|)_+^{k-1}t^\alpha e^{-\beta t}dt$	$k = 2,3,\ldots, \quad \alpha = 0,1,\ldots$
Gaussian function	$\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-\|x-x_j\|^2}{2\sigma^2}}$	$\sigma > 0$

product, which is the predominant operation in GMRES, can be computed somewhat locally. Using a gaussian function as the basic function, it holds that A has the following entries:

$$A_{ij} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}\right). \tag{7}$$

Since the gaussian function decays rapidly, for a suitable choice of σ, the entries of A in (6) which correspond to the interaction of distant points can be neglected. Such resulting sparsity of A depends on the relative size of the domain compared to the standard deviation σ of the gaussian. If σ is kept constant while the size of the domain increases with N, the calculation load will scale as $\mathcal{O}(N)$. The communication required to compute $A_k x_{\Omega_k}$ is also restricted to a constant number of entries corresponding to neighbor points. Therefore, RASM entails a high level of parallelism at each GMRS iteration, and a good scalability for surface reconstruction problems with a domain size as large as hundreds (or even thousands) of σ [22].

A remarkable implementation feature of PetRBF is the use of the parallel numerical library PETSc (Portable, Extensible Toolkit for Scientific Computation) [1]. All vectors and matrices can be distributed via PETSc routines in a such way that each process stores only a local portion of data. In particular, this can be done by defining x and b as `Vec` PETSc objects. Overlapping and non-overlapping sub-domains can be handled by means of index sets (`IS`). A PETSc `IS` object is a global index that identifies the entries in each sub-domain and is distributed among the processes in the same way as the vectors. Recalling that the interpolation matrix has entries which depends only on the vector x of data sites and σ in the gaussian (eq. 7), it turns out that A can be represented as a `MatShell` PETSc object, that allows to perform matrix computations in a matrix-free way, i.e. without actually storing the matrix. Inner products, norms, and scalar products are computed by specific PETSc routines, and the solution of the linear system is obtained by calling the `KSPSolve` PETSc solver.

B. GPU implementation

Our basic idea consists in using the new PETSc GPU feature in the latest version of PETSc [1]. In fact, we have developed an improved version of PetRBF that extends the original code to GPU environments. GPU support to PETSc has been introduced by means of the CUDA framework and exploits the open source libraries THRUST [19] and CUSP [14]. THRUST is a collection of data parallel primitives that provide high level abstractions to describe efficient computations on GPU. CUSP is a sparse linear algebra toolkit for performing matrix operations and solving linear systems via CUDA on GPUs. They allows a transparent access to the GPU, without radically changes to the existing source code of PETSc. In PETSc GPU, two new subclasses `VecCUSP` and `MatCUSP` of the `Vector` and `Matrix` classes have been defined, which in turns rely on CUBLAS, CUSP, and THRUST routines to perform matrix and vector operations on the GPU. CUSP natively supports several sparse matrix formats:

- Coordinate list (COO)
- Compressed Sparse Row (CSR)
- Diagonal (DIA)
- ELLPACK (ELL)
- Hybrid (HYB)

It is worth noting that in PETSc GPU by using the routines `VecSetType()` and `MatSetType()` we can select the desired sparse matrix format simply with the command line option `-mat_cusp_storage_format <format>`, and switch from a CPU version to a GPU version by means of the command line parameters `-vec_type cusp` and `-mat_type aijcusp`.

Our GPU parallel implementation of the implicit surface reconstruction method in Algorithm 1 focuses on steps 2 and 3, i.e. construction of the RBF interpolant for the extended dataset (2), and evaluation of such RBF interpolant on a given set of evaluation points. The most expensive operation in step 2 is the matrix-vector multiplication at each iteration of the RASM preconditioned GMRES. The choice of a suitable PETSc object for the sparse matrix A turns out to be crucial for the efficiency of the implementation of step 2 on the GPU. According to the results of specific experiments, we decided to use the Diagonal (DIA) sparse matrix format in CUSP, and the related CUSP sparse matrix operations. Besides taking advantage of very reliable and efficient CUSP routines, the resulting code exhibits a very low overhead for data transfer from host memory (CPU) to device memory (GPU). By contrast, in implementing step 3 we represented the matrix A as a `MatShell` PETSc object and then we used the CUDA kernel reported in Algorithm 3 below. This choice is justified by the fact that the matrix-vector multiplication in the evaluation step must be performed only once, so that the time needed to compute the non-zero entries of A, using the Algorithm 2 as in step 2, would nullify all the advantages of the efficient CUSP routines. Moreover, our CUDA kernel makes use of the shared memory portion of the device memory, whose low latency improves the performance of the computation.

Algorithm 2 Pseudo-code for the construction of the interpolation matrix

1: **for each** subdomain Ω_i **do**
2: **for each** point x_i in the subdomain Ω_i **do**
3: **for each** point x_j in the truncation area of Ω_i **do**
4: Set $A_{ij} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$
5: **end for**
6: **end for**
7: **end for**

IV. Experimental results

In this section we present some results of our method for surface reconstruction. The results were computed using a system equipped with an Intel Core i7-940 CPU (2,93 GHz, 8M Cache). The middleware framework is OS Linux kernel 2.6.32-28 and PETSc developer version 3.3.

First, we investigate the impact of the parameter σ on the quality of reconstruction. As we showed in §III.B, our method is very efficient for small values of σ compared to the domain size. However, besides efficiency, even the quality of the result also depends on the values of σ. This is because the accuracy

Algorithm 3 CUDA code for the evaluation step

```
__shared__ float sharedXi[BLOCK_SIZE];
__shared__ float sharedGi[BLOCK_SIZE];
int bx = blockIdx.x;
int tx = threadIdx.x;
int i = blockIdx.x * BLOCK_SIZE +
threadIdx.x;
float pf = 0;
float coeff = 0.5f/(sigma*sigma);
for (unsigned int m = 0; m <
(col-1)/BLOCK_SIZE+1; m++) {
sharedXi[tx] = Xi[m*BLOCK_SIZE + tx];
__syncthreads();
for (unsigned int k = 0; k <
BLOCK_SIZE; k++) {
dx = Xj[i]-sharedXi[k];
pf += sharedGi[k]*exp(-(dx*dx)*coeff;}}
Pf[i] = pf/M_PI*coef;
```

of the interpolation model depends on the ratio between the density of the point cloud and σ.

The experiments carried out in [22] were designed only for equally-spaced lattice point distributions, and the point density was measured by the spacing h between the points. In that case, a good choice of σ, in terms of performance and accuracy, is that yielding $h/\sigma \approx 1$.

For non-uniform unorganized data, more appropriate density measures can be devised. One is the so-called *separation distance* defined as

$$q_{\mathcal{X}} = \frac{1}{2} \min_{i \neq j} \|x_i - x_j\|_2. \tag{8}$$

As shown in Fig. 3, $q_{\mathcal{X}}$ geometrically represents the radius of the largest (hyper)sphere that can be drawn around each point in such a way that no (hyper)sphere intersects the others; that is why it is sometimes called *packing radius*. Another measure, popular in approximation theory, is the so-called *fill distance*:

$$h_{\mathcal{X},\Omega} = \sup_{x \in \Omega} \min_{x_j \in \mathcal{X}} \|x - x_j\|_2. \tag{9}$$

It indicates how well the set $\mathcal{X}$ fills the domain Ω. A geometric interpretation of the fill distance is given by the radius of the biggest empty (hyper)sphere that can be placed among the data sites in Ω (see Fig. 3); for this reason, it is also referred to as *covering radius*. Hence for scattered data, using (8) and (9), the heuristic "optimal" ratio $h/\sigma \approx 1$ can be expressed as follows:

$$\sigma \approx 2q_{\mathcal{X}} \tag{10}$$

and

$$\sigma \approx h_{\mathcal{X},\Omega}\sqrt{2}. \tag{11}$$

A. Tests on synthetic dataset

The experiments on synthetic dataset deal with a sphere, that allows to easily calculate (8) and (9). In order to perform

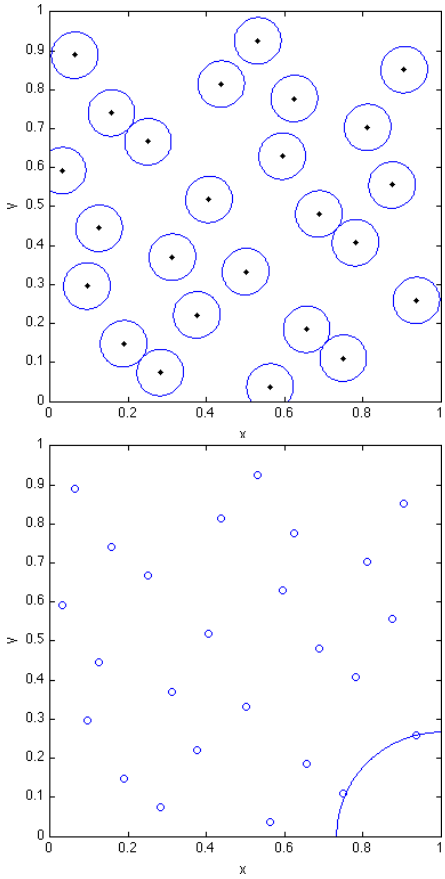


Fig. 3. Geometric interpretation of separation distance (on the left) and fill distance (on the right) for 25 Halton points on the domain $\Omega = [0,1]^2$ ($q_{\mathcal{X}} \approx 0.0597$ and $h_{\mathcal{X},\Omega} \approx 0.2667$).

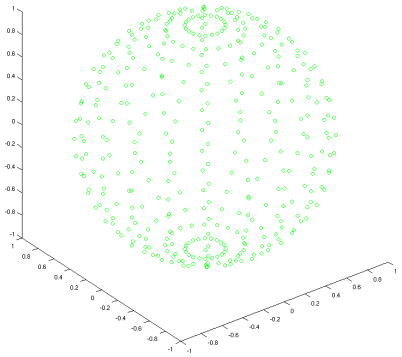


Fig. 4. 382 point cloud from the unit sphere centered in the origin.

a test consistent with a real dataset, the widely scattered point cloud in Fig. 4 was selected.

As shown in Fig. 5 (top left), a too small value of σ leads to a surface that actually interpolates the given point cloud but whose reconstruction quality is unsatisfactory. Notice that, even using (10) as in [22], one couldn't achieve a better result (see Fig. 5, top right). On the contrary, by using (11) and increasing the value of σ (see Fig. 5, bottom left) the quality of the reconstruction improves up to the desired level (see Fig.

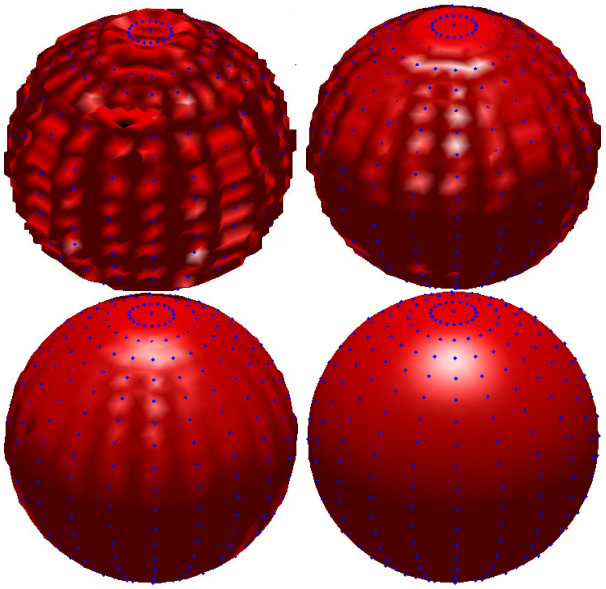


Fig. 5. Reconstructed sphere for different values of σ. (from left to right $\sigma = 0.025$, $\sigma = 2q_{\mathcal{X}} = 0.048$, $\sigma = 0.065$, $\sigma = h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 0.157$)

5, bottom right). It is interesting to note that for intermediate values of σ, though the quality is not globally satisfactory, it appears to be locally adequate, mainly in those parts of the surface where the points are at a distance $d \approx \sigma$.

1) Tests on incomplete data:

We assessed the sensitivity of our method to the lack of information by means of an incomplete dataset. We began with a dataset composed of 50% randomly chosen points from the previous dataset; Fig. 6 shows the new the point cloud. In this case, it turns out that an "optimal" value of σ is

$$\sigma = h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 0.328.$$

This optimal value corresponds to the actual fill distance for the new dataset and provides a successful reconstruction of the sphere (see Fig. 6, bottom). We remark that the old value $\sigma = 0.157$ would lead to a reconstructed surface which is not the desired sphere (see Fig. 6, top).

Another interesting test deals with a point cloud belonging to the (upper) semi-sphere:

$$\mathcal{M}^+ = \{(x,y,z) \in \mathbb{R}^3 : \ x^2 + y^2 + z^2 = 1, z \geq 0\}.$$

If we set $\sigma = h_{\mathcal{X},\mathcal{M}^+}\sqrt{2} = 0.157$, then we can reconstruct the $\mathcal{M}^+$ surface (see Fig.7, top) from which the dataset was extracted. If we assume that the point cloud came from the whole sphere $\mathcal{M}$, then the value of the fill distance would change to $h_{\mathcal{X},\mathcal{M}} = \sqrt{2}$. The latter value of the fill distance gives an optimal value of $\sigma = h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 2$, which leads to the reconstruction of the whole sphere, as shown in Fig. 7, bottom.

B. Tests on real dataset

We briefly discuss the results of some tests concerning real dataset, namely the Stanford Bunny model. This is composed

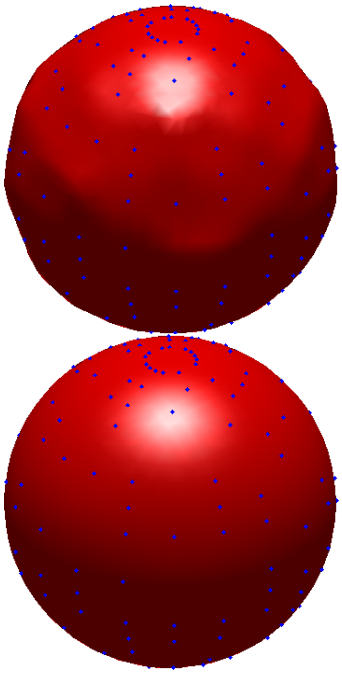


Fig. 6. Reconstructed sphere from incomplete data. (on the left $\sigma = 0.157$ and on the right $\sigma = h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 0.328$)

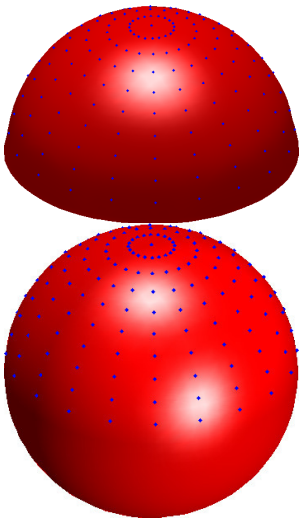


Fig. 7. Reconstructed surface from upper semi-sphere data. (on the left $\sigma = h_{\mathcal{X},\mathcal{M}^+}\sqrt{2} = 0.157$ and on the right $\sigma = h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 2$)

of $N = 8171$ points, giving an extended dataset of $N_{ext} = 3N = 24513$ points. In order to select a suitable value of σ, we need to calculate the value of the fill distance for the given dataset. On real datasets, where the geometry of the surface is either unknown or very complex, this task can be a real challenge. Recalling that the fill distance measures the data density in the membership domain, we introduce a new measure defined as

$$h_{max} = \max_j \min_{i \neq j} \|x_i - x_j\|_2.$$

This represents the largest value of the distances between each point and its nearest neighbor point. For a dataset without multiple *connected components*, the fill distance can be approximated by

$$h_{\mathcal{X},\mathcal{M}} \approx h_{max}\sqrt{2}. \tag{12}$$

As in the case of synthetic dataset, a too small value of σ results in a low quality reconstructed surface (see Fig. 8, top left)), intermediate values lead to reconstructions which are adequate only where there is a high density of points (see Fig.8, top right), while the choice (12) of $\sigma \approx h_{\mathcal{X},\mathcal{M}}$ provides the best result (see Fig. 8, bottom).

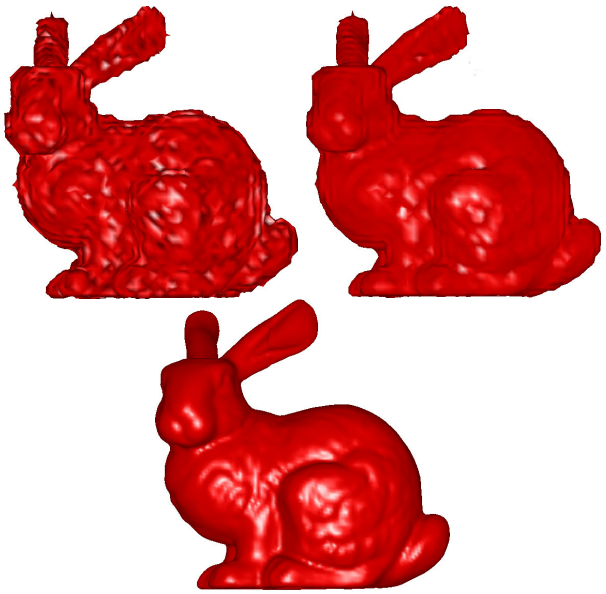


Fig. 8. Reconstructed bunny for different values of σ. (from left to right $\sigma = 0.0007$, $\sigma = 0.0012$ and $\sigma = h_{max} \approx h_{\mathcal{X},\mathcal{M}}\sqrt{2} = 0.0033$)

C. Tests on performance

We report some results on the performance of the GPU implementation of our method compared to its CPU implementation. CPU times refer to the execution on one core of the i7-940 CPU, and GPU times refer to execution on the Nvidia Fermi C1060 GPU, with 4Gb of RAM. The middleware software consists of PETSc developer version 3.3, compiled with GPU support, CUDA release 4.2, CUSP version 0.3.0 and THRUST version 1.5.2.

We have used synthetic point clouds of increasing size number N, i.e with an increasing the density. Since we want to emphasize the advantages in exploiting the GPU, we choose a constant value of σ ($\sigma = 0.157$). Notice that if the value of σ scaled with the density, then the problem would scale as $\mathcal{O}(N)$, giving rise to a less noticeable GPU acceleration.

In Tab. II, we report the execution times on a single CPU and a single GPU, and also the resulting speed-ups for the interpolant construction (step 2 of the reconstruction algorithm), varying the size N of the dataset. To make a fair comparison, we fixed the number of iterations in GMRES to 50. Results

in Tab. II show that, even though the RASM preconditioner is not available in CUSP, the GPU implementation achieves a 6.6 sped-up.

TABLE II
SPEED-UPS FOR THE INTERPOLANT CONSTRUCTION STEP ON GPU.

N	CPU	GPU	Speed up
1323	0,67551	0,1404	4,81
4686	34,567	5,6472	6,12
15625	451,75	71,57	6,31
24036	5398,4	815,09	6,63

In Tab. III we reported the execution times on CPU and GPU, and the resulting speed-ups for the interpolant evaluation (step 3), varying the size N of the point cloud and the size M of the evaluation grid. As expected, these execution times are substantially lower than those obtained for the step 2, but in this case the GPU is fully exploited and larger speed-ups are achieved.

TABLE III
SPEED-UPS FOR THE EVALUATION STEP ON GPU.

$M\backslash N$		1323	4686	15625	24036
15625	CPU	0,11571	0,20605	0,78172	0,97539
	GPU	0,012831	0,033589	0,12167	0,13699
	Speed up	**9,01**	**6,13**	**6,42**	**7,12**
125000	CPU	0,60406	1,5657	5,9558	7,433
	GPU	0,029907	0,09974	0,32034	0,43697
	Speed up	**20,19**	**15,69**	**18,59**	**17,01**
421875	CPU	1,9098	5,2612	19,946	25,037
	GPU	0,084603	0,2892467	0,89695	1,1272
	Speed up	**22,57**	**18,18**	**22,23**	**22,211**
1000000	CPU	4,4389	12,505	47,594	59,501
	GPU	0,18456	0,60741	2,0629	2,4968
	Speed up	**24,05**	**20,58**	**23,07**	**23,83**
1953125	CPU	8,8515	24,431	92,589	116,39
	GPU	0,35668	1,0662	3,9195	4,8035
	Speed up	**24,81**	**22,91**	**23,62**	**24,23**
3375000	CPU	15,018	42,166	160,17	199,45
	GPU	0,59846	1,7525	6,4671	7,9468
	Speed up	**25,09**	**24,06**	**24,76**	**25,09**

V. CONCLUSION

We proposed a parallel method based on radial basis functions for surface reconstruction on GPU. Our implementation relies on GPU-parallel scientific libraries, in order to take full advantage of the computing power of the GPU device. We showed that reconstruction quality and performance of the method are strongly related to the gaussian RBF parameter σ. We proposed an optimal heuristic estimate of such parameter based on suitable density measures of the point cloud. Finally, the observed speed-ups and running times confirm that the RBF interpolation can be a very effective approach to solve large scale surface reconstruction problems.

REFERENCES

[1] Satish Balay, William Gropp, Lois Curfman McInnes, and Barry F Smith. Petsc, the Portable, Extensible Toolkit for scientific computation. *Argonne National Laboratory*, 2:17, 1998.

[2] Alex Beutel, Thomas Mlhave, Pankaj K. Agarwal, Natural neighbor interpolation based grid DEM construction using a GPU. Proc. of the 18th SIGSPATIAL International Conf. on Advances in Geographic Information Systems. pp 172-181. ACM New York, 2010

[3] Jonathan C Carr, Richard K Beatson, Jon B Cherrie, Tim J Mitchell, W Richard Fright, Bruce C McCallum, and Tim R Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proc. of the 28th annual Conf on Computer graphics and interactive techniques*, pages 67–76. ACM, 2001.

[4] Jonathan C. Carr, W. Richard Fright, and Richard K Beatson. Surface interpolation with radial basis functions for medical imaging. *Medical Imaging, IEEE Transactions on*, 16(1):96–107, 1997.

[5] P.C. Curtis. n-parameter families and best approximation. *Pacific Journal of Mathematics*, 9(4):1013–1027, 1959.

[6] S. Cuomo, P. De Michele, R. Farina, F. Piccialli, A Smart GPU Implementation of an Elliptic Kernel for an Ocean Global Circulation Model, *Applied Mathematical Sciences*, Vol. 7, no. 61, 3007 - 3021 (2013).

[7] S. Cuomo, P. De Michele, R. Farina, A CUBLAS-CUDA implementation of PCG method of an ocean circulation model, *AIP Conf. Proc.*,1389, pp. 1923-1926; doi:http://dx.doi.org/10.1063/1.3636988 (2011).

[8] R. Franke. Scattered data interpolation: Tests of some methods. *Math. Comput.*, 38(157):181–200, 1982.

[9] Sarah F Frisken, Ronald N Perry, Alyn P Rockwood, and Thouis R Jones. Adaptively sampled distance fields: a general representation of shape for computer graphics. In *Proc. of the 27th annual Conf on Computer graphics and interactive techniques*, pages 249–254. ACM Press/Addison-Wesley Publishing Co., 2000.

[10] F. Farina, S. Cuomo, P. De Michele, F.Piccialli, An inverse preconditioner for a free surface ocean circulation model *AIP Conf. Proc.*, 1493, pp. 356-362; doi:http://dx.doi.org/10.1063/1.4765513 (2012).

[11] Jan Klein and Gabriel Zachmann. Point cloud surfaces using geometric proximity graphs. *Computers & Graphics*, 28(6):839–850, 2004.

[12] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proc. of the 27th annual Conf. on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 131–144. ACM Press/Addison-Wesley Publishing Co., 2000.

[13] J.C. Mairhuber. On haar's theorem concerning chebychev approximation problems having unique solutions. *Proc. of the American Mathematical Society*, pages 609–615, 1956.

[14] Nathan Bell and Michael Garland. Cusp: Generic parallel algorithms for sparse matrix and graph computations, 2012. Version 0.3.0.

[15] F. Piccialli, S. Cuomo, P. De Michele, A Regularized MRI Image Reconstruction based on Hessian Penalty Term on CPU/GPU Systems, *Procedia Computer Science*, 18, 2643-2646 (2013).

[16] Joachim Pouderoux, Jean-Christophe Gonzato, Ireneusz Tobor, and Pascal Guitton. Adaptive hierarchical rbf interpolation for creating smooth digital elevation models. In *Proc. of the 12th annual ACM international workshop on Geographic information systems*, GIS '04, pages 232–240. ACM, 2004.

[17] Oliver Schall and Marie Samozino. Surface from scattered points. In *a Brief Survey of Recent Developments. 1st International Workshop on Semantic Virtual Environments, Page (s)*, pages 138–147, 2005.

[18] J Süßmuth, Q Meyer, and G Greiner. Surface reconstruction based on hierarchical floating radial basis functions. In *Computer Graphics Forum*, volume 29, pages 1854–1864. Wiley Online Library, 2010.

[19] Jared Hoberock and Nathan Bell. Thrust: A parallel template library, 2010. Version 1.5.2.

[20] Greg Turk, Huong Quynh Dinh, James F O'Brien, and Gary Yngve. Implicit surfaces that interpolate. In *Shape Modeling and Applications, SMI 2001 International Conf on.*, pages 62–71. IEEE, 2001.

[21] Greg Turk and James F O'brien. Variational implicit surfaces. 1999.

[22] Rio Yokota, L.A. Barba, and Matthew G. Knepley. Petrbf a parallel o(n) algorithm for radial basis function interpolation with gaussians. *Computer Methods in Applied Mechanics and Engineering*, 199(25):1793 – 1804, 2010.

[23] H. Wendland. *Scattered data approximation*, volume 2. Cambridge University Press Cambridge, 2005.

[24] Hong-Kai Zhao, Stanley Osher, and Ronald Fedkiw. Fast surface reconstruction using the level set method. In *Variational and Level Set Methods in Computer Vision, 2001. Proc. IEEE Workshop on*, pages 194–201. IEEE, 2001.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 441–446

Towards an Efficient Multi-Stage Riemann Solver for Nuclear Physics Simulations

Sebastian Cygert,
Joanna Porter-Sobieraj
Warsaw University of Technology
Faculty of Mathematics and Information Science
Koszykowa 75, 00-662 Warsaw,
Poland
Email: j.porter@mini.pw.edu.pl

Daniel Kikoła
Purdue University
Department of Physics
525 Northwestern Ave.,
West Lafayette, IN 47907,
United States
Email: dkikola@purdue.edu

Jan Sikorski,
Marcin Słodkowski
Warsaw University of Technology
Faculty of Physics
Koszykowa 75, 00-662 Warsaw,
Poland
Email: slodkow@if.pw.edu.pl

***Abstract*—Relativistic numerical hydrodynamics is an important tool in high energy nuclear science. However, such simulations are extremely demanding in terms of computing power. This paper focuses on improving the speed of solving the Riemann problem with the MUSTA-FORCE algorithm by employing the CUDA parallel programming model. We also propose a new approach to 3D finite difference algorithms, which employ a GPU that uses surface memory. Numerical experiments show an unprecedented increase in the computing power compared to a CPU.**

I. Motivation

NUMERICAL hydrodynamics has been used in many scientific and engineering applications for decades, mostly due to its relative simplicity. Even a complicated, dynamic system can be described with a small set of relatively simple hyperbolic conservation laws in this framework. All the information about the physical properties of a system is contained in a single equation of state, which is the relationship between the thermodynamic properties of the analyzed system. Knowledge of the details of the interactions that take place on the microscopic level is not required. However, there is one strong assumption underlying the use of hydrodynamics: the system has to be at least in the local thermodynamic equilibrium, which means that the thermodynamic quantities for any point are approximately constant around that point.

Recently relativistic hydrodynamics has been applied in studies of a new field of physics: high energy nuclear science. The goal of high energy nuclear science is to study the interactions between the basic constituents of matter; quarks and gluons. In normal conditions, quarks and gluons are bound together to form nucleons: protons and neutrons. However, the forces binding quarks together can be subjected to a sufficiently high energy density, leading to a transition from ordinary nuclear matter to a new state, where quarks and gluons behave like quasi-free particles. Such a *soup* of quarks and gluons is called a Quark-Gluon Plasma (QGP), and it is hypothesized to have existed in the early universe, a few millionths of a second after the Big Bang. The energy density of nuclear matter created in relativistic heavy ion collisions at the Relativistic Heavy Ion Collider (RHIC) at Brookheaven National Laboratory, or the Large Hadron Collider (LHC) at CERN, is sufficiently high that a phase transition to a QGP is expected in such reactions.

In the collisions of heavy nuclei at RHIC or LHC, a decrease in the amount of nuclear matter occurs with exposure to the extreme energy density. The first phase of collisions consists of interactions with a large momentum transfer. Then the system expands, equilibrates and forms the QGP. Relativistic hydrodynamics can then be employed to extract the properties of the QGP. Relativistic ideal fluid dynamics has indeed been used in the theoretical modeling of the QGP and remarkable agreement between experimental data and simulations has been found, leading to the unexpected conclusion that *hot* nuclear matter behaves like a nearly frictionless liquid (i.e. with extremely low viscosity) [1], [2].

The initial success of relativistic hydrodynamics stimulated further development of the numerical codes needed for a more precise understanding of the fundamental properties of the quark and gluon dynamics in the QGP.

Firstly, fully (3+1)-dimensional simulations (3 spatial dimensions + time) are necessary to describe the system's evolution without any assumptions regarding its symmetries. Such numerical problems are well defined and numerical methods are available; however, (3+1)-dimensional simulations are extremely expensive in terms of computing power. Moreover, the fluctuations in the pre-QGP phase might have an impact on the dynamics of the QGP, therefore *event-by-event* studies (where an *event* is a single collision between heavy ions) with fluctuating initial conditions and with a large amount of statistics are of particular interest and require a fast and efficient computer code. Furthermore, studies of jets (which are narrow beams of particles with a high momentum) and their propagation through the hot nuclear medium can provide information about the fundamental properties of the QGP (transport coefficients, for instance). However, such studies require an accurate representation of relativistic flows and shock waves. This requires a larger numerical grid in the simulations, which in turn increases the computing time

This work was supported in part by Dean's Grant (2012) at the Faculty of Physics Warsaw University of Technology

978-83-60810-53-8/$25.00

needed significantly.

Due to all these factors, there is a constantly increasing demand for computing resources in relativistic hydrodynamics simulations. Graphics Processing Unit (GPU) computing is a promising solution for this problem, and offers an unprecedented increase in computing power compared to standard CPU simulations. In this paper, we present the concept of an implementation of 3+1 ideal hydrodynamics simulations carried out on a GPU using an NVIDIA CUDA framework and test results for selected physics problems.

II. Equation System for the Riemann Problem

A. Hydrodynamics Equations

Equations of relativistic hydrodynamics can be written in a conservative form:

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}+\frac{\partial G(U)}{\partial y}+\frac{\partial H(U)}{\partial z}=0 \qquad (1)$$

where $U=(E,M_x,M_y,M_z,R)$ is a vector of conserved quantities in the *laboratory rest frame*, E is the energy density, M_i is the momentum density in the i-th Cartesian coordinate and R is a conserved charge density (e.g. a baryon number). F,G,H are vectors of fluxes of those quantities in the x,y,z directions, defined as:

$$F(U)=\begin{pmatrix}(E+p)v_x\\ M_xv_x+p\\ M_yv_x\\ M_zv_x\\ Rv_x\end{pmatrix}$$

$$G(U)=\begin{pmatrix}(E+p)v_y\\ M_xv_y\\ M_yv_y+p\\ M_zv_y\\ Rv_y\end{pmatrix} \qquad (2)$$

$$H(U)=\begin{pmatrix}(E+p)v_z\\ M_xv_z\\ M_yv_z\\ M_zv_z+p\\ Rv_z\end{pmatrix}$$

where v is the velocity and p is pressure, defined by an equation of state: $p=p(e,n)$; e and n are the energy and charge density in the *fluid rest frame* (i.e. in a frame where velocity vanishes, $v=(0,0,0)$).

B. Numerical Scheme

In numerical applications, all continuous fields have to be represented in a finite number of degrees of freedom, e.g. on a fixed numerical grid. In our program we use a finite-difference scheme on a Cartesian grid. Since non-conservative methods (i.e. methods based on non-conservative variables) have been shown to fail (do not converge to a correct solution) if a shock wave is present in the solution [3], a conservative method is used.

There are two standard approaches to solving the problem (1): *dimensional splitting* and a *finite volume method*. Derivation of dimensional splitting methods is based on Taylor series expansions and may give incorrect results for discontinuous solutions [4], thus a finite volume method was chosen. For a three-dimensional problem, such a scheme reads:

$$\begin{aligned}U^{n+1}_{i,j,k} &= U^n_{i,j,k}+\frac{\Delta t}{\Delta x}\left(F_{i-\frac12,j,k}-F_{i+\frac12,j,k}\right)\\ &+\frac{\Delta t}{\Delta y}\left(G_{i,j-\frac12,k}-G_{i,j+\frac12,k}\right)\\ &+\frac{\Delta t}{\Delta z}\left(H_{i,j,k-\frac12}-H_{i,j,k+\frac12}\right)\end{aligned} \qquad (3)$$

where $U^n_{i,j,k}$ represents a conserved quantity at the discrete time t_n; Δt and Δx, Δy, Δz are time and space steps, respectively, and $F_{i-\frac12,j,k}$, ... , $H_{i,j,k+\frac12}$ are numerical fluxes through cell boundaries.

The central point of a particular scheme is the construction of intercell fluxes $F_{i-\frac12,j,k}$, ... , $H_{i,j,k+\frac12}$. There are two distinct approaches to this problem: the *upwind* and *centered* schemes.

The main feature of upwind schemes is that they explicitly exploit information about wave propagation contained in the equations, usually by solving a one-dimensional Riemann problem locally. The accuracy of such schemes is highly dependent on the choice of a particular Riemann solver, which should ideally be *complete* (i.e. take into account all characteristic fields present in the exact solution).

On the other hand, centered methods do not solve the Riemann problem directly, and therefore are usually simpler and more general, at the cost of accuracy (given that there is a complete Riemann solver available).

In order to obtain a general and accurate algorithm, we use a hybrid MUSTA (MUlti-STAge) approach [5], [6]. This utilizes a centered flux in a predictor-corrector loop, solving the Riemann problem numerically, i.e. without using a priori information about waves.

The algorithm in a one-dimensional case is as follows:

1) In order to calculate flux $F_{i+\frac12}$ we introduce auxiliary variables $U^{(l)}_L$ and $U^{(l)}_R$ and their fluxes $F^{(l)}_L$ and $F^{(l)}_R$.
2) Set $U^0_L=U_i$, $U^0_R=U_{i+1}$.
3) Calculate $F^{(l)}_L=F(U^{(l)}_L)$ and $F^{(l)}_R=F(U^{(l)}_R)$ using (1).
4) Calculate $F^{(l)}_{i+\frac12}$ using a centered flux, $U^{(l)}_L$, $U^{(l)}_R$, $F^{(l)}_L$ and $F^{(l)}_R$. If l reached a predefined value, stop.
5) Solve Riemann problem locally:

$$\begin{aligned}U^{(l+1)}_L &= U^{(l)}_L-\tfrac{\Delta t}{\Delta x}\left(F^{(l)}_{i+\frac12}-F^{(l)}_L\right)\\ U^{(l+1)}_R &= U^{(l)}_R-\tfrac{\Delta t}{\Delta x}\left(F^{(l)}_R-F^{(l)}_{i+\frac12}\right)\end{aligned} \qquad (4)$$

6) Go back to step 3.

One drawback to using such an algorithm is that it is numerically more expensive than other, more conventional, algorithms, for instance the SHASTA (SHarp And Smooth Transport Algorithm) [7], [8].

As a centered flux, we used the FORCE (First ORder CEntered) scheme:

$$F^{\text{force}}_{i+\frac{1}{2}} = \frac{1}{2}\left(F^{\text{lw}}_{i+\frac{1}{2}} + F^{\text{lf}}_{i+\frac{1}{2}}\right) \tag{5}$$

where $F^{\text{lw}}_{i+\frac{1}{2}}$ is the Lax-Wendroff type flux (in terms of MUSTA auxilliary variables):

$$F^{\text{lw}}_{i+\frac{1}{2}} = F\left(\frac{1}{2}(U_L + U_R) - \frac{1}{2}\frac{\alpha\Delta t}{\Delta x}(U_R - U_L)\right) \tag{6}$$

and $F^{\text{lf}}_{i+\frac{1}{2}}$ is the Lax-Friedrichs type flux:

$$F^{\text{lf}}_{i+\frac{1}{2}} = \frac{1}{2}(F_L + F_R) - \frac{1}{2}\frac{\Delta x}{\alpha\Delta t}(U_R - U_L) \tag{7}$$

In a three-dimensional case $\alpha = 3$, but other values may also be considered.

To achieve second order accuracy in space and time, we extend our algorithm with MUSCL-Hancock scheme. The basic idea of this scheme is to use more cells to interpolate inter-cell values and evolve them half a time step. The algorithm is:

1) Replace cell average values U^n_i by a piece-wise linear function inside i-th cell:

$$U_i(x) = U^n_i + \frac{(x - x_i)}{\Delta x}\Delta_i \tag{8}$$

where Δ_i is a slope vector and will be defined later. In the local coordinates the points $x = 0$ and $x = \Delta x$ correspond to boundaries of the cell $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$. The values at these points are $U^L_i = U^n_i - \Delta_i/2$ and $U^R_i = U^n_i + \Delta_i/2$.

2) Propagate U^L_i and U^R_i by a time $\frac{1}{2}\Delta t$:

$$\begin{aligned}
\tilde{U}^L_i &= U^L_i + \frac{1}{2}\frac{\Delta t}{\Delta x}(F(U^L_i) - F(U^R_i)) \\
&\quad + \frac{1}{2}\frac{\Delta t}{\Delta y}(G(U^L_i) - G(U^R_i)) \\
&\quad + \frac{1}{2}\frac{\Delta t}{\Delta z}(H(U^L_i) - H(U^R_i)) \\
\tilde{U}^R_i &= U^R_i + \frac{1}{2}\frac{\Delta t}{\Delta x}(F(U^L_i) - F(U^R_i)) \\
&\quad + \frac{1}{2}\frac{\Delta t}{\Delta y}(G(U^L_i) - G(U^R_i)) \\
&\quad + \frac{1}{2}\frac{\Delta t}{\Delta z}(H(U^L_i) - H(U^R_i))
\end{aligned} \tag{9}$$

3) Use $\tilde{U}^L_i$ and $\tilde{U}^R_i$ as U^0_L and U^0_R in MUSTA.

A simple choice for the *slope* Δ_i in (8) is:

$$\Delta_i = \frac{1}{2}(U^n_{i+1} - U^n_{i-1}) \tag{10}$$

which indeed results in a second-order accurate algorithm. However, as predicted by Godunov's theorem, it has the unpleasant effect of producing spurious oscillations in the vicinity of strong gradients.

To solve this issue, a number of flux limiting and slope limiting methods have been proposed. We employed a slope limiting method; instead of Δ_i as in (10) we use:

$$\tilde{\Delta}_i = \xi(r_i)\Delta_i \tag{11}$$

in (8), where ξ is called the *slope limiter*, and r_i is defined as:

$$r_i = \frac{U_i - U_{i-1}}{U_{i+1} - U_i} \tag{12}$$

There are a number of possible choices for ξ, each with its own characteristics and features. One possibility is the MINBEE limiter:

$$\xi_{\text{mb}}(r) = \max(0, \min(1, r)) \tag{13}$$

and there is another, called SUPERBEE:

$$\xi_{\text{sb}}(r) = \max(0, \min(2r, 1), \min(r, 2)) \tag{14}$$

Introducing non-linearity in the scheme provides less oscillations near high gradients and retains good accuracy in smooth areas of the solution.

III. Implementation Notes

A. GPUs – an Overview

Recent developments in GPU technology have transformed them into very powerful devices offering a notable speed increase compared to traditional CPUs in high performance computing. The reason behind this lies in the difference in structure between these two processors. GPUs use many more resources for arithmetic operations at the expense of cache and flow control.

The basic idea of a GPU is that it is built around an array of Streaming Multiprocessors (SMs). Compute Unified Device Architecture (CUDA) allows a programmer to define a special function *kernel* which is executed on a GPU device by a number of threads which are organized into blocks. Each thread block executes in parallel on a single SM independently and in undefined order.

CUDA threads can access several memory spaces. *Global memory*, which has a very big latency, can be accessed by all the threads. Threads within one block can cooperate through *shared memory*. Shared memory is expected to be much faster than global memory and many applications benefit from using it. Registers, whose limited number is distributed to threads by a streaming multiprocessor, offer the lowest latency. By default, all the variables are placed in registers for as long as the latter are available. When there is a lack of registers, other variables go in the local memory which resides in device memory and provides the same high memory latency. This is called *register spilling* and causes a notable slow down for applications.

The main drawback of shared memory is its limited size (48 KB in contemporary GPUs). This is sufficient for many applications, but for others - such as image processing where millions of pixels are transformed in parallel - it is not. For such cases data may be put into texture memory which resides in device memory and is cached in the texture cache. As a

result, data read from texture memory cost one read from global memory on cache miss, and give almost immediate access otherwise. The texture cache is optimized for 2D spatial locality. Texture memory offers only data reading while surface memory offers both read and write operations.

Threads are executed in groups of 32 parallel threads called warps, of which all execute the same instruction scheduled by the warp scheduler. If, due to conditional sentences, different threads within a warp follow different paths, then the scheduler visits each path taken sequentially, disabling threads that are not active in the current path. This is called branching and effects slow down in execution.

A common bottleneck for many GPU applications is memory access latency due to the limited size of cache, especially when using global device memory. However, those latencies can be hidden by a warp scheduler. At every instruction issue time, a warp scheduler selects a warp that is ready to execute its next instruction, if any, and issues the instruction to the active threads of the warp. It can easily be seen that CUDA performs this most effectively when there are plenty of threads to be executed.

B. Data Organization

GPUs have recently been widely used for many advanced computations. Typical examples have involved 3D finite difference computations using the most popular sliding-window approach, which operates using shared memory [9] - [12]. Some papers have also studied using texture memory to optimize the solution in fluid dynamics [13], [14].

In our approach surface memory is applied to hold the simulation data. Surface memory retains all the benefits of texture memory, but it also works in write mode. It is available for NVIDIA GPUs with a compute capability of 2.x or higher. Currently, surface memory does not support double precision floating-point arithmetic and all the calculations are done solely in single precision.

Surface memory offers several benefits over the popular sliding-window approach. Due to the limited size of shared memory, only a limited amount of the data can be held in it and hence loop tiling has to be performed. An input grid is divided into smaller blocks (*tiles*) that fit into shared memory, the threads copy these parts of the data from global to shared memory and perform computations in the latter. Thus, the shared memory can be seen as a manually managed cache. It can easily be noticed that this results in data access redundancy which can be computed using the formula $(n * m + k(n + m))/(n*m)$, where n and m stand for a block's size and k is the order of stencil [10]. In contrast to shared memory, which is highly limited in size, the maximum number of elements we can bound to surface memory is 65536x32768x2048 [15]. Therefore, in practice surface memory is limited only by its size and in most cases all the data can be kept in surface memory.

Surface memory allows the algorithm to decrease register usage. In our approach, the numerical scheme requires a single cell to contain more than one simple variable. When we multiply the size of a cell by the number of cells needed in stencil computation it is easy to notice that the registers are quite heavily used in this algorithm.

Using surface memory it is also easier to modify and test different numerical schemes. Code which uses shared memory is usually dedicated to just one kind of stencil. Changing the order or *direction* of the stencil results in changing many lines of code. In the case of surface memory, since all data are held in it and global indexing is used, this change may be done at once and handling special cases is kept to a minimum. Hence, it is a more general approach.

C. GPU Algorithm

In this section we present the idea behind the algorithm which uses surface memory.

Algorithm 1 Data processing schema for a thread (i, j) and an order-4 stencil computation

```
for n in 1..N do
  for k in 3..Z-Dimension −2 do
    Load neighbor cells from surface memory.
    Compute cell U(i, j, k).
    Write result to surface memory.
    Synchronize threads.
  end for
end for
```

The presented algorithm was built based on the idea of the sliding-window algorithm. All threads work on a 2D *xy*-slice of the grid, benefiting from optimization for the 2D spatial locality of surface memory, and iterate through the *Z*-axis. The *Compute cell* method refers to the MUSTA-FORCE algorithm. In this algorithm we use two surfaces which are employed alternately for read/write operations. It is possible to use just one surface but this would put more pressure on registers and require additional synchronizations. Since memory usage is not a problem in our simulation we decided to stay with two surfaces being used.

IV. Experimental Results

We examined the performance of the parallel GPU code and compared it to the performance of the sequential CPU code. The goal of our research was just to verify the usefulness of GPUs for solving equations of relativistic hydrodynamics and to approximate the order of magnitude of the possible speed-up. Therefore, we measured the time needed to perform the simulations on a single GPU and a single CPU core. This allowed us to estimate the time required to perform massive physics simulations.

The numerical experiments were executed on an Intel Pentium B960, 2.2 GHz processor with an NVIDIA GeForce 610 1 GB graphics card with Compute Capability 2.1. The figures show the time taken for a hydrodynamic simulation, using the MUSTA-FORCE algorithm approach for various configurations of input data.

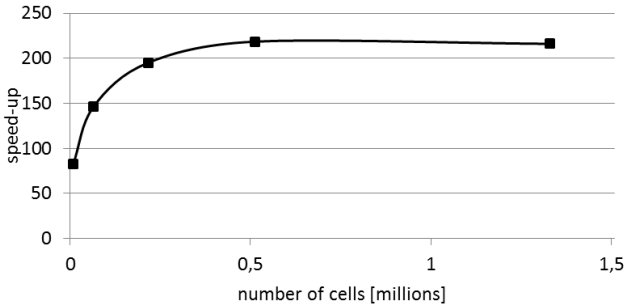


Fig. 1. The speed-up gained by using surface memory compared to CPU implementation for the MUSTA-FORCE algorithm as a function of the total number of cells

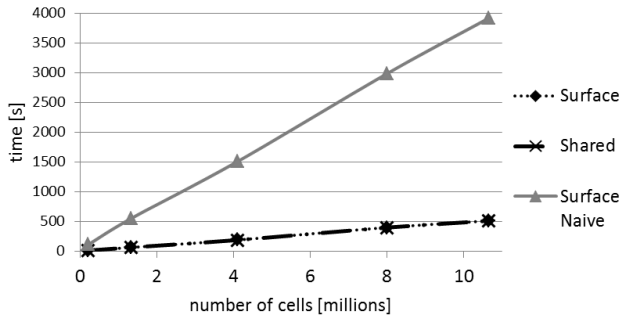


Fig. 2. Execution time for shared memory and surface memory approaches for 100 steps of the MUSTA-FORCE algorithm as a function of the total number of cells

With the GPU, the tests were carried out for 100 time steps of the MUSTA-FORCE algorithm, using grids of dimensions 60^3, 110^3, 160^3, 200^3, and 220^3. A single cell, containing a vector U, occupies 20 bytes of memory. The maximum grid that fitted within the memory limitations was 240^3.

The simulations on the CPU were conducted on smaller grids. We interrupted the calculations for grids when the time exceeded a few hours, because it would have taken days to perform massive tests on the CPU.

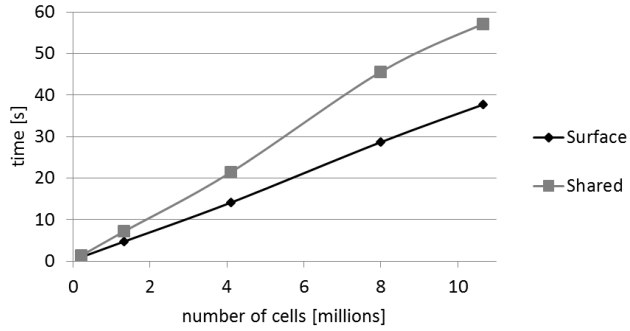


Fig. 3. Execution time for 100 steps of a generalized 3D finite difference algorithm

Fig. 1 shows the acceleration factor gained by using a GPU instead of a CPU. The GPU implementation speed-up is over 200 for bigger numbers of cells. For a grid of size 110^3, the GPU simulation took 1 minute while the CPU required over 3 hours. This proves that finite difference computations are a perfect example of an algorithm that fits the parallel computation concept. The whole algorithm can easily be divided into small parts that single threads can perform in parallel.

Note that the simulation on the CPU is very slow. The estimated effort for 100 time steps and a grid with 220 cells in each dimension is almost 30 hours. Such large grids are necessary in the case of studies of ultra-relativistic flows and strong shocks. Moreover, event-by-event simulations require samples of thousands of such simulations (events). Because of this, the total computing time needed for such analysis (assuming 1,000 events) amounts to as much as 3 years, which makes such a study extremely hard, if not impossible, without parallel computing. This example illustrates how expensive, in terms of computation and memory usage, relativistic hydrodynamic simulations are.

The next thing we examined was various implementations on a GPU. Fig. 2 presents a comparison between the execution time for the sliding-window algorithm using shared memory, and surface memory implementations. For shared memory we used a 16x16 data tile, which we found to be the most effective size. The first, naive surface memory implementation used exactly the same approach as with the shared memory algorithm. Thus it can be concluded that surface memory is about 8 times slower. The revised version of the algorithm with surface memory, presented in section 3c, used all its benefits - keeping all data in surface memory, lower usage of registers, and smaller branching. It decreased the simulation time significantly. As a result the timing of the surface memory and sliding-window approach turned out to be almost identical.

Profiling of the application showed that here we are faced with register spilling which causes a serious slow down. This is due to the fact that the MUSTA-FORCE and MUSCL algorithms we have used, both use many temporary cells interpolated during computations. Now, when there only a maximum of 63 registers per thread, and each cell takes 5 of them, a lot of data need to be kept in local memory. Usage of local memory instead of registers is one of the biggest limitations in current GPUs. Our tests showed that just increasing the size of a single cell kept in memory without any other extra computation cost, made the time of simulation increase 5-fold.

Since one of our goals was to evaluate the effectiveness of the surface memory approach for 3D finite difference computations we prepared another version of the application, which performs an interpolation between the cells instead of the whole MUSTA-FORCE algorithm. Fig. 3 shows that the sliding-window approach is more than 50% slower than the surface algorithm. This proves our thesis that the results in Fig. 2 are affected by register pressure. Profiling of this

application shows that achieved occupancy is higher for the surface algorithm (0.45 instead of 0.32), there is lower usage of registers (in the surface algorithm all data is kept in registers while in the second approach 8 variables are also kept in local memory), and that there is a slightly smaller number of branches (12.5% instead of 14.5% on average). These numbers and the test results prove that despite the fact that surface memory is slower than shared memory, the benefits it offers allow the application to achieve better performance.

V. Conclusion

In this paper the possibilities of using GPUs for developing a solver for the Riemann problem have been examined. We studied two methods of 3D finite difference computation – a sliding-window algorithm using shared memory and our new approach based on surface memory.

First of all, the GPU proved to be a good choice for 3D finite difference computations. Such a problem scales perfectly with parallel computations and thus is very effective. Our implementation is over 200 times faster than a sequential implementation on a CPU. This number shows that graphics cards offer greater computational power for problems that can be divided into independent subproblems.

We have investigated the usefulness of our novel approach using surface memory. Because the amount of memory available is very big all the data can be kept in surface, which thus decreases data redundancy and pressure on registers. On the other hand, shared memory is in general faster than surface memory. As a result, in our application using the MUSTA-FORCE algorithm, both implementations have very comparable speeds. However, as we showed, the results were affected by register spilling caused by the high memory cost of the algorithm used. To investigate the effectiveness of surface memory in 3D finite difference methods we prepared a simplified version of an algorithm that minimized register usage. As a result, the surface memory approach turned out to be faster than the one with the sliding-window approach. It should also be stressed that surface memory implementation is more general and, in contrast to the shared memory approach, can easily be changed to use any other kind and order of isotropic, or anisotropic, stencil in any direction.

In this paper we showed that GPUs are very effective for hydrodynamics simulations in comparison to CPUs. The current GPU implementation allows a device to perform a massive number of simulations in a reasonable time. This was previously impossible, even in our preliminary parallel CPU implementation with the use of a cluster computer and MPI. The designed GPU algorithm, based on surface memory, is easy to modify and proved to be a valuable new tool for high energy nuclear science.

References

[1] J. Adams *et al.*, “Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions,” *Nucl.Phys.*, vol. A757, pp. 102–183, 2005.

[2] “RHIC Scientists Serve Up "Perfect" Liquid,” Brookhaven National Laboratory Press Release. [Online]. Available: http://www.bnl.gov/newsroom/news.php?a=1303

[3] T.Y. Hou, P.G. LeFloch, Why non-conservative schemes converge to the wrong solutions: error analysis, *Math. of Comput.*, 62: 497-530, 1994.

[4] E. F. Toro, Multi-stage predictor-corrector fluxes for hyperbolic equations. Isaac Newton Institute for Mathematical Sciences Preprint Series NI03037-NPA , University of Cambridge, UK, 2003.

[5] E.F. Toro, MUSTA: A multi-stage numerical flux, *Applied Numerical Mathematics*, 56(10-11), pp.1464-1479, 2006.

[6] E.F. Toro, V.A. Titarev, MUSTA fluxes for systems of conservation laws, *Journal of Computational Physics*, 216(2), pp.403-429, 2006.

[7] J.P. Boris, D.L. Book, Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works, *J. Comput. Phys.*, 11(1), pp.38-69, 1973.

[8] J. Gerhard, V. Lindenstruth, M. Bleicher, Relativistic hydrodynamics on graphic cards, *Computer Physics Communications*, 184(2), pp. 311-319, 2013.

[9] G. Zumbusch: Vectorized Higher Order Finite Difference Kernels, In: Proc. of the 11th international conference on Applied Parallel and Scientific Computing, pp. 343-357, 2012.

[10] P. Micikevicius: 3D fnite difference computation on GPUs using Cuda. In: Proc. 2nd Workshop on General Purpose Processing on Graphics Processing Units, 2009.

[11] T. Nagaoka, S.Watamabe, A GPU-Based Calculation Using the Three-Dimensional FDTD Method for Electromagnetic Field Analysis in 32nd Annual Internation Conference of the IEEE EMBS Buenos Aires, 2010.

[12] V. Demir, A. Z. Elsherbeni, Compute Unified Architecture (CUDA) Based Finite-Difference Time-Domain (FDTD) Implementation in Aces Journal vol. 25, 2010.

[13] E. Elsen, P. LeGresley, E. Darve, Large calculation of the flow over a hypersonic vehicle using a GPU, *J. Comput. Phys.*, 227(24), pp. 10148-10161, 2008.

[14] E. Phillips, M. Fatica, Implementing the Himeno benchmark with CUDA on GPU clusters. In IEEE International Parallel & Distributed Processing Symposium, pp. 1-10, 2010.

[15] NVIDIA Corporation: NVIDIA CUDA Programming Guide Version 5.0, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 447–454

Application of AVX (Advanced Vector Extensions) for Improved Performance of the PARFES – Finite Element Parallel Direct Solver

Sergiy Fialko
Tadeusz Kościuszko Cracow University of Technology
ul. Warszawska 24 St., 31-155 Kraków, Poland
Email: sfialko@poczta.onet.pl

***Abstract*—The paper considers application of the AVX (Advanced Vector Extensions) technique to improve the performance of the PARFES parallel finite element solver, intended for finite element analysis of large-scale problems of structural and solid mechanics using multi-core computers. The basis for this paper was the fact that the *dgemm* matrix multiplication procedure implemented in the Intel MKL (Math Kernel Library) and ACML (AMD Core Math Library) libraries, which lays down the foundations for achieving high performance of direct methods for sparse matrices, does not provide for satisfactory performance with the AMD Opteron 6276 processor, Bulldozer architecture, when used with the algorithm required for PARFES. The procedure presented herein significantly improves the performance of PARFES on computers with processors of the above architecture, while maintaining the competitiveness of PARFES with the Intel MKL *dgemm* procedure on computers with Intel processors.**

I. Introduction

THE PARFES (Parallel Finite Element Solver) is a sparse direct method for solving linear equation sets with sparse symmetric matrices, which arise when the finite element method is applied to structural and solid mechanics problems, is presented in [7], [8]. The method is developed to be used in FEA software focused on multi-core shared memory computers. PARFES supports core mode (CM) as well as two out of core modes – OOC and OOC1. In the core mode, the solver only utilizes random access memory (RAM), demonstrating good performance and speed up when the number of threads increases. If the dimension of the problem exceeds the RAM capacity, the method switches to the OOC mode, in which disk storage is used, and the amount of I/O operations is minimal. Performance and speed up deteriorate slightly compared to the CM. If the amount of RAM is not sufficient for the OOC mode, PARFES switches to OOC1. In this mode, the number of I/O operations is greatly increased; however, the RAM amount requirements are low. The performance and speed up degrade significantly, but this method allows solving problems of several million equations using desktop and laptop computers.

The option to use disk memory is the advantage of PARFES compared to PARDISO (Parallel Direct Solver), which is described in [16] and presented in the Intel MKL library [11]. Although PARDISO formally supports the OOC mode, practice showed that in this mode, this method is considerably inferior both to PARFES, and the multifrontal method where small tasks are concerned [1], [5], [10], and simply crashes when used for larger problems [7], [15].

In contrast to the multifrontal method, PARFES demonstrates significantly higher performance and speed up, and smaller RAM requirements (in OOC1 mode) [7], [8].

This paper describes further development of PARFES for the use with Intel AVX instructions [14] that implement computation vectorization elements with 256-bit registers, allowing to perform four multiplications or four additions of double type values in one CPU cycle.

It was discovered that the *dgemm* matrix multiplication procedure as implemented in Intel MKL 11.0 [12] does not provide for satisfactory performance of PARFES on a computer with a 16-core AMD Opteron 6276 CPU 2.3/3.2 GHz processor, Bulldozer architecture. For test 1: $\mathbf{C} = \mathbf{C} - \mathbf{A}\cdot\mathbf{B}$, where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are 8 000 × 8 000 square matrices, the performance of this procedure is 3 958 MFLOPS with a single thread and 35 013 MFLOPS with 16 threads. The performance of the same procedure as implemented in ACML 15.2.0 (AMD Core Math Library) [2] is 14 203 MFLOPS and 94 852 MFLOPS respectively.

However, when solving test 2 (Fig. 1, 2) it was found that the performance of this algorithm degrades (see Table 1), and the threads run in the OS kernel mode for a considerable amount of time.

```
#pragma omp parallel for
for(ib=0; ib<Nb; ++ib)
{
    ip = omp_get_thread_numb();
    Cib = Cib − Aib·B;
}
```

Fig.1 Algorithm for test 2

Matrices $\mathbf{C}$ and $\mathbf{A}$ have a block structure (Fig. 2), *ip* is the thread number, and *ib* is the block number. Inside the loop, the single-treaded version of the *dgemm* procedure (ACML [2]) is used. The arrows indicate the packing of data in the respective matrices.

This work was supported by Narodowy Centrum Nauki on the basis of decision DEC-2011/01/B/ST6/00674.

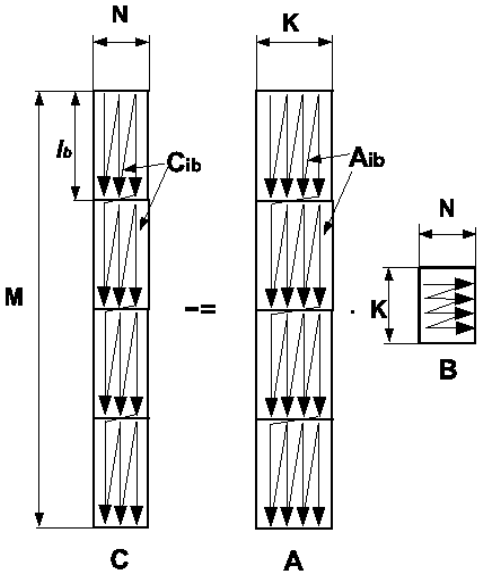


Fig. 2 Structure of **A**, **B**, **C** matrices in test 2

Test 2 is a good simulation of the PARFES correction procedure [7], [8], when the *jb* block-column (matrix **C**) is updated by the *kb* block-column (matrix **A**) located to the left of the former.

Thus, it was decided to develop a new procedure, *microkern_8x4_AVX*, which would allow achieving high performance with processors that support AVX instructions on the × 64 platform.

II. Factorization Stage

A.Problem definition

Let us consider the direct method for solving linear equation sets.

$$\mathbf{KX} = B,\quad K = K^T,\quad X = \{x_i\},\quad B = \{b_i\},\quad i \in [1, nrhs], \quad (1)$$

where **K** is the symmetric sparse stiffness matrix; **X** and **B** are solution vectors and right-hand parts for multiple load cases; and *nrhs* is the number of right-hand parts. The decomposition is sought in the form of

$$K = L \cdot S \cdot L^T, \quad (2)$$

where **L** is the lower triangular matrix and **S** is the sign diagonal that summarizes the Cholesky decomposition method into a class of indefinite matrices. After factorization (2), forward substitution, diagonal scaling and back substitution are carried out:

$$\begin{aligned} L \cdot Y &= B \rightarrow Y \\ S \cdot Z &= Y \rightarrow Z \\ L^T \cdot X &= Z \rightarrow X \end{aligned} \quad . \quad (3)$$

B.Sparse matrix analysis

First of all the adjacency graph for nodes of the finite element model is reordered to reduce the number of non-zero entries in the factorized stiffness matrix. The number of non-zero entries and the non-zero structure of the sparse lower triangular matrix **L** depend on the reordering method used [3].

Each node of FE model, which has *dof* degrees of freedom, produces a dense submatrix with the dimensions *dof* × *dof*. Therefore, the physical formulation of the problem leads to the division of the original sparse matrix into dense submatrices of relatively small dimensions. To achieve high performance, we should enlarge the dimension of these blocks, and do so in a way that provides for the minimal number of zero entries appearing as the result of such procedure. To this end, we use the algorithm presented in [8]. As a result, matrix **L** is divided into dense rectangular blocks, and the blocks located on the main diagonal are filled completely. The blocks located below the main diagonal may be filled either completely or partially. Memory is not allocated to empty blocks, and for partially filled blocks, only non-zero rows are taken into consideration (Fig. 3).

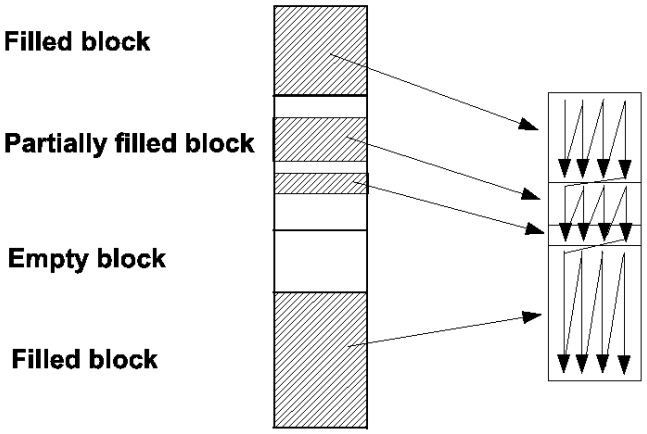


Fig.3. Block-column consisting of empty, partially and completely filled blocks. The packing of data in column major storage is shown to the right.

A more detailed description of the method is provided in [7], [8].

C.Numerical factorization

The algorithm used in this method of left-looking block factorization for the CM mode is shown in Fig. 4, 5.

Factorization is performed in a loop going over *jb* block-columns, the current block column *jb* is corrected by the fully factored block columns located to the left (p.2). Nb is the number of block-columns (p. 1, Fig. 5).

To avoid a situation when two or more threads attempt to modify the same block $\mathbf{A}_{ib,jb}$ in a *jb* block-column, all blocks of current block row are mapped to the same thread. To evenly distribute the processor load, the weight of each block row is calculated (the number of non-zero elements in this block-row), the block rows are sorted in the descending weight order, and then mapped to the threads alternately; with that, the current block row is assigned to the thread with the currently-minimal amount of computation.

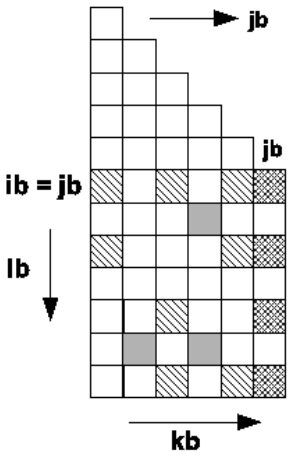


Fig. 4. Left-looking factorization of jb block-column. All block-columns located to the left of jb (kb < jb) are fully factorized.

1. **do** *jb=1,Nb*
2. update of block-column *jb*
prepare parallel tasks Q[*ip*] for update of block-column *jb*
#pragma omp parallel
while(Q[*ip*])
$\{\mathbf{L}_{jb,kb}, \mathbf{L}_{ib,kb}, kb\} = (Q[ip]/\{\mathbf{L}_{jb,kb}, \mathbf{L}_{ib,kb}, kb\})$
$\mathbf{A}_{ib,jb} = \mathbf{A}_{ib,jb} - \mathbf{L}_{ib,kb} \cdot \mathbf{S}_{kb} \cdot \mathbf{L}_{jb,kb}^T$
$(kb \in List_k[jb];\ ib \in L_{kb})$
end while
end of parallel region
end of update
3. factoring of block-column *jb*
$\mathbf{A}_{jb,jb} = \mathbf{L}_{jb,jb} \cdot \mathbf{S}_{jb} \cdot \mathbf{L}_{jb,jb}^T$
#pragma omp parallel for $(ib \in L_{jb})$
$\mathbf{L}_{jb,jb} \cdot \mathbf{S}_{jb} \cdot \mathbf{L}_{ib,jb}^T = \mathbf{A}_{ib,jb}^T \rightarrow \mathbf{L}_{ib,jb}^T$
end of factoring
4. prepare *List_k* for block-columns, which are located
to the right of block-column *jb* and will be updated by it
end **do**

Fig. 5. Looking-left block factorization algorithm

As a result, a queue of tasks Q[*ip*] is created for each *ip* thread, *ip* = 0, 1, …, *np*-1, *np* is a number of threads. Each queue element $\{\mathbf{L}_{jb,kb}, \mathbf{L}_{ib,kb}, kb\}$ contains pointers to the factorized matrix blocks $\mathbf{L}_{jb,kb}$, $\mathbf{L}_{ib,kb}$ and the *kb* index of the sign diagonal block $\mathbf{S}_{kb}$. The *jb* block-column is corrected only by those block-columns that have non-zero blocks $\mathbf{L}_{jb,kb}^T$ in the block row *ib* = *jb* (*kb* ∈ *List_k*[*jb*]).

In the parallel region each thread runs a *while* loop going over its own queue of tasks Q[*ip*] until it is exhausted.

The nearest element is popped from the queue and immediately deleted: $\{\mathbf{L}_{jb,kb}, \mathbf{L}_{ib,kb}, kb\} = (Q[ip]/\{\mathbf{L}_{jb,kb}, \mathbf{L}_{ib,kb}, kb\})$. Then, the task $\mathbf{A}_{ib,jb} = \mathbf{A}_{ib,jb} - \mathbf{L}_{ib,kb} \cdot \mathbf{S}_{kb} \cdot \mathbf{L}_{jb,kb}^T$ is performed. Conditions *ib* ∈ $\mathbf{L}_{kb}$ and *ib* = ∈ $\mathbf{L}_{jb}$ mean that the *ib* index accepts only those values that correspond to the non-zero blocks in the *kb* and *jb* block-columns respectively.

The details of this algorithm are presented in [7].

To ensure high performance, we can use the *dgemm* matrix multiplication procedure, or the microkern_8x4_AVX procedure presented in this paper.

When the *jb* block-column is completely corrected, it is factorized (p.3), and then the *jb* index is pushed to the *List_k* block-column list, which will be updated by the *jb* block-column at the next factorization steps (p. 4).

Therefore, performance at the numerical factorization stage is mainly determined by the performance of the matrix multiplication procedure. Test 2 (see Fig. 1, 2) simulates correction of the *jb* block-column by block-columns located to the left of it. Therefore, this was the test mainly used to test out the microkern_8x4_AVX procedure. Following [6], we will refer to the procedure microkern_8x4_AVX, whose code is written based on the AVX instructions, as microkernel.

D.Microkernel microkern_8x4_AVX

The proposed approach uses the same idea as in the development of the microkernel, based on SSE2 [6], [9], [10]. To achieve high performance, it is necessary to use cache blocking, register blocking, computing vectorization, data repacking in order to reduce the number of cache misses, and to unroll the inner loop to maximize the use of the processor pipelines. Since in the PARFES method, the maximum dimension of block l_b is 120, the l_b, K, N dimensions do not exceed this value. Therefore, division of matrices $\mathbf{A}_{ib}$, **B** and $\mathbf{C}_{ib}$ into blocks (cache blocking) is not required, and the dimension of TLB (translation look aside buffer) will not be exceeded [6].

Modern processors supporting AVX have 256-bit YMM registers, and 16 registers are available on the ×64 platform. Four floating-point double precision words can be loaded into each register, and four additions or four multiplications are carried out per each clock of processor. The register blocking diagram for the ×64 platform is shown in Fig. 6.

The result is stored in 8 registers intended for the elements of matrix $\mathbf{C}_{ib}$. Two registers are used for elements of matrix $\mathbf{A}_{ib}$, one register – for elements of matrix **B** and one register is required to store the intermediate multiplication results. The block dimension is $m_r \times n_r = 8 \times 4$. When the inner loop runs, the elements of matrices $\mathbf{A}_{ib}$ and **B** are repacked to ensure their locations in neighboring RAM addresses. This reduces the number of cache misses and allocates the data in the cache extremely densely. We denote: **AA = repack($\mathbf{A}_{ib}$), BB = repack(B),** where **Dest = repack(Source)** means repacking from array **Source** to array **Dest** (Fig. 6, bottom). The elements of matrices $\mathbf{A}_{ib}$, **B** and $\mathbf{C}_{ib}$ are located in the RAM column-major storage. Prefetch instructions are applied to hide memory system latency.

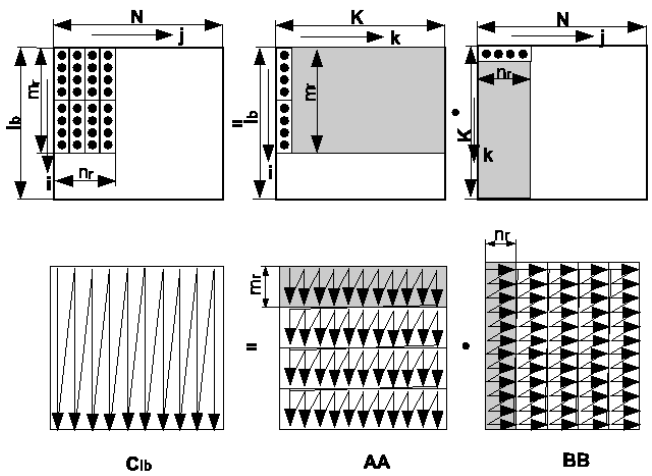


Fig. 6. Diagram of register blocking (top) and data repacking (bottom)

The AVX is used to accelerate the transmission of data when matrices $\mathbf{A}_{ib}$, **B** are repacked into arrays **AA** and **BB** respectively. The pseudo code presenting the microkern_8x4_AVX procedure is shown in Fig 7.

1. **Procedure Pack_BB: B = repack(BB)** (fig. 6, bottom).

2. **Procedure microkern_8x4_AVX:**

$\mathbf{C}_{ib} = beta \cdot \mathbf{C}_{ib} + alpha \cdot \mathbf{A}_{ib} \cdot \mathbf{B}$ (in the future index *ib* is omitted)

AA = repack(A)

```
for(j=0; j<N; j+=nr)
{
   //pBB0 = BB+K*j;  //point to current
   //vertical pane of BB
    for(i=0; i<lb; j+=mr)
    {
      pAA=AA+i*K; //point to current
                  //horizontal pane of AA
      pC=C+ldc*j+i; //point to Cij, ldc =
                    //lb.
      pBB = pBB0;

      //move Cij, Ci+1,j, … , Ci+7, j to cache
      //untill CPU run internal loop
      //move Ci,j+1, Ci+1,j+1, … , Ci+7, j+1 to cache
      //untill CPU run internal loop
      _mm_prefetch((const char *)(pC+ldc),
                   _MM_HINT_T0);
      _mm_prefetch((const char *)(pC+2*ldc),
                   _MM_HINT_T0);
      c1 = _mm256_setzero_pd();  //c1 ← 0
      ..................................
      c8 = _mm256_setzero_pd();  //c8 ← 0
      for(k=0; k<K; k+=16)
      {
        _mm_prefetch((const char *)(pAA+mr),
                     _MM_HINT_T0);
        _mm_prefetch((const char *)(pBB+2*nr),
                     _MM_HINT_T0);
        a0 = _mm256_load_pd(pAA);
        a1 = _mm256_load_pd(pAA+4);
        b0 = _mm256_broadcast_sd(pBB);
        b1 = _mm256_broadcast_sd(pBB+1);
        b2 = _mm256_broadcast_sd(pBB+2);
        b3 = _mm256_broadcast_sd(pBB+3);

        mul = _mm256_mul_pd(a0, b0);
        c1  = _mm256_add_pd(c1, mul);
        mul = _mm256_mul_pd(a1, b0);
        c2  = _mm256_add_pd(c2, mul);

        mul = _mm256_mul_pd(a0, b1);
        c3  = _mm256_add_pd(c3, mul);
        mul = _mm256_mul_pd(a1, b1);
        c4  = _mm256_add_pd(c4, mul);

        mul = _mm256_mul_pd(a0, b2);
        c5  = _mm256_add_pd(c5, mul);
        mul = _mm256_mul_pd(a1, b2);
        c6  = _mm256_add_pd(c6, mul);

        mul = _mm256_mul_pd(a0, b3);
        c7  = _mm256_add_pd(c7, mul);
        mul = _mm256_mul_pd(a1, b3);
        c8  = _mm256_add_pd(c8, mul);
        //and so on 15 times

        pAA += 16*mr;
        pBB += 16*nr;
      }//end k loop

// put alpha*A*B to c1 – c8
      mul =_mm256_set_pd(alpha,alpha,alpha,alpha);
      c1  = _mm256_mul_pd(c1, mul);
      c2  = _mm256_mul_pd(c2, mul);
      c3  = _mm256_mul_pd(c3, mul);
      c4  = _mm256_mul_pd(c4, mul);
      c5  = _mm256_mul_pd(c5, mul);
      c6  = _mm256_mul_pd(c6, mul);
      c7  = _mm256_mul_pd(c7, mul);
      c8  = _mm256_mul_pd(c8, mul);

      if(beta)
      {
        //put alpha*AAi*BBj + beta*CCij to c1 – c8
        b0  = _mm256_set_pd(beta,beta,beta,beta);
        a0  = _mm256_loadu_pd(pC);
        a1  = _mm256_loadu_pd(pC+4);
        mul = _mm256_mul_pd(b0, a0);
        c1  = _mm256_add_pd(c1, mul);
        mul = _mm256_mul_pd(b0, a1);
        c2  = _mm256_add_pd(c2, mul);

        a0  = _mm256_loadu_pd(pC+ldc);
        a1  = _mm256_loadu_pd(pC+ldc+4);
        mul = _mm256_mul_pd(b0, a0);
        c3  = _mm256_add_pd(c3, mul);
        mul = _mm256_mul_pd(b0, a1);
        c4  = _mm256_add_pd(c4, mul);

        a0  = _mm256_loadu_pd(pC+2*ldc);
        a1  = _mm256_loadu_pd(pC+2*ldc+4);
        mul = _mm256_mul_pd(b0, a0);
        c5  = _mm256_add_pd(c5, mul);
        mul = _mm256_mul_pd(b0, a1);
        c6  = _mm256_add_pd(c6, mul);

        a0  = _mm256_loadu_pd(pC+3*ldc);
        a1  = _mm256_loadu_pd(pC+3*ldc+4);
        mul = _mm256_mul_pd(b0, a0);
        c7  = _mm256_add_pd(c7, mul);
        mul = _mm256_mul_pd(b0, a1);
        c8 = _mm256_add_pd(c8, mul);
      }//end if(beta)

      //unload c1 – c8 to matrix C
```

```
    _mm256_storeu_pd(pC,    c1);
    _mm256_storeu_pd(pC+4,  c2);
    pC += ldc;
    _mm256_storeu_pd(pC,    c3);
    _mm256_storeu_pd(pC+4,  c4);
    pC += ldc;
    _mm256_storeu_pd(pC,    c5);
    _mm256_storeu_pd(pC+4,  c6);
    pC += ldc;
    _mm256_storeu_pd(pC,    c7);
    _mm256_storeu_pd(pC+4,  c8);
  }//end i loop
}//end j loop
```

Fig. 7. microkern_8x4_AVX

Here, for ease of understanding the basic idea of the method, we consider only the case when M is a multiple of m_r, N multiple of n_r, and K is a multiple of 16. In the real microkernel, submatrices of dimension $M_1 \times K$ and $K \times N_1$ are extracted from matrices **A** and **B**, where M_1 and N_1 assume the greatest value with the following limitations: ($M_1 \le M$) ^ (M_1 % m_r) = 0, ($N_1 \le N$) ^ (N_1 % n_r) = 0, where a%b means that the remainder after the division of a by b is zero. Therefore, matrices $\mathbf{A}_{ib}$, **B** are divided into blocks, in which the largest submatrices are a multiple of m_r and n_r respectively. The YMM register blocking scheme (Fig. 6) is applied specifically for these submatrices. For the remaining small submatrices, simpler multiplication methods are used.

Matrix B is repacked in a separate procedure, allowing us to use it only once for each *kb* block-column. To produce register blocking, indexes i, j are increased by increments of m_r, n_r correspondingly. The pointers pAA and pBB are set to the beginning of the horizontal strip of matrix AA and the vertical strip of BB (Fig. 6, 8), before loops with indexes i, j are initiated.

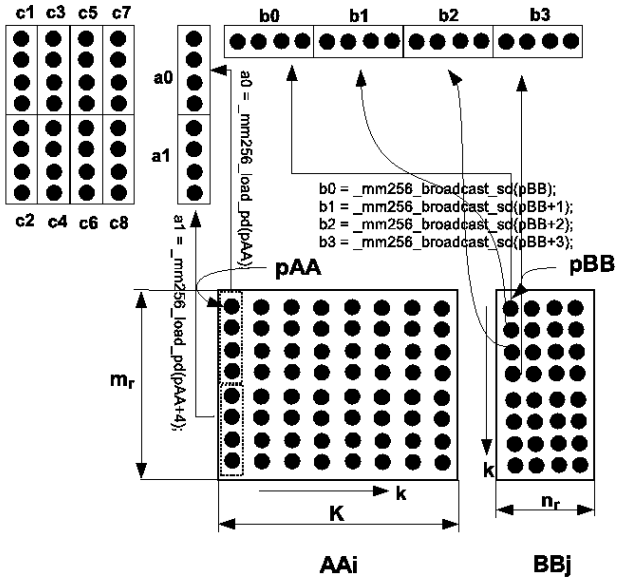


Fig. 8. Map of YMM register's loading

Registers c1 through c8 are zeroed before the loop with index *k* is started. The inner loop with index k is unrolled 16 times. The eight consecutive elements of array **AA** (the entire column of horizontal strip – Fig. 6) are loaded into registers a0, a1 by means of the instruction _mm256_load_pd(...). Each of the four consecutive elements of array **BB** (the entire vertical strip row) is sent to registers b0, b1, b2, b3 correspondingly, by means of the instruction _mm256_broadcast_sd(...) (Fig. 7, 8). As a result, the first element from the vertical strip row is found four times in register b0, the second – four times in register b1, etc.

The contents of register a0 are multiplied by the contents of register b0, and the result is placed into register mul – instruction mul = _mm256_mul_pd(a0, b0). Instruction c1 = _mm256_add_pd (c1, mul) adds up the contents of registers c1 and mul, and sends the result into register c1. Then, the contents of register a1 are multiplied by the contents of register b0, and the result is added to the contents of register c2. The contents of registers a0, a1 are multiplied by b1, and the results are added to the contents of registers c3, c4, etc. respectively. At the end of the loop with index *k*, registers c1 through c8 contain the fully computed elements of the $m_r \times n_r$ block of matrix $\mathbf{C}_{ib}$, which constitute the result of multiplying the horizontal strip $m_r \times K$ of matrix $\mathbf{A}_{ib}$, repacked into array **AA**, by the vertical strip $K \times n_r$ of matrix **B**, repacked into array **BB**.

This result is multiplied by scalar factor *alpha*. If the *beta* coefficient is non-zero, the 8 elements c_{ij}, $c_{i+1,j}$, ..., $c_{i+7,j}$ are loaded into registers a0, a1 by using _mm256_loadu_pd(...). The elements of matrix $\mathbf{A}_{ib}$ are loaded from array **AA** by using instruction _mm256_load_pd (...), because memory for array **AA** is allocated with a 32 byte alignment. Memory for matrix **C** is allocated without the 32 byte alignment, so here we use the _mm256_loadu_pd (...). The elements of matrix **C** held in registers a0, a1 are multiplied by factor *beta* and added to the contents of registers c1 and c2. Then, eight elements from the next column of matrix $\mathbf{C}_{ib}$ – $c_{i,j+1}$, $c_{i+1,j+1}$, ..., $c_{i+7,j+1}$ are loaded into registers a0 and a1, multiplied by factor *beta*, and added to the contents of registers c3, c4, etc. Transition to the next column of matrix $\mathbf{C}_{ib}$ is made by offsetting $ldc = l_b$ of pC pointer. While the current iteration is running, the prefetch instruction is applied to transmit the elements of matrix $\mathbf{C}_{ib}$ from RAM to the cache, as required for the next iteration of the loop with index *i*.

As a result, registers c1 through c8 hold the accumulated result of *alpha*·**AA**i***BB**j+*beta*·**CC**ij, where **AA**i, **BB**j are, respectively, the horizontal strip of matrix $\mathbf{A}_{ib}$, determined by the value of index *i*, and the vertical strip of matrix **B**, defined by the value of index *j*, and **CC**ij – the corresponding block of matrix $\mathbf{C}_{ib}$. Instructions _mm256_storeu_pd (...) unload data from registers c1 through c8 to the corresponding elements of matrix $\mathbf{C}_{ib}$.

III. Numerical results

A. Test 2

The results of test 2, described in the introduction (Fig. 1, 2), have been obtained on two computers and are shown in Table 1.

The first computer has a 16-core AMD Opteron 6276 CPU 2.3/3.2 GHz processor, 64 GB DDR3 RAM, and runs Windows Server 2008 R2 Enterprise SP1, 64 bit. The second computer has a 4-core Intel i7 2760QM CPU 2.4/3.5 GHz processor, 8 GB DDR3 RAM, and runs Windows 7 Professional SP1, 64 bit.

For the ACML 15.2.0 procedure, column for computing on 16 threads (the computer with AMD processor) contains two values: the first (41 469 MFLOPS) corresponds to solving the problem by parallelizing only within the *dgemm* procedure, using its multi-threaded version; while the second (10 061 MFLOPS) – to the use of the single-threaded version of *dgemm* in a parallel OpenMP loop. The first value is used to estimate the top performance of the AMD Opteron 6276 CPU for this test, since the ACML library is best adapted to AMD processors. The second value confirms that the *dgemm* procedure from the ACML library does not work properly in the mode required by PARFES.

A comparison of the results (Table 1) showed that the proposed microkern_8x4_AVX procedure successfully solved this problem on the computer with AMD Opteron 6276 processor, as well as on the computer with Intel i7 2760QM processor.

Next, we consider two real-life problems, taken from the collection of SCAD Soft – a Software Company (www.scadsoft.com) developing software for civil engineering. SCAD is FEA software, which is widely used in the CIS region and has a certificate of compliance to local regulations.

I. Problem 1

A design model of multistorey building contains 2 546 400 equations, consists of triangular, quadrilateral shell finite elements, as well as spatial frame ones (Fig. 9).

The original stiffness matrix contains 27 927 845 nonzero entries, and lower triangular factorized matrix – 1 124 085 204 nonzero entries. METIS reordering method [13] has been used.

The duration and performance of the factorization stage is presented in Table 2. As in the preceding example, the *dgemm* procedure from the Intel MKL 11.0 library does not achieve the desired performance on the computer with AMD Opteron 6276 processor. The *dgemm* procedure from the ACML library works well on a single thread, but when PARFES implements multithreading, the procedure is not performing its task. The microkern_8x4_AVX procedure proposed in this paper demonstrates good results with a single thread as well as during multi-threading. It is interesting to note that on a computer with Intel i7 2760QM processor, this procedure was not inferior to the *dgemm* procedure from the Intel MKL 11.0 library.

B. Problem 2

A design model of soil-structure interaction problem contains 2 989 476 equations (Fig. 10, 11) and consists of triangular, quadrilateral shell finite elements, as well as spatial frame and volumetric finite elements simulating the behavior of the ground.

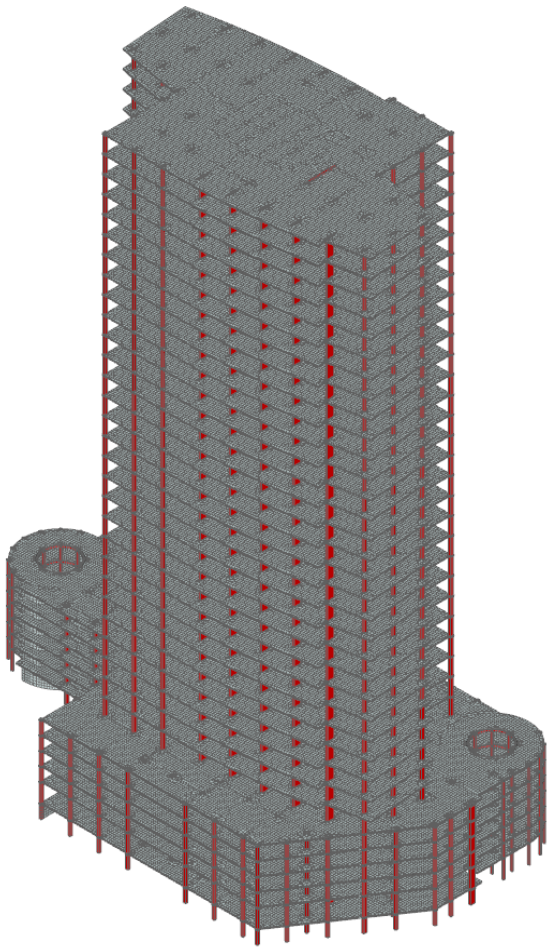


Fig. 9. Problem 1. Design model of multistorey building (2 546 400 equations).

This problem is very challenging for direct methods, because the soil prism, simulated by volumetric finite elements, generates a relatively dense part of a sparse matrix. Of all the methods available for reordering – the minimum degree algorithm MMD [4], the nested dissection method ND [3], the parallel section method [3] in conjunction with the MMD – the most efficient method for this task is METIS [13]. The number of nonzero elements in original matrix is 68 196 176 and in the lower triangular matrix – 4 966 055 936 (37 GB).

The duration of the numerical factorization phase and the performance obtained on the computer with AMD Opteron 6276 processor is shown in Table 3. The solution of this problem on the computer with Intel i7 2760QM processor and 8 GB of RAM was not effective due to the small amount of core memory. PARFES was run in the OOC1 mode, performing a large number of I/O operations. For this reason, performance analysis of the matrix multiplication procedure is not applicable.

The suggested microkernel procedure is slightly inferior to the *dgemm* procedure from the ACML 15.2.0 library on a single thread, but greatly outperforms it on 16 threads. In all cases, the proposed procedure is faster than the *dgemm* procedure from Intel MKL.

TABLE 1.
PERFORMANCE (MFLOPS) OF ALGORITHM **C = C – A·B** FOR MATRICES M × N × K = 2 000 000 × 120 × 120 (A – MATRIX M × K, B – K × N, C – M × N)

Procedure	AMD Opteron 6276 CPU 2.3/3.2 GHz		Intel i7 2760QM CPU 2.4/3.5 GHz	
	Single thread	16 threads	Single thread	4 threads
dgemm MKL 11.0	2 377	24 945	17 921	40 563
dgemm ACML 15.2.0	6 837	41 469 / 10 061	–	–
microkern_8x4_AVX	6 373	50 571	17 582	41 025

TABLE 2.
PROBLEM 1. DURATION (S) AND PERFORMANCE (MFLOPS) OF PARFES ON THE NUMERICAL FACTORIZATION STAGE (PROBLEM 1)

Procedure	AMD Opteron 6276 CPU 2.3/3.2 GHz				Intel i7 2760QM CPU 2.4/3.5 GHz			
	Single thread		16 threads		Single thread		4 threads	
	Duration, s	Perform.	Duration, s	Perform.	Duration, s	Perform.	Duration, s	Perform.
dgemm MKL 11.0	1 139	3 619	160	25 789	330	12 554	191	21 787
dgemm ACML 15.2.0	718	5 743	542	7 628	–	–	–	–
microkern_8x4_AVX	753	5 477	118	34 843	294	14 196	157	27 649

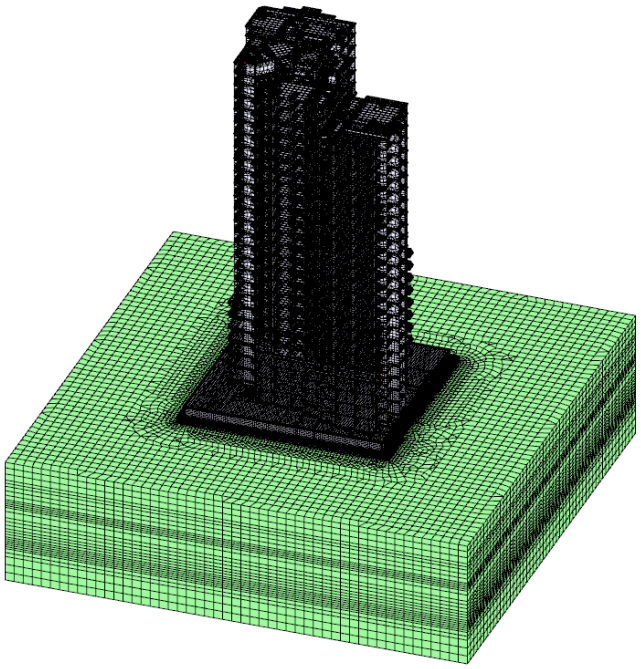


Fig. 10. Problem 2. Design model of soil-structure interaction problem (2,989,476 equations).

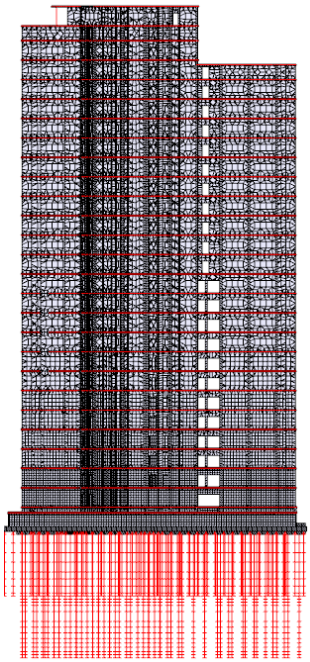


Fig. 11. The pile foundation (ground is hidden)

TABLE 3.
PROBLEM 2. DURATION (S) AND PERFORMANCE (MFLOPS) OF PARFES ON THE NUMERICAL FACTORIZATION STAGE. COMPUTER WITH AMD OPTERON 6276 PROCESSOR (PROBLEM 2)

Procedure	Single thread		16 threads	
	Duration, s	Performance, MFLOPS	Duration, s	Performance, MFLOPS
dgemm MKL 11.0	18 992	3 138	2 123	28 068
dgemm ACML 15.2.0	12 871	4 630	10 897	5 476
microkern_8x4_AVX	13 541	4 400	1 481	40 216

The speed up with the increase in the number of processors is depicted in Fig. 12.

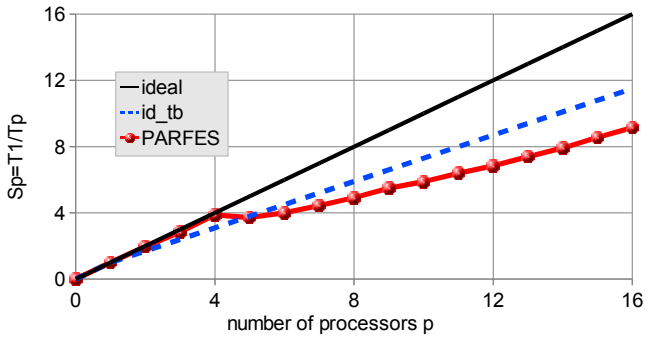


Fig. 12. Speed up with the increase in the number of threads. Ideal – the ideal speed up, id_tb – the ideal speed up on processors with Turbo Core support, PARFES – the real speed up.

The straight line of the "ideal" speed up passes through the points {0, 0}, {1, 1}, {2, 2}, … . This means that if the problem is solved using p threads, we would like to solve it p times faster than when using one thread. The id_tb curve approximates the ideal speed up for processors that support the Turbo Core mode – when a small number of cores is loaded, the processor increases the clock frequency, and when the number of loaded cores increases, reduces the frequency to the nominal value of 2.3 GHz. This curve is represented by a square parabola passing through the points {0, 0}, {1, 1}, {16, 11.5}. The ordinate of the last point was obtained as 16 × (minimum clock frequency of the processor) / (maximum clock frequency of the processor) = 16·3.2/2.3 = 11.5.

When using up to 4 threads, the speed up of PARFES is almost perfect. We explain the anomaly at p = 5 by the features of the Turbo Core control on this processor, because testing of PARFES on computers with different processors [7], [8] does not produce such behavior. The speed up of the method when p > 4 is stable up to p = 16, although lower than for the id_tb curve.

IV. Conclusion

Developing the microkernel procedure, based on AVX, in the parallel direct solver PARFES designed to solve problems of structural and solid mechanics that arise as a result of applying the finite element method, significantly accelerates matrix factorization on computers with AMD Opteron 6276 processor, Bulldozer architecture, while maintaining high performance and competitiveness with the Intel MKL *dgemm* procedure on computers with Intel processors.

References

[1] P. R Amestoy, I. S Duff, and J. Y L'Excellent, "Multifrontal parallel distributed symmetric and unsymmetric solvers," *Comput. Meth. Appl. Mech. Eng*., vol. 184, pp 501–520, 2000.

[2] ACML 15.2.0. URL: http://developer.amd.com/tools/cpu-development/amd-core-math-library-acml/ (accessed 17.11.2012).

[3] A. George and J. W. H. Liu, *Computer solution of sparse positive definite systems*. New Jersey : Prentice-Hall, Inc. Englewood Cliffs, 1981.

[4] A. George and J. W. H. Liu, "The Evolution of the Minimum Degree Ordering Algorithm," SIAM Rev, vol. 31, pp. 1–19, March, 1989.

[5] N. I. M Gould, Y. Hu and J. A. Scott, "A numerical evaluation of sparse direct solvers for the solution of large sparse, symmetric linear systems of equations," Technical report RAL-TR-2005-005, Rutherford Appleton Laboratory, 2005.

[6] K. Goto and R. A. Van De Geijn, "Anatomy of High-Performance Matrix Multiplication," *ACM Transactions on Mathematical Software*, vol. 34 (3), pp. 1–25, 2008.

[7] S. Fialko, "PARFES: A method for solving finite element linear equations on multi-core computers," *Advances in Engineering software,* vol. 40, 12, pp. 1256 – 1265, 2010.

[8] S. Fialko, "Parallel Finite Element Solver for Multi-Core Computers", *Federated Conference on Computer Science and Information Systems*, September 9–12, 2012, Wrocław, Poland. IEEE Xplore Digital Library, 978-83-60810-51-4, IEEE Catalog Number CFP1285N-USB, pp. 1 – 8. URL: http://proceedings.fedcsis.org/2012/pliks/101.pdf .

[9] S. Fialko, "The block substructure multifrontal method for solution of large finite element equation sets," *Technical Transactions*, 1-NP, issue 8, pp. 175 – 188, 2009.

[10] S. Fialko, *The direct methods for solution of the linear equation sets in modern FEM software*. Moscow: SCAD SOFT, 2009 (in Russian).

[11] Intel® Math Kernel Library Reference Manual. Document Number: 630813-029US. URL: http://www.intel.com/software/products/mkl/docs/WebHelp/mkl.htm.

[12] Intel MKL 11.0 release notes. URL: http://software.intel.com/en-us/articles/intel-mkl-110-release-notes/ (accessed 17.11.2012).

[13] G. Karypis and V. Kumar, "METIS: Unstructured Graph Partitioning and Sparse Matrix Ordering System,". Technical report, Department of Computer Science, University of Minnesota, Minneapolis, 1995.

[14] Optimize for Intel® AVX Using Intel® Math Kernel Library's Basic Linear Algebra Subprograms (BLAS) with DGEMM Routine. URL: http://software.intel.com/en-us/articles/optimize-for-intel-avx-using-intel-math-kernel-librarys-basic-linear-algebra-subprograms-blas-with-dgemm-routine/ (accessed 19.11.2011).

[15] D. Pardo, Myung Jin Nam, Carlos Torres-Verdín, Michael G. Hoversten and Iñaki Garay, "Simulation of marine controlled source electromagnetic measurements using a parallel Fourier hp-finite element method," *Comput. Geosci*., vol. 15, pp. 53–67, 2011.

[16] O. Schenk, K. Gartner, "Two-level dynamic scheduling in PARDISO: Improved scalability on shared memory multiprocessing systems," *Parallel Computing,* vol. 28, pp. 187–197, 2002.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 455–462

Library for Matrix Multiplication-based Data Manipulation on a "Mesh-of-Tori" Architecture

Maria Ganzha, Marcin Paprzycki
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland
Email: firstname.lastname@ibspan.waw.pl

Stanislav Sedukhin
University of Aizu
Aizu Wakamatsu, Japan
Email: sedukhin@u-aizu.ac.jp

***Abstract*—Recent developments in computational sciences, involving both hardware and software, allow reflection on the way that computers of the future will be assembled and software for them written. In this contribution we combine recent results concerning possible designs of future processors, ways they will be combined to build scalable (super)computers, and generalized matrix multiplication. As a result we propose a novel library of routines, based on generalized matrix multiplication that facilitates (matrix / image) manipulations.**

I. Introduction

SINCE the early 1990's one of the important factors limiting computer performance became the ability to feed data to the, increasingly faster, processors. Already, in 1994 authors of [1] discussed problems caused by the increasing gap between the speeds of memory and processors. Their work was followed, among others, by Burger and Goodman ([2]), who were concerned with the limitations imposed by the memory bandwidth on the development of computer systems. In 2002, P. Machanick presented an interesting survey ([3]) in which he considered the combined effects of doubling of processor speed (predicted by Moore's Law) and the 7% increase in memory speed, when compared in the same time scale.

The initial approach to address this problem was through introduction of memory hierarchy for data reuse (see, for instance, [4]). In addition to the registers, CPUs have been equipped with small fast cache memory. As a result systems with 4 layers of latency were developed. Data could be replicated and reside in (1) register, (2) cache, (3) main memory, (4) external memory. Later on, while the "speed gap" between processors and memory continued to widen, multi-processor computers gained popularity. As a result, systems with an increasing number of latencies have been built. On the large scale, data element could be replicated and reside in (and each subsequent layer means increasing / different latency of access): (1) register, (2) level 1 cache, (3) level 2 cache, (4) level 3 cache, (5) main memory of a (multi-core / multi-processor) computer, (6) memory of another networked computer (node in the system), (7) external device. Obviously, such complex structure of a computer system resulted in need for writing complex codes to efficiently use it. Data blocking and reuse became the method of choice for solution of large computational problems. This method was applied not only for multi-processor computers, but also computers with processors consisting of multiple computational units (e.g. cores, processors, etc.). In this context, let us note that as the number of computational units per processor is systematically increasing, the inflation adjusted price of a processor remains the same. As a result, the price per computational operation continues to decrease (see, also [5]).

While a number of approaches have been proposed to deal with the memory wall problem (e.g. see discussion of 3D memory stacking in [6]), they seem to only slow down the process, rather than introduce a radical solution. Note that, introduction of multicore processors resulted in (at least temporary) sustaining the Moore's Law and thus further pushing the performance gap (see, [3], [7]). Here, it is also worth mentioning recent approach to reduce memory contention via data encoding (see, [8]). The idea is to allow for hardware-based encoding and decoding of data to reduce its size. Since this proposal is brand new, time will tell how successful it will be. Note, however, that also this proposal is in line with the general observation that "computational hardware" (i.e. encoders and decoders) is *cheap*, and should be used to reduce volume of data *moved* between processor(s) and memory.

Let us now consider one of the important areas of scientific computing – computational linear algebra. Obviously, here the basic object is a matrix. While, one dimensional matrices (vectors) are indispensable, the fundamental object of majority of algorithms is a 2D, or a 3D, matrix. Upon reflection, it is easy to realize that there exists a conflict between the structure of a matrix and the way it is stored and processed in most computers. To make the point simple, 2D matrices are rectangular (while 3D matrices are cuboidal). However, they are *stored* in one-dimensional memory (as a long vector). Furthermore, in most cases, they are *processed* in a vector-oriented fashion (except for the SIMD-style array processors). Finally, they are sent back to be *stored* in the one-dimensional memory. In other words, data arrangement natural for the matrix is neither preserved, nor taken advantage of, which puts not only practical, but also theoretical limit on performance of linear algebra codes (for more details, see, [9]).

Interestingly, similar disregard to the natural arrangement of data concerns also many "sensor systems." Here, the input image, which is square or rectangular, is read out serially,

978-83-60810-53-8/$25.00

pixel-by-pixel, and is send to the CPU for processing. This means that the transfer of pixels destroys the 2D integrity of data (an image, or a frame, start to exist **not** in their natural layout). Separately, such transfer introduces latency caused by serial communication. Here, the need to transfer large data streams to the processor may prohibit their use in applications, which require (near) real-time response [10]. Note that large data streams exist not only in scientific applications. For instance modern digital cameras capture images consisting of 22.3×10^6 pixels (Cannon EOS 5D Mark III [11]) or even 36.3×10^6 pixels (Nikon D800 [12]). What is even more amazing, recently introduced Nokia phone (Nokia 808 PureView [13]) has camera capturing 41×10^6 pixels.

For the scientific / commercial sensor arrays, the largest of them seems to be the 2-D pixel matrix detector installed in the Large Hadron Collider in CERN [14]. It has 10^9 sensor cells. Similar number of sensors would be required in a CT scanner array of size approximately $1m^2$, with about 50K pixels per $1cm^2$. In devices of this size, for the (near) real-time image and video processing, as well as a 3-D reconstruction, it would be natural to load data directly from the sensors to the processing elements (for immediate processing). Thus, a focal-plane I/O, which can map the pixels of an image (or a video frame) directly into the array of processors, allowing data processing to be carried out immediately, is highly desired. The computational elements could store the sensor information (e.g. a single pixel, or an array of pixels) directly in their registers (or local memory of a processing unit). Such an architecture has two potential advantages. First, cost can be reduced because there is no need for memory buses or a complicated layout. Second, speed can be improved as the integrity of input data is not destroyed by serial communication. As a result, processing can start as soon as the data is available (e.g. in the registers). Note that proposals for similar hardware architectures have been outlined in [15], [16], [17]. However, all previously proposed focal-plane array processors were envisioned as a mesh-based interconnect, which is good for the local data reuse (convolution-like simple algorithms), but is not proper to support the global data reuse (matrix-multiplication-based complex algorithms).

Separately, it has been established that computational linear algebra can be extended through the theory of algebraic semirings, to subsume large class of problems (e.g. including a number of well known graph algorithms). The theoretical mechanism is named Algebraic Path Problem (APP). As shown, for instance, in ([18]), there is an interesting link between the arithmetical fused multiply and add (FMA) operation, which is supported in modern hardware, and the FMAs originating from other semirings that are not. Specifically, if it was possible to modify the standard FMA to include operations from other semirings (in a way similar to the proposals of KALRAY; [19]), and thus develop a generalized FMA, it could be possible to speed-up large class of APP problems at least 2 times ([18]).

Let us now assume the existence of a computational unit that satisfies the above requirements: (1) accepts input from the sensor(s) and transfers it directly to its operational registers / local memory; (2) is capable of generalized FMA operations. The latter requirement means that such FMA should store (in its registers) constants needed to efficiently perform FMA operations originating from various semirings. Let us name it the *extended generalized FMA; EG FMA*. Recall, that the cost of computational units (of all types) is systematically decreasing ([5]). Therefore, cost of the *EG FMA* unit should not be much higher than that of a standard FMAs found in today's processors. Hence, it is easy to imagine m(b)illions of them "purchased" for a reasonable price. As stated above, such *EG FMAs* should be connected into a square array that will match the shape of the input data. Let us now describe how such system can be build.

II. Mesh-of-tori interconnection topology

Since early 1980's a number of topologies for supercomputer systems have been proposed. Let us omit the unscalable approaches, like a bus, a tree, or a star. The more interesting topologies (from the 1980's and 1990's) were:

- hypercube – scaled up to 64000+ processor in the Connection Machine CM-1,
- mesh – scaled up to 4000 processors in the Intel Paragon,
- processor array – scaled up to 16000+ processor in the MassPar computer,
- rings of rings – scaled up to 1000+ processors in the Kendall Square KSR-1 machines
- torus – scaled up to 2048 units in the Cray T3D

However, all of these topologies suffered from the fact that at least some of the elements were reachable with a different latency than the others. This means, that algorithms implemented on such machines would have to be asynchronous, which works well, for instance, for ising-model algorithms similar to [20], but is not acceptable for a large set of computational problems. Otherwise, extra latency had to be introduced by the need to wait for the information to be propagated across the system.

To overcome this problem, recently, a new (*mesh-of-tori*; *MoTor*) multiprocessor system topology has been proposed ([21], [22]). The fundamental (*indivisible*) unit of the *MoTor* system is a μ-Cell. The μ-Cell consists four computational units connected into a 2×2 doubly-folded torus (see, Figure 1). Logically, an individual μ-Cell is surrounded by so-called membranes that allow it to be combined into larger elements through the process of cell-fusion. Obviously, collections of μ-Cells can be split into smaller structures through cell division. In Figure 1, we see total of 9 μ-Cells logically fused into a single macro-μ-Cell consisting of 4 μ-Cells (combined into a 2×2 doubly folded torus), and 5 separate (individual) μ-Cells. Furthermore, in Figure 2 we observe all nine μ-Cells combined into a single system (a 3×3 doubly folded torus). Observe that, when the 2×2 (or 3×3) μ-Cells are logically fused (or divided), the newly formed structure remains a doubly folded torus. In this way, it can be postulated that the single μ-Cell represents the "image" of the whole system. While in earlier publications (e.g. [22], [23], [24], [25]

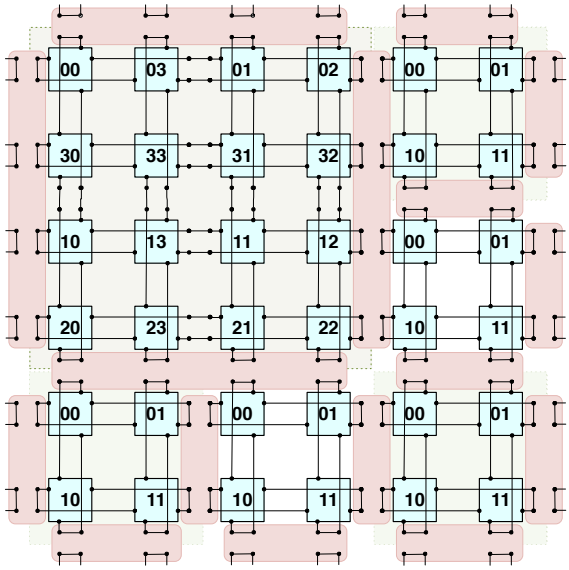


Figure 1. 9 μ-Cells fused into a single 2×2 "system," and 5 separate μ-Cells

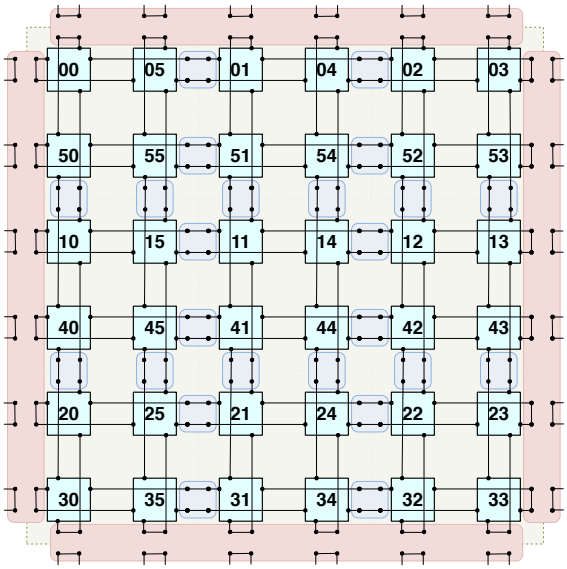


Figure 2. 9 μ-Cells fused into a single 6×6 EG FMA system

the computational units were mostly treated as "theoretical entities," in the context of this paper we assume that each one of them is the *EG FMA* described above. However, analysis of cell connectivity in Figure 1 shows that the model *EG FMA* proposed in the previous section has to be complemented by four interconnects that allow construction of the *MoTor* system. Therefore, from here on, we will understand the *EG FMA* in this way. Furthermore, we will keep in mind that the *MoTor* architecture is build from *indivisible* μ-Cells, each consisting of four, interconnected into a doubly folded torus *EG FAMs*.

Let us now observe that the proposed *MoTor* topology has similar restriction as the array processors from the early 1990's. To keep its favorable properties, the system must be square. While this was considered an important negative (flexibility limiting) factor in the past, this is no longer the case. When the first array processors were built and used, arithmetical operations and memory were "expensive." Therefore, it was necessary to avoid performing "unnecessary" operations (and maximally reduce the memory usage). Today, when GFlops costs about 50 cents (see, [5]) and this price is systematically dropping, and when laptops come with 8 Gbytes of RAM (while some cell phones come with as much as 64 Gbytes of flash memory on a card), it is data movement / access / copying that is "expensive" (see, also [26]). Therefore, when matrices (images) are rectangular (rather than square), it is reasonable to assume that one could just pad them up, and treat them as square. Obviously, since the μ-Cell is a single indivisible element of the $MoTor$ system, if the matrix is of size $N \times N$ then N has to be even.

Observe that there are two sources of inspiration for the $MoTor$ system: (i) matrix computations, and (ii) processing data from, broadly understood, sensor arrays (e.g. images). Furthermore, we have stated that the extended generalized FMA can contain a certain number of data registers to store (a) the needed scalar elements originating from various semirings, (b) elements of special matrices needed for matrix transformations (see, below), as well as (c) data that the FMA is to operate on. However, we also consider the possibility that each FMA may have a "local memory" to allow it to process "blocks of data." This idea is based on the following insights. First, if we define a pixel as " the smallest single component of a digital image" (see, [27]), then the data related to a single pixel is very likely to be not larger than a single 24 bit number. Second, in early 2013 the largest number of FMA units combined in a single computer system was 5.2×10^6. This means that, if there was a one-to-one correspondence between the number of FMA units and the number of "sensed pixels" then the system could process stream of data from a 5.2 Megapixel input device (or could process a matrix of size $N \simeq 2200$).

Let us now consider, development of the *MoTor*-based system. In the initial works, e.g. in [22], links between cells have been conceptualized as programmable abstract links (μ-Cells were surrounded by logical membranes that could be fused or divided as needed, to match the size of the problem). Obviously, in an actual system, the abstract links and membranes could be realized logically, while the whole system would have to be hard-wired to form an actual $MoTor$ system of a specific size. Therefore to build a large system with M^2 μ-Cells (recall the assumption that the mesh will constitute a square array), it can be expected that their groups will be combined into separate "processors," similarly to multicore / multi-FMA processors of today. As what concerns cell fusion and division, it will be possible to assemble sub-system(s) of a needed size, by logically splitting and/or fusing an appropriate

number of cells within the $MoTor$ system. However, it is worthy to stress that, while the theoretical communication latency across the mesh-of-tori system is uniform, this may not be the case when the system will be assembled from processors constituting logical (and in some sense also physical) macro-μ-Cells. In this case it may be possible that the communication within the processor (physical macro-μ-Cell) will be slightly faster than between processors. Therefore, the most natural split would be such that would involve complete macro-μ-Cells (processors). However, let us stress that, the design of the mesh-of-tori topology does not distinguish between the connections that are "within a chip" and "between the chips." Therefore, the communication model used in the algorithms described in [22], [23], and considered in subsequent sections, is *independent* of the hardware configurations.

Finally, let us consider the input from the sensor array (or sending a matrix) into the mesh-of-tori type system. As shown in [22], any input that is in the canonical (square matrix) arrangement, is not organized in a way that is needed for the matrix processing in a (doubly folded) torus. However, adjusting the data organization (e.g. to complete a 2D $N \times N$ DFT), requires 2 matrix multiplications (left and right multiplication by appropriate transformation matrices, see below). These two multiplications require $2N$ time steps on a *MoTor* architecture. Next, after the processing is completed, the canonical arrangement can be restored, by reversing the original transformation. Here, again, two multiplications are needed and their cost is $2N$ time steps. For the details about the needed transformations and their realizations as a triple matrix multiplication, see [22].

III. Data manipulations in a $MoTor$ system

Let us summarize points made thus far. First, we have refreshed arguments that there is an unfulfilled need for computer systems that (1) have focal-plane I/O that, among others, can feed data from sensors directly to the operand registers / memory of extended FMA units (generalized to be capable of performing arithmetical operations originating from different semirings), (2) operate on matrices treating them as square (or cuboidal, e.g. tensor) objects, (3) are developed in such a way that (i) minimizes data movement / access / copying / replication, (ii) maximizes data reuse, and (iii) is aware of the fact that arithmetical operations are cheap in comparison with any form of data "movement." Such systems are needed not only to process data originating from the Big Hadron Collider, but also for everyday electronics. Here, it is worth mentioning that virtual reality and 3D media (part of the new enterprises, so called *creative industries*) are in the latter category, illustrating where the computational power is going to be needed in an increasing rate, beyond the classic domains of scientific computing.

Second, we have briefly outlined key features of the, recently proposed, mesh-of-tori topology, which has some favorable features and naturally fits with the proposed *EG FMAs*. Furthermore, we have pointed to some issues that are likely to be encountered in the development of (large-scale) $MoTor$ based systems. Let us now assume, that the, just proposed, $MoTor$ computer systems have been built. In ([22], [23], [25]) it was shown that a large number of matrix operations / manipulations can be unified through the use of a generalized matrix multiply-and-update (MMU) operation. However, in this context it is important to realize that one more problem we are facing today is the increasing complication of codes that are being developed to take advantage of current computer architectures (see, for instance [28], [26]). This being the case, a return to simplicity is needed. In the remaining sections of this paper we will illustrate how a MATLAB / MATHEMATICA style (meta-level) approach can be used to build a library of matrix manipulation operations that, among others, can be implemented on the proposed $MoTor$ computer architecture. This library will be uniformly based on the "fused" matrix multiply-and-update (MMU) operation.

A. Basic operations

To proceed, we will use the generalized matrix multiply and update operation in the form (as elaborated in [28]):

$$\mathtt{C} \leftarrow \mathtt{MMU}[\otimes, \oplus](\mathtt{A}, \mathtt{B}, \mathtt{C}) : \mathtt{C} \leftarrow \mathtt{C} \oplus \mathtt{A}^{\mathtt{N/T}} \otimes \mathtt{B}^{\mathtt{N/T}}. \qquad (1)$$

Here, A, B and C are square matrices of (even) size N (recall the, above presented, reasons for restricting the class of matrices); while the $\otimes, \oplus$ operations originate from a scalar semiring; and $\mathtt{N/T}$ specify if a given matrix is to be treated as being in a canonical (`N`) or in a transposed (`T`) form, respectively.

In what follows, we present a collection of matrix / image manipulations that can also be achieved through matrix multiplication. While they can be implemented using *any* matrix multiplication algorithm, we use this as a springboard to further elaborate the idea of *MoTor* system, and a library of routines that can complement it. Note that, for simplicity of discussion (and due to the lack of space), in what follows we only discuss the special case when a single *EG FMA* stores scalar data elements. However, as discussed in [22], [23], [25], [24] all matrix manipulations can be naturally extended to blocked algorithms. Therefore, we actually do not contradict our earlier assumption that each *EG FMA* holds a block of data (e.g. pixel array, or a block of a matrix).

1) Reordering for the mesh-of-tori processing: Let us start from the above mentioned fact that the canonical form of the matrix (image) fed to the $MoTor$ system through the focal-plane I/O is not correct for further (parallel) processing on a doubly folded torus. As shown in [23], the proper format can be obtained by corresponding linear transform through two matrix-matrix multiplications. Specifically, matrix product in the form $M \leftarrow R \times A \times R^T$, where A is the original / input $(N \times N)$ matrix that is to be transformed, M is the matrix in the format necessary for further processing on the mesh-of-tori system, and R is the format rearranging matrix (for details of the structure of the R matrix, consult [23]). Taking into account the implementation of the generalized MMU, proposed in [28], the needed transformation is:

$$M = R \cdot A \cdot transpose(R). \quad (2)$$

Note that on the $MoTor$ system: (a) operation $R \cdot A$ is performed in place and requires N time steps, (b) operation $A \cdot transpose(R)$ is performed in place, requires N time steps and is implemented as a parallel matrix mutiplication with a different data movement (operation scheduling) that the standard multiplication. In other words, the matrix arrangement remains unchanged and well-known problems related to row vs. column matrix storage (see, for instance, [29]) do not materialize (for more details, see [30]).

Observe that when instantiating the $MoTor$ system, it is assumed that appropriate matrix R will be pre-loaded into the macro-μ-Cell, upon its creation. In earlier work, see for instance [28] and references to earlier work of S. Sedukhin collected there, it was assumed that the generalized FMA will store in its operand registers the (scalar) constants originating from implemented semiring(s) and representing elements $\bar{0}$ and $\bar{1}$. Here, we assume that, in addition to the separate operand registers dedicated to each $\bar{0}, \bar{1}$ element originating from the semiring implemented in the hardware, in a separate operand register an appropriate element of a transformation matrix needed to perform operations summarized in this paper will be pre-loaded. It is in this way, that the matrix R will be pre-loaded into the $MoToR$ system.

Observe that in the $MoTor$ system, the size of the "logical" system (macro-μ-Cell) can vary with time and be changed through cell fusion and division. This means that, after a group of μ-Cells is fused (split), some of the "transformation matrices" will have to be re-instantiated. However, this will not concern matrices like `ONES` (see, below) that preserve format during cell fusion / division. Nevertheless, matrix R will have to be re-initialized each time cells are fused / divided. We summarize these considerations, for the "special matrices" identified in this paper, in Table I.

When considering the implementation of the transformation, observe that the information about the R matrix does not need to be known to the user (only to the implementer). This being the case, we can define the $Canonical\_to\_MoTor$ function that will have as its input the matrix A in the canonical form, and as its output matrix M in the form ready to use in the $MoTor$ system. This function will perform operations from equation 2, while hiding matrix R from the user. Obviously, an inverse function $MoTor\_to\_Canonical$, that will perform operation $A \leftarrow R^T \cdot M \cdot R$ (with the same matrix R), will restore the matrix to its original (canonical) format in $2N$ time steps. Obviously, these two functions make sense only in the context of the $MoTor$ system. Specifically, such transformations can be performed on any computer system, but they are useless if that system has a different topology.

2) Row and column permutations: It is a well known fact, that row and column permutations can be represented as matrix-matrix multiplications. Specifically, permutation of rows of matrix A is achieved through left hand side multiplication by an appropriate matrix P ($A' \leftarrow P \cdot A$), while column permutation is achieved through the right hand side multiplication by an appropriate matrix Q ($A' \leftarrow A \cdot Q$). Both matrices P and Q are identity matrices, with two elements (corresponding to appropriate rows or columns) modified. While the matrix multiplication is typically treated as a convenient notation used in theoretical linear algebra, in computational practice row and column permutations are usually implemented as vector operations. However, in the $MoTor$ approach, row and column permutations will be achieved through actual $N \times N$ matrix multiplication, performed in place, in $O(N)$ time steps.

As far as the implementation is concerned, matrix P will be initially stored in the meta-μ-Cell as a copy of the identity matrix(I). Note that this means that this matrix will be actually represented in separate operand registers of appropriate *EG FMAs*. Let us now introduce function $Row\_permute(A, i, j)$. This function, when called, will send information to appropriate four *EG FMAs* $((i,i), (j,j), (i,j), (j,i))$ to flip their values from $\bar{0}$ to $\bar{1}$ and vice-versa. Depending on the implementation, this should be achieved in no more than 4 time steps. Next, the actual matrix multiplication will be performed. Finally, the four processing units (belonging to matrix P) that changed their values, will revert to the original ones (again, in no more than 4 time steps). The same approach will be used in the case of column permutation (function $Column\_permute(A, i, j)$). Observe that, while there exist algorithms that ask for ($A' \leftarrow P \cdot A \cdot Q$), these two operations (left and right side multiplication) do not have to be performed simultaneously (in this paper we do not consider a more general case of performing concurrently triple matrix multiplications). This means that actually, we do not need to store two separate matrices P and Q. All that is needed is a single `PERMUT` matrix than can be used by the implementer to support both operations (this matrix is not being made available to the user). Potential use of the actual identity matrix (in place of the separate `PERMUT` matrix) needs to be further considered, before such decision could be made. As what concerns the effect of cell fusion and splitting, it should be obvious from Figures 1 and 2 that it is during this process it is necessary to reinitialize both the identity and the `PERMUT` matrices.

3) Scalar data replication (broadcast): Let us now consider replication of a data element across all processors in the system (operation that in MPI [31] is known as *MPI_BCAST*). As shown in [23], this can be achieved through two matrix multiplications. The appropriate triple has the form $B \leftarrow$ `LBCAST` $* D *$ `RBCAST`, where B is the resulting matrix with all of its elements equal to the replicated one; D is a zero matrix with the element to be replicated in position $d(i, j)$; `LBCAST` is a matrix with all zeros except of the ones in column i; and `RBCAST` is a zero matrix with ones in row j. Obviously, on the $MoTor$ system (since the broadcast involves 2 matrix multiplications) it will be completed in place, in $2N$ time steps.

Based on the material presented thus far, we propose the following implementation of broadcast of a selected element across all processors of a $MoTor$ system. Assume that ma-

trices LBCAST and RBCAST are zero matrices with ones in column 1 and row 1 respectively. Furthermore, matrix D is a copy of the $\bar{0}$ matrix, which is going to be used in both multiplications. In the first step, the selected element is send to the *EG FMA* located in position $(1,1)$ of the $MoTor$ system and stored in the operand register corresponding to $D(1,1)$. Next, the MMU operation is invoked twice ($B \leftarrow$ LBCAST $*\, D *$ RBCAST) within a $ElBcast$ function, which has the form: $ElBcast(element)$, where the $element$ specifies element that should be replicated. As a result, in $2N$ time steps, the selected element is replicated to all elements of matrix B and thus made available across the $MoTor$ system. Finally, the $D(1,1)$ element is zeroed. Note that, while matrices LBCAST and RBCAST have to be reinstantiated, the matrix D, being a copy of the zero matrix remains unchanged during cell fusion / division. Possibility of use of the actual zero matrix, instead of matrix D has to be further evaluated. Obviously, matrices LBCAST and RBCAST are available only to the implementer, while being hidden from the user.

4) Global reduction and broadcast: The next matrix operation that can be formulated in terms of matrix multiplications, is the *global reduction and broadcast*. As seen in [23], when the standard arithmetic is applied, and matrix A is multiplied from both sides by a matrix of ones (matrix with all elements equal to one, let us name it ONES), then the resulting matrix will have its elements equal to the sum of all elements of A. On a mesh-of-tori system, this can be implemented in place, in $2N$ time steps, when the matrix ONES is available (preloaded) in all *EG FMAs* of the system.

However, recall that our approach is based on use of the generalized MMU. Thanks to this, we can apply operations originating from different semirings. Here, particularly interesting would be semirings, in which the addition and multiplication operations are defined as $(\times, max)$ or $(\times, min)$. In this case, the "generalized reduction and broadcast" operation is going to consist of two generalized MMUs (represented in notation from the equation 1):

$$\texttt{MMU}[\otimes,\oplus](\texttt{A},\texttt{ONES},\texttt{TEMP}) : \texttt{TEMP} \leftarrow \texttt{TEMP} \oplus \texttt{A} \otimes \texttt{ONES};$$
$$\texttt{MMU}[\otimes,\oplus](\texttt{ONES},\texttt{TEMP},\texttt{RESULT}) :$$
$$\texttt{RESULT} \leftarrow \texttt{RESULT} \oplus \texttt{ONES} \otimes \texttt{TEMP}.$$

Here, operations $[\otimes,\oplus]$ are defined in an appropriate semiring, while matrices TEMP and RESULT are initialized as copies of the $\bar{0}$ matrix (zero matrix for a given semiring). Finally, ONES is a matrix of all ones, where the "one" element originates from a given scalar semiring (its element $\bar{1}$).

Under these assumptions it is easy to see that we can define at least three functions that will have the same general form, while being based on different semirings: $AddBcast$ – realizing summation of all elements in a matrix and broadcasting the result to all processors (function based on the standard arithmetic); $MaxBcast$ finding the largest element in a matrix and broadcasting it to all processors (based on the $(\times, max)$ semiring); and $MinBcast$ finding the smallest element in the matrix and broadcasting it to all processors (based on the $(\times, min)$ semiring). Each of these functions will be completed in place, in $2N$ time steps, through 2 generalized matrix multiplications. Observe that, for all practical purposes, matrix ONES does not have to be instantiated. It consists of $\bar{1}$ elements that, according to our assumptions, are already stored in operand registers of the *EG FMAs*. Finally, note that (regardless of the way it will finally be instantiated in the *MoTor* system) matrix ONES is independent of the size of the macro-μ-Cell and remains unchanged and available after cell fusion / division operation. As previously, all information about very "existence" of matrices ONES and TEMP, and initialization of matrices TEMP and RESULT is going to be hidden from the user.

B. Matrix (image) manipulations

Let us now consider three simple matrix *manipulations* that can be achieved with help of matrix multiplication. While they are presented as matrix operations, their actual value can be seen when the underlying matrices represent images (e.g. each matrix element represents a pixel, or a block of pixels).

1) Upside-down swap: Image (matrix) upside down swap can be achieved by multiplying the matrix from the left hand side by the $SWAP$ matrix, which has the following form:

$$SWAP = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}$$

Obviously, we assume that on the $MoTor$ system, matrix SWAP will be instantiated when macro-μ-Cell(s) will be created (appropriate elements will be stored in separate operand registers of the *EG FMAs*). However, it will not be made available to the user. This means that the upside-down swap will be achieved by calling a $UDswap(A)$ function, and completed in place, in N time steps. Matrix SWAP will have to be re-initialized after each μ-Cell fusion or division.

2) Left-right swap: The left-right image (matrix) swap can be achieved the same way as the upside-down swap, with the only difference being that the image matrix A is going to be multiplied by the SWAP matrix from the right hand side. Therefore, on the $MoTor$ system, the left-right swap will be completed in place, in N time steps, by calling the $LRswap(A)$ function. All the remaining comments, concerning the SWAP matrix, presented above, remain unchanged.

3) Rotation: Interestingly, combining the two swaps into a single operation (multiplication of a given image / matrix A from left and right by the matrix SWAP) results in rotation of the matrix / image A by $180°$. Obviously, from the above follows that on the $MoTor$ system, this operation can be completed in place, in $2N$ steps using two matrix multiplications, by calling an appropriately defined $Rotate(A)$ function.

IV. Towards library of matrix multiplications based data manipulations

Let us now summarize the above considerations from the point of view of development of a library of operations that can be performed on matrices / images though generalized matrix

Table I
SUMMARY OF FUNCTIONS PROPOSED FOR THE LIBRARY

Functionality	Function	Matrices	Re-instantiate
Reorder Canonical to $MoTor$	$Canonical\_to\_MoTor(A)$	R	yes
Reorder $MoTor$ to Canonical	$MoTor\_to\_Canonical(A)$	R	yes
Row permutation	$Row\_permute(A,i,j)$	PERMUT	yes
Column permutation	$Column\_permute(A,i,j)$	PERMUT	yes
Replication	$ElBcast(element)$	LBCAST, RBCAST, D	yes & no
Addition and broadcast	$AddBcast(A)$	ONES	no
Max and broadcast	$MaxBcast(A)$	ONES	no
Min and broadcast	$MinBcast(A)$	ONES	no
Upside-down swap	$UDswap(A)$	SWAP	yes
Left-right swap	$LRswap(A)$	SWAP	yes
Rotation by 180°	$Rotate(A)$	SWAP	yes

multiplication. In Table I we combine proposals presented thus far. There, we present the functionality, the proposed function name, the "special matrices" that have to be instantiated within the *MoTor* system to complete the operations, and information if these matrices have to be reinitialized after μ-cell fusion / splitting operation. Observe that, while the first two functions are directly connected with the *MoTor* architecture, the remaining ones can be seen as "system independent." In other words, they can be implemented for any computer architecture, taking full advantage of the underlying architecture.

This latter observation deserves further attention, and some points have to be made explicit. Only the transformations from the canonical to the $MoTor$ format and back are $MoTor$ architecture specific. The remaining functions are system independent. While the above considerations have in mind the $MoTor$ architecture, the proposed functions use only matrix multiplication and thus can be implemented to run on any computer architecture, using its best of breed matrix multiplication algorithm. This being the case, and taking into account discussion presented in Section I, it may be desirable to implement functions from Table I on existing computers, using state-of-the-art matrix multiplication algorithms and consider their efficiency.

A. Object oriented realization

Let us now recall that our main goal is to consider functions from Table I in the context of the *MoTor* architecture. However, we also see them as a method of simplifying code writing (by introducing matrix operations represented in the style similar to that found in MATLAB / MATHEMATICA). This being the case we assume that there may be multiple ways of implementing these routines, and that they are likely to be vendor / hardware specific. Nevertheless, at the time of writing of this paper, object oriented programming is one of more popular ways of writing codes in scientific computing and image processing. Furthermore, this means the possible trial implementations, suggested above, are likely to be tried using this paradigm. This being the case, we have decided to conceptualize the top-level object-oriented representation of the library of routines from Table I. Since different OO languages have slightly different syntax (and semantics), we use a generic notation, distinguishing information that needs to be made available in the interface and in the main class. We start from the interface (see, also [28]).

```
/* T - type of matrix element */

interface Matrix_interface {
        public Matrix 0(n) {/*generalized zero matrix*/}
        public Matrix I(n) {/*generalized identity matrix*/}
        public Matrix operator + /*generalized  A+B*/
        public Matrix operator * /*generalized A*B*/
        public Matrix Canonical_to_Motor(A);
        /*reordering for themesh-of-tori processing*/
        public Matrix Motor_to_Canonical (A);/*inverse of
        the reordering for the mesh-of-tori processing*/
        public Matrix transpose(A);
        {/*transposition of matrix A*/}
        /*generalized permutation of column / row
                i and j in matrix A*/
        public Matrix Column_Permut (A,i,j);
        public Matrix Row_Permut (A,i,j);
        /*generalized element broadcast*/
        public Matrix ElBcast(element);
        public Matrix AddBcast (A);/*generalized summation
        of all elements of A and broadcast*/
        /*broadcast the largest element of A*/
        public Matrix MaxBcast (A);
        /*broadcast the smallest element of A*/
        public Matrix MinBcast (A);
        /*Matrix (Image) Manipulation*/
        public Matrix UDswap (A);/*upside-down swap*/
        public Matrix LRswap (A);/*left-right swap*/
        /*image vertical rotation*/
        public Matrix Rotate (A);...
}
```

Just defined interface is to be used with the following class *Matrix*. This class summarizes the proposals outlined above.

```
class Matrix inherit scalar_Semiring
                implement Matrix_interface {
  T: type of element;/*double, single,...*/
  private Matrix R (n);/*matrix for MoTor
        transformation*/
  private Matrix ONES (n) /* matrix of ones*/
  private Matrix PERMUT(i,j,n)/*identity matrix
       with interchanged columns i and j*/
  //anti-diagonal matrix of ones
  private Matrix SWAP (n)
       // Methods
  public Matrix 0(n) {/*0 matrix*/}
  public Matrix I(n) {/*identity matrix*/}
  public Matrix transpose(A:Matrix){
                /*MMU-based transposition of A*/}
  public Matrix operator + {A,B:Matrix}
                {return MMU(A,I(n),B,a,b)}
  public Matrix operator * {A,B:Matrix}
                {return MMU(A,B,matrix_0,a,b)}
  public Matrix Column_Permut (A,i,j){
                 return MMU(PERMUT(i,j,n),A,O(n))}
  public Matrix Row_Permut(A,i,j){
                return MMU(A,PERMUT(i,j,n),O(n))}
```

```
    public Matrix Canonical_to_Motor (A){
                M = R * A * transpose(R);
                return M}
    public Matrix UDswap (A);/*upside-down swap*/
                {return MMU(O(n),SWAP,A)}
    public Matrix LRswap (A);/*left-right swap*/
                {return MMU(O(n), A,SWAP)}
          /*image vertical rotation*/
          public Matrix Rotate (A);
                {A=MMU(O(n),SWAP,A);
                 return MMU(O(n),A,SWAP)}
          ...
    private MMU(A,B,C:Matrix(n)){
      return "vendor/implementer specific
          realization of MMU = C + A*B where
      + / * are from class scalar_Semiring"}
...
}
```

Obviously, just defined class and interface allow us to write codes in the suggested manner. Here, the matrix operations (image manipulations) can be performed by calling very simple functions, and hiding all implementation details from the user.

V. Concluding remarks

The aim of this paper was to reflect on current trends in computational sciences. We have focused our attention on selected trends in hardware and software design, to visualize possible designs of future processors and ways they will be combined to build scalable (super)computers. In this context, the mesh-of-tori architecture is one of the more promising concepts for design of large-scale computer systems. This provided us with background, against which we have considered the role of generalized matrix multiplication. As a result we proposed a novel library of routines, based on generalized matrix multiplication that allows for data (matrix / image) manipulations. In the future we plan to implement the proposed library on the virtual *MoTor* system.

Acknowledgment

Work of Marcin Paprzycki was completed while visiting the University of Aizu.

References

[1] K. Boland and A. Dollas, "Predicting and precluding problems with memory latency," *IEEE Micro*, vol. 14, no. 4, pp. 59–67, 1994.

[2] D. Burger, J. R. Goodman, and A. Kagi, "Memory bandwidth limitations of future microprocessors," in *Proceedings of the 23rd Annual International Symposium on Computer Architecture*. New York, NY, USA: ACM, 1996, pp. 78–89, doi:http://doi.acm.org/10.1145/232973.232983.

[3] P. Machanick, "Approaches to addressing the memory wall," http://homes.cs.ru.ac.za/philip/Publications/Techreports/2002/Reports/memory-wall-survey.pdf, 2002.

[4] R. van der Pas, "Memory hierarchy in cache-based systems," http://www.sun.com/blueprints/1102/817-0742.pdf, Sun Microsystems, Tech. Rep., 2002.

[5] Wikipedia, "Flops," http://en.wikipedia.org/wiki/FLOPS.

[6] P. Jacob, A. Zia, O. Erdogan, P. M. Belemjian, J.-W. Kim, M. Chu, R. P. Kraft, J. F. McDonald, and K. Bernstein, "3D memory stacking; mitigating memory wall effects in high-clock-rate and multicore CMOS 3-D processor memory stacks," *Proceedings of the IEEE*, vol. 97, no. 1, January 2009.

[7] F. Alted, "Why modern CPUs are starving and what can be done about it," *Computing in Science and Engineering*, vol. 12, pp. 68–71, 2010, doi:http://doi.ieeecomputersociety.org/10.1109/MCSE.2010.51.

[8] A. Wegener, "Numerical encoding shatters exascale's memory wall," http://www.hpcadvisorycouncil.com/events/2013/Stanford-Workshop/pdf/Presentations/Day2013.

[9] F. G. Gustavson, "Cache blocking for linear algebra algorithms," in *Parallel Processing and Applied Mathematics*, ser. Lecture Notes in Computer Science, R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Waśniewski, Eds. Springer Berlin Heidelberg, 2012, vol. 7203, pp. 122–132.

[10] D. Fey and D. Schmidt, "Marching-pixels: a new organic computing paradigm for smart sensor processor arrays," in *CF '05: Proceedings of the 2nd conference on Computing frontiers*. New York, NY, USA: ACM, 2005, pp. 1–9, doi:http://doi.acm.org/10.1145/1062261.1062264.

[11] "Canon EOS 5D," http://www.usa.canon.com/cusa/consumer/products/cameras/slr_cameras/eos_5d_mark_iii, 2013.

[12] "Nikon D800," http://www.nikonusa.com/en/Nikon-Products/Product/Digital-SLR-Cameras/25480/D800.html, 2013.

[13] "Nokia 808 pureview," http://reviews.cnet.com/smartphones/nokia-808-pureview-unlocked/4505-6452_7-35151907.html, 2013.

[14] E. H. M. Heijne, "Gigasensors for an attoscope: Catching quanta in CMOS," *IEEE Solid State Circuits Newsletter*, vol. 13, no. 4, pp. 28–34, 2008.

[15] S. Chai and D. Wills, "Systolic opportunities for multidimensional data streams," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 388–398, 2002.

[16] S. Kyo, S. Okazaki, and T. Arai, "An integrated memory array processor architecture for embedded image recognition systems," in *Computer Architecture, 2005. ISCA '05. Proceedings. 32nd International Symposium on*, June 2005, pp. 134–145.

[17] Á. Zarándy, *Focal-Plane Sensor-Processor Chips*. Springer, 2011. [Online]. Available: http://books.google.co.jp/books?id=wpCsjwEACAAJ

[18] S. G. Sedukhin and T. Miyazaki, "Rapid*closure: Algebraic extensions of a scalar multiply-add operation," in *CATA*, 2010, pp. 19–24.

[19] "Kalray multi-core processors," http://www.kalray.eu/.

[20] P. Altevogt and A. Linke, "Parallelization of the two-dimensional ising model on a cluster of ibm risc system/6000 workstations," *Parallel Computing*, vol. 19, no. 9, pp. 1041–1052, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167819193900964

[21] S. Sedukhin, T. Miyazaki, K. Kuroda, H. Oi, and Y. Okuyama, "Arithmetic Processing Unit, Patent Application," Filled on September 2007. [Online]. Available: {http://worldwide.espacenet.com/publicationDetails/biblio?adjacent=true&locale=en_EP&FT=D&date=20070927&CC=JP&NR=2007249744A&KC=A}

[22] A. A. Ravankar and S. G. Sedukhin, "Mesh-of-tori: A novel interconnection network for frontal plane cellular processors," *2013 International Conference on Computing, Networking and Communications (ICNC)*, pp. 281–284, 2010.

[23] A. A. Ravankar, "A new "mesh-of-tori" interconnection network and matrix based algorithms," Master's thesis, University of Aizu, September 2011.

[24] A. Ravankar and S. Sedukhin, "Image scrambling based on a new linear transform," in *Multimedia Technology (ICMT), 2011 International Conference on*, 2011, pp. 3105–3108.

[25] A. A. Ravankar and S. G. Sedukhin, "An O(n) time-complexity matrix transpose on torus array processor," in *ICNC*, 2011, pp. 242–247.

[26] J. L. Gustafson, "Algorithm leadership," *HPCwire*, vol. Tabor Communications, April 06, 2007. [Online]. Available: http://www.hpcwire.com/features/17898659.html

[27] "Wikipedia pixel," http://en.wikipedia.org/wiki/Pixel, March 2013.

[28] S. G. Sedukhin and M. Paprzycki, "Generalizing matrix multiplication for efficient computations on modern computers," in *Parallel Processing and Applied Mathematics*, ser. Lecture Notes in Computer Science, R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Waśniewski, Eds. Springer Berlin Heidelberg, 2012, vol. 7203, pp. 225–234.

[29] M. Paprzycki, "Parallel Gaussian elimination algorithms on a Cray Y-MP," *Informatica*, vol. 19, no. 2, pp. 235–240, 1995.

[30] S. G. Sedukhin, A. S. Zekri, and T. Myiazaki, "Orbital algorithms and unified array processor for computing 2D separable transforms," in *Parallel Processing Workshops, International Conference on*. Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 127–134.

[31] M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra, *MPI-The Complete Reference, Volume 1: The MPI Core*, 2nd ed. Cambridge, MA, USA: MIT Press, 1998.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 463–469

Automatic Connections in IEC 61131-3 Function Block Diagrams

Marcin Jamro and Dariusz Rzonca
Rzeszow University of Technology
Department of Computer and Control Engineering
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
Email: {mjamro,drzonca}@kia.prz.edu.pl

***Abstract*—IEC 61131-3 standard defines five languages for programming industrial controllers. They support both textual and graphical development approaches. In case of Function Block Diagram graphical language, diagrams consist of a set of elements connected with lines, which have various length and shape. Development of an editor supporting diagrams design involves implementation of an algorithm, which is able to automatically find a suitable connection between blocks. In the paper an appropriate application of A* algorithm is proposed. The authors have ensured that the proposed solution is efficient and work smoothly. Relations between implementation details and performance are discussed. Achieved results caused that the mechanism has been introduced into graphics editors available in CPDev engineering environment for programming controllers.**

***Index Terms*—A* algorithm, graphics editors, IEC 61131-3, searching path.**

I. Introduction

GRAPHICS editors for various diagrams allow the user to create connections between symbolic blocks. The simplest approach is to draw an exact line or polyline by the user. It can be cumbersome and lead to errors, especially when the block, which is a start or end point for the connection, is moved later, because the connection is not updatable. A better solution is to draw connections between blocks automatically and also update them when necessary, without user attention. This scenario makes creation of the diagram easier and limits a number of errors. It is consistent with a process of diagram creation that starts from designing the first working version of control algorithm, without focusing on legibility of the diagram, and then moving elements to get more readable design that is easier to understand and maintain.

To implement such an approach, a dedicated algorithm is required. It complies with some rules corresponding to the diagram type (e.g. restrictions on overlapping lines and crossing other blocks) and takes into account user preferences (directions changed rarely). Moreover, the algorithm should determine all paths on the fly, immediately after creating or moving an element.

Results obtained during research caused an implementation of the mechanism of automatic connections finding in graphics editors inside the CPDev engineering environment. They work smoothly also on devices with limited resources and allow the designer to create and modify connections in almost imperceptible way. A process of updating connections between multiple elements typically takes only a few milliseconds.

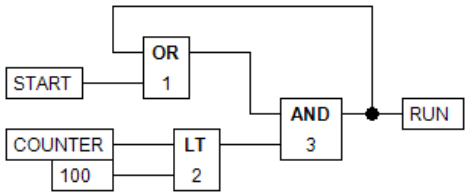


Fig. 1. FBD program that represents the following formula: $RUN = (RUN \vee START) \wedge (COUNTER < 100)$.

In this paper, the algorithm with an application for Function Block Diagram (FBD) graphical language from IEC 61131-3 standard is presented. The article is organized as follows. The second section reviews FBD language and CPDev engineering environment. A problem of path searching in a graph with analysis of A* algorithm modifications is described in the third section. The results of measurements and a short information about test software is presented in the fourth section.

II. FBD in CPDev engineering environment

Third part of the IEC 61131 standard [1] defines five programming languages for industrial controllers. Textual languages include ST (Structured Text) and IL (Instruction List). FBD (Function Block Diagram) and LD (Ladder Diagram) are graphical, while SFC (Sequential Function Chart) is mixed and requires parts implemented in other languages.

In the paper, the authors focus on FBD, that is a graphical language allowing users to create control programs in a visual way. POUs (Program Organization Units, i.e. programs, function blocks, and functions) defined in this language consist of rectangles that represent variables, constants, instances of function blocks, and functions. All of them are connected with lines as shown in Fig. 1.

Graphical languages benefit from visual programming. Their features include legibility of diagrams, easiness of program understanding or modification, and a possibility of attaching printouts directly to the documentation. Engineering tools (e.g. Beckhoff TwinCAT [2], CoDeSys [3], Control Builder F [4]) contain graphics editors that allow users to design programs graphically. They support drawing and

978-83-60810-53-8/$25.00

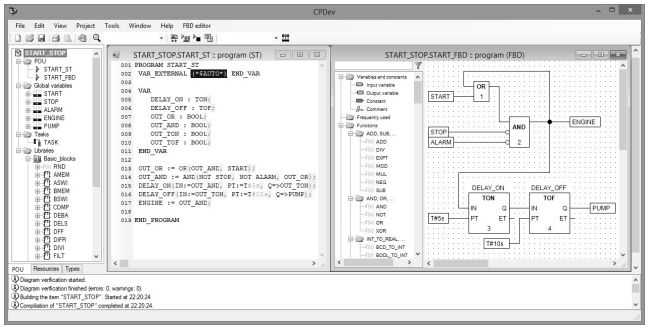


Fig. 2. CPDev IDE with editors of ST and FBD languages.

updating connections between blocks automatically, using some proprietary solutions, without any clues of the used algorithms and implementation details. Development of the CPDev engineering environment required another solution, thus the authors proposed using the A* algorithm with some adjustments and tuning of parameters, as described later.

CPDev (Control Program Developer) is an engineering environment [5] developed in the Department of Computer and Control Engineering at Rzeszow University of Technology (Poland). It can be used for programming PLC, and PAC controllers, mini-DCSs [6], and secure NCS systems [7], according to IEC 61131-3 standard [1]. The CPDev environment is universal and generates code in the form that can be executed on various target platforms, including AVR, ARM, x86, and FPGA. It is achieved by using a virtual machine executing an intermediate code [8]. The environment is open for controller constructors and engineers that can implement low-level procedures and add them to the virtual machine. Developers using CPDev can create own libraries with POUs and reuse them in multiple projects. CPDev consists of several parts, including integrated development environment (Fig. 2), compilers, translators, configuration tools, testing application [9], and visualization mechanism [10]. CPDev has been applied for ship control and monitoring systems from Praxis Automation Technology B.V. (Leiderdorp, the Netherlands) [11] and for small PAC controller in measurement-and-control systems from LUMEL S.A. (Zielona Gora, Poland) [12].

All kinds of POUs, i.e. programs, function blocks, and functions, can be created using CPDev graphics editors [13]. They are equipped with typical functionalities such as basic edition operations (adding, moving, copying, pasting), translation to ST code, conversion to XML format, and printing accordingly to a template. The editors provide also an execution mode to run programs with support of tracing variable values and breakpoints.

Automatic connection finding is one of the most important features. It generates a connection between two elements on the diagram (variables, functions, or function blocks) automatically. Therefore, the user can focus on implementation of the control software, without paying special attention to connections. In CPDev graphics editors the lines are drawn automatically, just after selecting the beginning and the end of the connection. The problem can be interpreted as finding a path in a graph. However, some simplifications and modifications have to be done due to short time requirement.

III. Path searching in a graph

The problem of finding the shortest path in a graph is one of typical problems in discrete mathematics [14]. The classical approach, like the Dijkstra algorithm [15], focuses on examining all possible paths to find the shortest one. However, such a solution is not performance efficient. As an extension to the classical approach, a number of BFS (*Best-First Search*) algorithms have been proposed. For the problem specified above, the authors have chosen the A* algorithm [16], which combines traditional approach (similar to Dijkstra's) and heuristic one involving a metric. The A* algorithm requires estimation of a distance between the current node and the target for every node in a graph. Such a distance must be predicted in an optimistic way, i.e. real distance cannot be shorter than estimated. Numerous modifications of A* algorithm have been described [17], as well as applications including finding optimal routing in wireless sensor networks [18], [19], shortest road on a map [20], or even solving motion correspondence problem in computer vision [21].

Basically, the A* algorithm in every step tries to go the most promising way, i.e. chooses such a neighbor node that is close both to the current track and to the target node. At first, estimated distance from the target is calculated for every node, denoted as *H score*. While traversing the graph, real distances from the start node to neighbors of the current one are evaluated. The *G score* for every node reflects the distance from the start point via the shortest path already examined. The G score for every node can be updated later if a shorter path to this node is found. The *F score* is a sum of G and H scores. It represents estimated cost of using this node while directing to the target. In every step a node with the lowest F score from so-called *open set* (containing nodes to be traversed) is selected, and added to the *closed set* (nodes already traversed).

Implementation of such an algorithm requires consideration of some data structures for these sets. Ordinary lists or hashtables can be used, but heaps or binary search trees (simple BST or self-balancing, e.g. red-black trees RBT) [22] usually turn out better. Selection of a node with the minimal F score from the open set (which is a priority queue) is one of the most time consuming parts in A* algorithm. The worst case computational complexity of such an operation is $O(n)$ for tables, $O(\log n)$ for balanced BST or RBT, and $O(1)$ for min heap. Choosing F score as a key for a complex data structure improves performance, but checking if the set already contains the node, by using structure with coordinates of the node as a key, is more convenient. Similarly, adding or removing a node from complex data structures, as well as updating their F score (and position in the structure) takes significantly more time. Thus, choosing the most efficient structure, as well as its key, in a given case requires some tests, as shown in the following section.

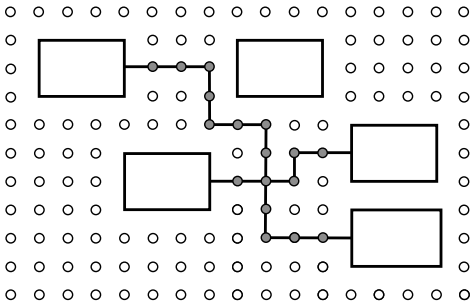


Fig. 3. Generated graph nodes.

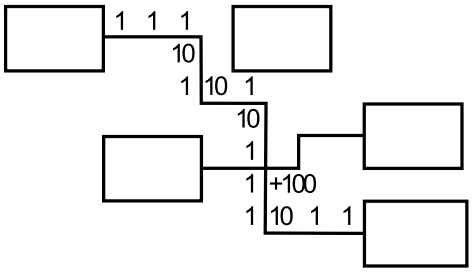


Fig. 4. Calculation of the path cost.

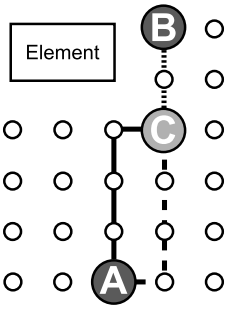


Fig. 5. Variants of the path between A and B nodes.

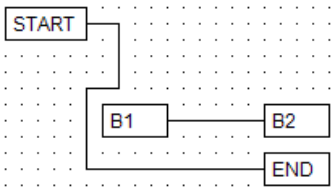


Fig. 6. Connections found with line cost intersection 50.

Another possibility for closed set implementation is addition of marks indicating whether the node has been already traversed. Such marks are included in internal class representation of the nodes. Thus, an additional data structure is unnecessary, which simplifies the implementation.

Finding an appropriate connection between elements in FBD diagrams can be considered as solving the shortest path problem in a graph. The connection should meet the following requirements:

- be found every time if elements are placed on the diagram correctly,
- pass round elements placed earlier,
- limit a number of intersections with other lines,
- change direction rarely,
- support additional margin around elements.

In our solution such a graph is created automatically. The nodes are simply diagram grid points, as shown in Fig. 3. Elements placed on the diagram (variables, functions, and function blocks) remove some nodes from the graph. Arcs connect the nodes vertically and horizontally, according to the neighborhood of the grid points. Initially weights of all arcs are equal. The start and end nodes (in the graph) are defined by positions of the elements, which should be connected.

A structure of the graph, which nodes represent grid points, is suitable for searching the shortest path using A* algorithm. The Manhattan (taxicab) distance is a natural choice for heuristic function to estimate a connection cost in such a case. The distance is calculated according to the formula $d(n_1, n_2) = |x_1 - x_2| + |y_1 - y_2|$, where $n_i = (x_i, y_i)$ denotes node i with coordinates x_i and y_i. Among other metrics, that could be also considered, the maximum (Chebyshev) one seems also promising, calculated as $d(n_1, n_2) = \max(|x_1 - x_2|, |y_1 - y_2|)$.

In order to find an optimal connection some additional conditions must also be considered while analyzing the diagram. Firstly, connection lines can not overlap, and can cross only when necessary. Overlapping can be easily solved by removing the arcs in the graph, which connect nodes representing fields that are already parts of any line. Crossing can be reduced by adding a high penalty for the G score. Secondly, lines in the diagram should go straight and change direction rarely. It can be achieved by setting an appropriate penalty on the G score, depending on a direction of actual path. An example of path cost calculation according to these rules is shown in Fig. 4.

It is worth mentioning, that G score rules sometimes lead to calculating different costs for a path between two nodes, depending on the previous path. Such a case is shown in Fig. 5. Considering the paths between A and B nodes, the algorithm can choose either one marked by a solid line or by dashed. In both cases the connection crosses C node. Two paths between A and C have the same cost, but the cost of the path between C and B is different, depending on the previous path. If the solid line between A and C is chosen, the connection changes direction in C towards B, thus some penalty for the G score is added.

A method for adding nodes to the structure representing the open set is another issue. In every step one node with the lowest F score in the open set is selected, but sometimes there may be many nodes with the same minimal F score. Selection of the node in such a case depends on the order of adding nodes in previous steps. There are several possibilities to be considered. New nodes can be added before or after previous ones, and they can be unsorted or sorted by direction. Such modifications affect searching time and shape of resulting path.

The costs of line intersection and direction change have

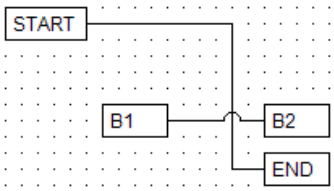


Fig. 7. Connections found with line cost intersection 5.

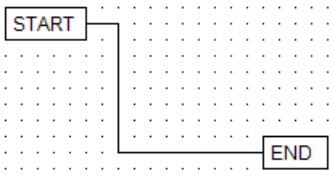


Fig. 8. Connection found with Manhattan metric.

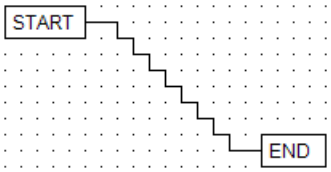


Fig. 9. Connection found with Maximum metric.

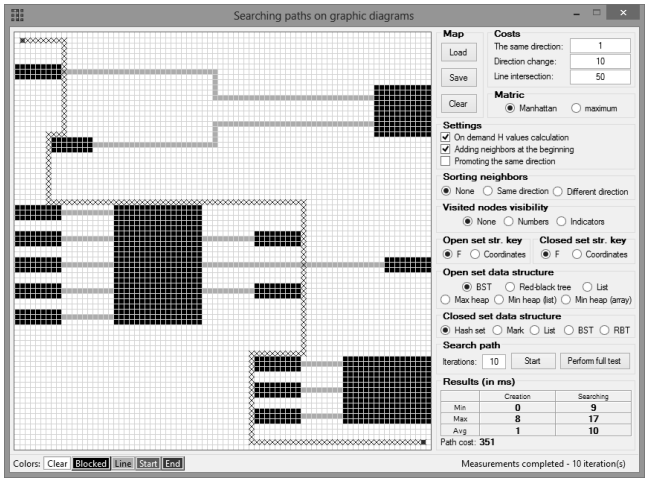


Fig. 10. The application for testing the mechanism of connection finding.

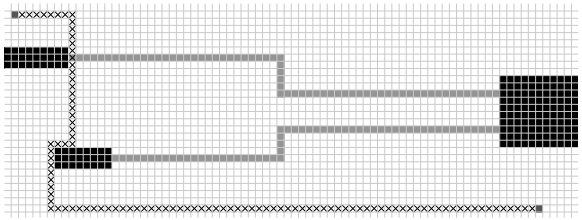


Fig. 11. The simple map for testing the mechanism of connection finding.

also an impact on the final path (Fig. 6, 7). If penalty for crossing a line is huge (as in Fig. 6), intersections will be avoided, but the path will go by roundabout way. The metric influences a shape of the path as well (Fig. 8, 9). For maximum metric the algorithm tries to minimize the distance in one of coordinates, the longest first. Such a metric can lead to generating ”stairs shaped” connection (Fig. 9) when a penalty for changing direction is low. Selection of appropriate cost values to find reasonable compromise in general case is not trivial.

IV. Test software and performance results

A. Test software

Test results have been measured using a dedicated software with user interface shown in Fig. 10. That makes it possible to adjust settings and perform measurements for various combinations of parameter values. All tests are run on the map implemented as a matrix of integer values. They represent current states of fields: *start*, *end*, *blocked*, *line*, or *clear*. Available settings include:

- cost values (e.g. direction change or line crossing)
- open set data structure (BST, RBT, list, max heap, and min heap)
- closed set data structure (hash set, mark, list, BST, and RBT)
- keys for open and closed set structure (F value or coordinates)

The testing application consists of two parts, i.e. map and settings panel. In the example from Fig. 10, the map presents a connection found by the mechanism between the start element (top-left) and the end (bottom-right). Some other elements, like blocks and variables from diagrams, are shown as black rectangles. They are connected by gray lines.

The measurements have been performed for a number of data sets, including simple and complex (Fig. 11, 12–15, respectively). The results have been calculated for various metrics, costs, structures, and keys for open and closed sets.

B. Simple map

The simple example from Fig. 11 consists of three blocks. Two of them represent input variables and the third one is an instance of function block. There are two lines from inputs to the block. The algorithm has to find a path between the start element (top-left) and the end (bottom-right) one. It is clear that for specified values of parameters the algorithm tries to avoid crossing lines, even if it requires to change direction and apply a longer path.

Improving performance is one of the most important reasons for testing the mechanism of finding connections. The results are presented in Table I. Following abbreviations have been assumed: *F* indicates F value, *C* – coordinates, *HS* – hash set structure, *Heap* – min heap structure, *Man* – Manhattan metric, and *Max* – Maximum metric. Cost is equal to 191 in all cases. Tests have been performed on PC with 2,5 GHz processor. Even for the simple example (Fig. 11) required times are different and depend on parameters, mainly on open and closed

TABLE I
RESULTS OF TESTING THE MECHANISM OF FINDING CONNECTIONS FOR THE SIMPLE MAP.

#	Open set		Closed set		Metr.	Required time [ms]
	Str.	Key	Str.	Key		
1	RBT	F	Mark	-	Man	0
2	BST	F	Mark	-	Man	1
3	BST	F	Mark	-	Max	1
4	RBT	F	Mark	-	Max	1
5	BST	F	HS	any	Man	1
6	RBT	F	HS	any	Man	1
7	Heap	F	Mark	-	Man	1
8	BST	F	BST	C	Man	3
9	List	F	Mark	-	Max	4
10	BST	F	RBT	F	Man	20
11	Heap	F	HS	F	Max	45
12	RBT	F	List	F	Max	64
13	List	F	List	C	Max	74
14	BST	F	BST	F	Man	118
15	List	C	BST	F	Man	125
16	Heap	C	BST	F	Max	174

set structures, as well as their keys. The metrics do not have such an important impact. The mechanism allows to calculate the path in the fastest way by using RBT or BST structures of the open set, and mark or hash set as a structure of the closed set (Table I, rows 1-6). The performance is significantly decreased by using a list or min heap as a structure of the open set, and BST, RBT, or list as a structure of the closed set (Table I, rows 15-16).

As mentioned in Section III, the results can be explained by internal concepts of various data structures and their computational complexity. In case of the open set it is important to select data structure that performs well while adding or removing an item, checking whether the structure contains specified item, or selecting an element with minimum value. For the closed set, only two operations should be well supported, i.e. adding an item and checking whether the structure contains specified item. Choosing a suitable combinations of structures for the open and closed sets significantly increases overall performance of the mechanism.

C. Complex map

The complex map (Fig. 12–15) consists of sixteen elements from the FBD diagram, i.e. ten input variables, three output variables, and three instances of function blocks. All of them are connected in more complicated way than before. The performance results for finding connections are presented in Table II. Cost is equal 351 in all cases.

The results confirm conclusions from the previous example. Again, the mechanism performs well with RBT or BST as a structure of the open set (Fig. 16), and with mark or hash set as a structure of the closed set (Fig. 17). However, differences between required times are higher, because significantly more operations must be performed while searching. In this case, the best combination of parameters finds the connection in 1 ms, but the worst in 781 ms.

TABLE II
RESULTS OF TESTING THE MECHANISM OF FINDING CONNECTIONS FOR THE COMPLEX MAP.

#	Open set		Closed set		Metric	Required time [ms]
	Str.	Key	Str.	Key		
1	RBT	F	Mark	-	Man	1
2	BST	F	Mark	-	Man	2
3	BST	F	HS	F	Man	3
4	RBT	F	HS	F	Man	3
5	Heap	F	Mark	-	Man	3
6	BST	F	RBT	C	Man	5
7	Heap	F	RBT	F	Max	103
8	Heap	F	BST	F	Max	781

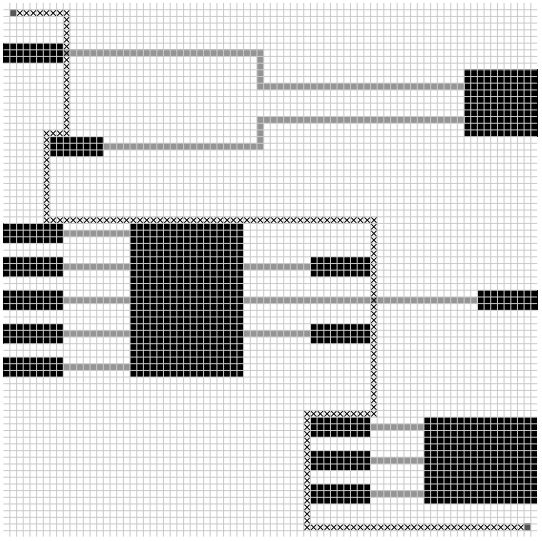


Fig. 12. Connections found for the Manhattan metric and 50 as a cost of line intersection.

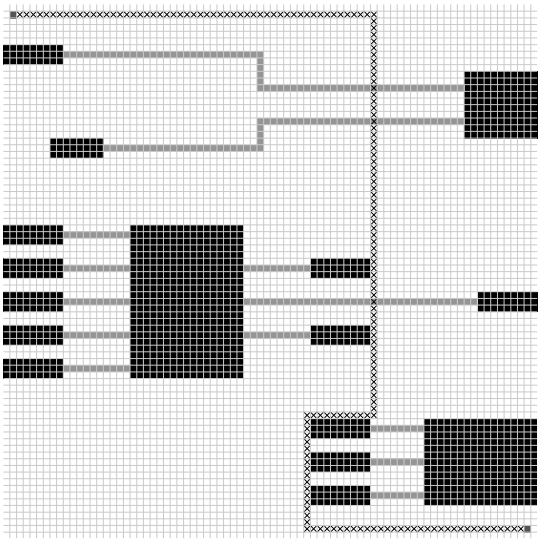


Fig. 13. Connections found for the Manhattan metric and 20 as a cost of line intersection.

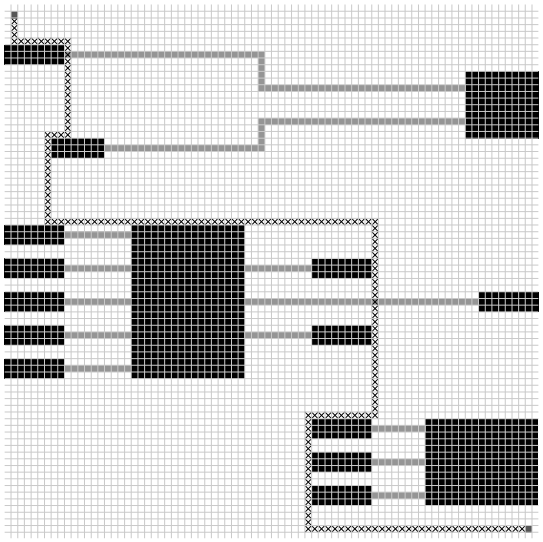


Fig. 14. Connections found for the Maximum metric and 50 as a cost of line intersection.

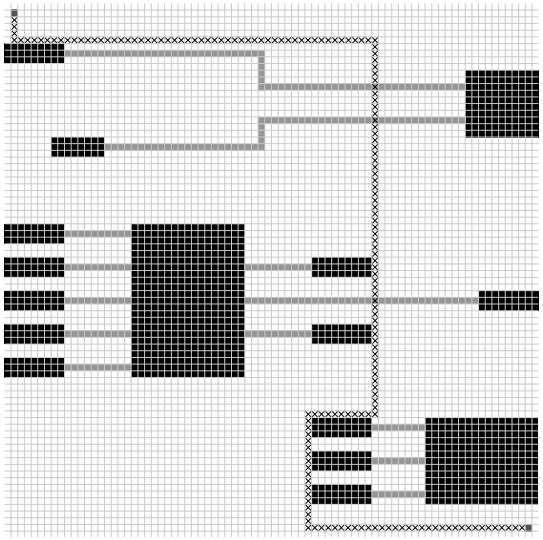


Fig. 15. Connections found for the Maximum metric and 20 as a cost of line intersection.

Using a proper open set structure (for fixed closed set structure) improves performance even a few times (Fig. 16). The difference is also seen in case of mark and hash set as structures of the closed set (Fig. 17). The mark can lead up to 50% increase of performance. Choosing a key for the closed set structure also affects performance, but only in case of structures other than mark.

A comparison between four combinations of structures of open and closed sets (RBT or BST, mark or hash set), depending on the metric, is shown in Fig. 18. The difference in case of the complex map is not high between BST and RBT used as the open set. Performance for mark and hash set as structures of the closed set is also similar. There is a difference in average time required to find connection when

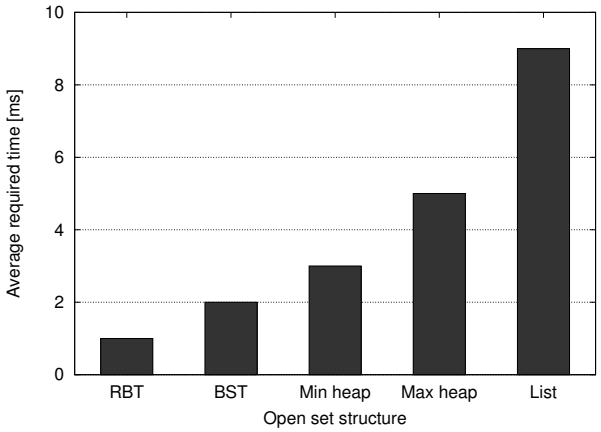


Fig. 16. Average required time for finding connection by open set structure (mark, F value as closed set structure and key, Manhattan metric).

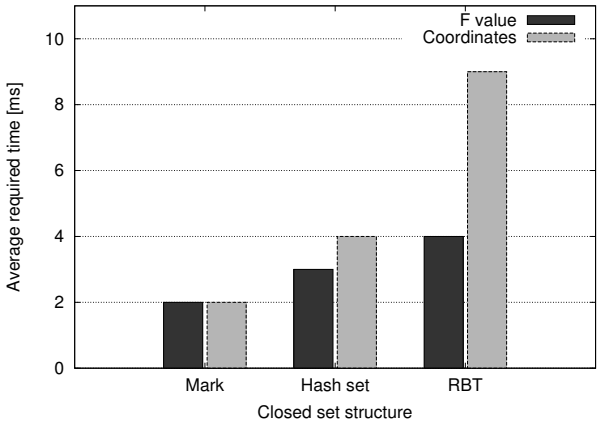


Fig. 17. Average required time for finding connection by closed set structure, open and closed set keys (BST as open set structure, Manhattan metric).

the mechanism uses Manhattan and maximum metrics. It is caused mainly by a different number of fields analyzed during searching. The difference is specific to the map.

Parameters have an impact not only on performance, but also on the path found. Depending on the metric and costs, it can change a direction more or less frequently, promote variants without crossing other lines, or even promote a specific direction. Differences are presented in Figs. 12–15 on a set of maps for the complex example.

The first connection (Fig. 12) is found for the Manhattan metric and 50 as a cost of crossing other lines. In this case the algorithm avoids crossings even by changing directions more frequently (see the top left part). In the second connection (Fig. 13) the cost of crossing is smaller and equals to only 20. It promotes crossing other lines instead of changing direction. Because of it, the first half of the connection differs significantly from the previous case. Similar conclusions can be made for the third and fourth connections (Figs. 14, 15), however, shape of the connection is different. The change

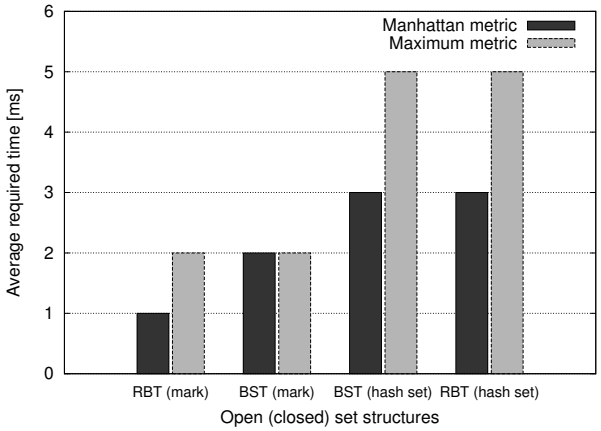


Fig. 18. Average required time for finding connection, depending on metric, open and closed set structures, with F values as keys.

is caused by the metric, either Manhattan (Figs. 12, 13) or maximum (Figs. 14, 15). The authors empirically found that values about 10-20 for direction change and 20-50 for line crossing give satisfactory results in case of FBD and LD diagrams.

V. Conclusion

The problem of finding a suitable path in a graph is typical in discrete mathematics. However, its application in FBD graphic diagrams to automatically find connections between elements requires some additional research and discussion. The mechanism meets a set of requirements which contain an execution in short time allowing to use on devices with limited resources. It supports a process of almost imperceptible updating and redrawing connections after moving any block on the diagram. Thus, the developer can easily adjust design of the diagram after creation of the first working version of POU. It can significantly increase legibility of the diagram and makes it easier to understand, modify, and maintain. Taken assumptions caused some modifications and tuning of the A* algorithm. As measured, values of parameters have significant impact on performance, length, and shape of the connection.

In the paper the authors considered some implementation details of A* algorithm, created a dedicated software for performance testing, and presented conclusions. The measurements indicate data structures that are efficient for the A* algorithm, and group of structures with performance problems. An impact on the shape of connection depending on path costs and metric is also described. All of these aspects are combined to tune properly the mechanism of finding connections in IEC 61131-3 Function Block Diagram editor implemented in the CPDev engineering environment.

References

[1] "IEC 61131-3 - Programmable controllers - Part 3: Programming languages," 2003.
[2] "Beckhoff TwinCAT website," 2013, http://www.beckhoff.com/english/twincat.
[3] "CoDeSys website," 2013, http://www.codesys.com.
[4] "Control Builder F website," 2013, http://www.abb.com/product/seitp334/ee37d357581192 adc12571ca00431c6e.aspx.
[5] "CPDev website," 2013, http://cpdev.kia.prz.edu.pl.
[6] D. Rzonca, A. Stec, and B. Trybus, "Data Acquisition Server for Mini Distributed Control System," in *Computer Networks*, ser. Communications in Computer and Information Science, A. Kwiecien, P. Gaj, and P. Stera, Eds. Springer Berlin Heidelberg, 2011, vol. 160, pp. 398–406.
[7] W. Rzasa, D. Rzonca, A. Stec, and B. Trybus, "Analysis of Challenge-Response Authentication in a Networked Control System," in *Computer Networks*, ser. Communications in Computer and Information Science, A. Kwiecien, P. Gaj, and P. Stera, Eds. Springer Berlin Heidelberg, 2012, vol. 291, pp. 271–279.
[8] D. Rzonca and B. Trybus, "Hierarchical Petri Net for the CPDev Virtual Machine with Communications," in *Computer Networks*, ser. Communications in Computer and Information Science, A. Kwiecien, P. Gaj, and P. Stera, Eds. Springer Berlin Heidelberg, 2009, vol. 39, pp. 264–271.
[9] M. Jamro, D. Rzonca, and B. Trybus, "Communication Performance Tests in Distributed Control Systems," in *Computer Networks*, ser. Communications in Computer and Information Science, A. Kwiecien, P. Gaj, and P. Stera, Eds. Springer Berlin Heidelberg, 2013, vol. 370, pp. 200–209.
[10] M. Jamro and B. Trybus, "IEC 61131-3 Programmable Human Machine Interfaces for Control Devices," in *Human System Interactions (HSI), 2013 6th International Conference on*, 2013, pp. 48–55.
[11] "Praxis Automation Technology B.V. website," 2013, http://www.praxis-automation.nl.
[12] "Lumel S.A. website," 2013, http://www.lumel.com.pl/en/.
[13] M. Jamro, "Graphics editors in CPDev environment," *Journal of Theoretical and Applied Computer Science*, vol. 6, no. 1, pp. 13–24, 2012.
[14] P. Festa, "Shortest Path Algorithms," in *Handbook of Optimization in Telecommunications*, M. G. Resende and P. M. Pardalos, Eds. Springer US, 2006, pp. 185–210.
[15] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
[16] P. Hart, N. Nilsson, and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," *Systems Science and Cybernetics, IEEE Transactions on*, vol. 4, no. 2, pp. 100 –107, July 1968.
[17] A. V. Goldberg, "Point-to-Point Shortest Path Algorithms with Preprocessing," in *SOFSEM 2007: Theory and Practice of Computer Science*, ser. Lecture Notes in Computer Science, J. Leeuwen, G. Italiano, W. Hoek, C. Meinel, H. Sack, and F. Plasil, Eds. Springer Berlin Heidelberg, 2007, vol. 4362, pp. 88–102.
[18] I. S. AlShawi, L. Yan, W. Pan, and B. Luo, "Lifetime Enhancement in Wireless Sensor Networks Using Fuzzy Approach and A-Star Algorithm," *Sensors Journal, IEEE*, vol. 12, no. 10, pp. 3010–3018, Oct. 2012.
[19] K. Rana and M. Zaveri, "A-Star Algorithm for Energy Efficient Routing in Wireless Sensor Network," in *Trends in Network and Communications*, ser. Communications in Computer and Information Science, D. Wyld, M. Wozniak, N. Chaki, N. Meghanathan, and D. Nagamalai, Eds. Springer Berlin Heidelberg, 2011, vol. 197, pp. 232–241.
[20] F. Hahne, C. Nowak, and K. Ambrosi, "Acceleration of the A*-Algorithm for the Shortest Path Problem in Digital Road Maps," in *Operations Research Proceedings 2007*, ser. Operations Research Proceedings, J. Kalcsics and S. Nickel, Eds. Springer Berlin Heidelberg, 2008, vol. 2007, pp. 455–460.
[21] K.-Y. Eom, J.-Y. Jung, and M.-H. Kim, "A heuristic search-based motion correspondence algorithm using fuzzy clustering," *International Journal of Control, Automation and Systems*, vol. 10, pp. 594–602, 2012.
[22] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*, 3rd ed. The MIT Press, 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 471–478

N-body simulation based on the Particle Mesh method using Multigrid schemes

P. E. Kyziropoulos
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Kimmeria, GR 67100 Xanthi, Greece
panakyzi@ee.duth.gr

C. K. Filelis-Papadopoulos
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Kimmeria, GR 67100 Xanthi, Greece
chripapa9@ee.duth.gr

G. A. Gravvanis
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Kimmeria, GR 67100 Xanthi, Greece
ggravvan@ee.duth.gr

***Abstract*—Through the last decades multigrid methods have been used extensively in the solution of large sparse linear systems derived from the discretization of Partial Differential Equations in two or three space variables, subject to a variety of boundary conditions. Due to their efficiency and convergence behavior, multigrid methods are used in many scientific fields as solvers or preconditioners. Herewith, we propose a new algorithm for N-body simulation, based on the V-Cycle multigrid method in conjunction with Generic Approximate SParse Inverses (GenAspI). The N-body problem chosen is in toroidal 3D space and the bodies are subject only to gravitational forces. In each time step, a large sparse linear system is solved to compute the gravity potential at each nodal point in order to interpolate the solution to each body and through the velocity Verlet method compute the new position, velocity and acceleration of each respective body. Moreover, a parallel version of the multigrid algorithm with a truncated approach in the parallel levels is utilized for the fast solution of the linear system. Furthermore parallel results are provided which depict the efficiency and performance for the proposed multigrid N-body scheme.**

I. Introduction

LET US consider the Partial Differential Equation (PDE) describing the gravity potential in a domain Ω, [19]:

$$\Delta\Phi = 4\pi G\rho, (x, y, z) \in \Omega \quad (1)$$

subject to Dirichlet boundary conditions

$$\Phi = 0, (x, y, z) \in \vartheta\Omega \quad (1.a)$$

where Ω denotes the region where the problem resides, $\partial\Omega$ is the boundary of the region, G is the Gravitational constant and ρ is the mass density in each nodal point computed by the mass of the neighboring bodies.

Applying the Finite Difference method, with the seven point stencil, for the PDE (1)-(1.a) results in solving a seven diagonal linear system,

$$A\varphi = f, \quad (2)$$

where A is a large sparse diagonally dominant symmetric matrix, f is the right hand side vector consisting of the forcing term and respective boundary conditions and φ is the gravity potential at each nodal point.

The linear system (2) derived from the discretization of the 3D PDE can be solved by the multigrid method. Multigrid methods have been used extensively, during the last decades, in a variety of scientific fields such as Computational Fluid Dynamics, Computational Economics and Partial Differential Equations, due to their near optimal computational complexity and convergence behavior, [3, 4, 5, 10, 15, 16, 17, 18, 21, 23, 24]. Multigrid methods are based on the observation that the low-frequency components of the error are not effectively damped by a stationary iterative method. However, the high frequency components of the error are quickly reduced towards zero within the first few iterations, [3, 4, 21]. In order to handle the low frequency components of the error, multigrid methods utilize a grid hierarchy consisting of coarser levels with higher mesh size (h) and by projecting the finer problem to the coarser levels the lower frequency components are becoming more oscillatory and can be damped efficiently by a stationary iterative solver. The vectors required in each level are transferred between the respective grid with the use of two operators: the prolongation and the restriction operator, which transfer vectors from coarser to finer grids and vice versa, [3, 4, 21]. The sequence in which the coarser grids are visited and the respective coarse grid corrections to the solution are obtained is referred to as the cycle strategy, [3, 4, 21]. A 3D Geometric multigrid hierarchy is depicted in Figure 1. In order to accelerate the convergence of the multigrid method the Dynamic Over / Under Relaxation (DOUR) scheme is used, [18], in conjunction with the GENeric Approximate SParse Inverse (GenAspI) matrix, as the smoother for the multigrid method, [8].

Approximate inverses have been used as preconditioners for various iterative schemes due to their inherent parallelism and convergence behavior, [5, 11, 12, 13, 20]. Moreover, approximate inverses have been used effectively in conjunction with the multigrid method for a variety of problems, [5, 9, 10, 14]. Recently, classes of Generic Approximate Inverses have been proposed, [8,14], that can handle any sparsity pattern of the coefficient matrix A, based on Incomplete LU factorization with zero fill-in. Unlike their predecessors, [11, 12, 13, 20], these Generic schemes are not limited by the structure of the coefficient matrix or the method used for the discretization and produce approximate inverses with sparsity patterns, based on Powers of Sparsi-

978-83-60810-53-8/$25.00

fied Patterns (PSM's), [6, 7], using adequate dropping strategies to further sparsify the initial sparsity pattern of the coefficient matrix A, thus leading to sparser more efficient approximate inverses. The GenAspI matrices are then computed using a modified procedure introduced in the GENeric Approximate Banded Inverses class (GenAbI), [14], according to an a priori known sparsity pattern. The GenAspI matrices are used as preconditioners in the damped Richardson scheme, in conjunction with the DOUR scheme, and V-cycle strategy to derive the GenAspI-MGV method which is used to obtain the gravity potential in each time step of the N-body simulation scheme proposed. The parallelization of the GenAspI-MGV method is performed with a new approach where the lower order levels are executed sequentially in order to avoid overhead in levels with low computational cost.

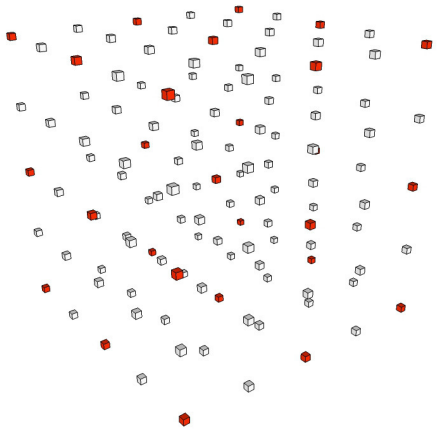


Fig. 1 Grid hierarchy of a three-dimensional PDE discretized with the Finite Differences method for mesh size h=1/2 (red points) and h=1/4 (red and white points).

The N-body simulation is a simulation of a dynamical system of particles, under the influence of, mainly the gravitational force. N-body simulations have become a fundamental tool in the study of complex physical systems, [19]. Starting from a basic physical interaction (e.g., gravitational, Coulomb) one can follow the dynamical evolution of a system of bodies, which represent the phase-space density distribution of the system. The greater the number of the particles used on the simulation results in more accurate and complete results, [19]. There are a lot of methods that can be used to calculate such kind of forces in a confined space. A direct approach to the problem called particle-particle simulation (P-P) assumes that on a simulation containing N bodies, there are N×(N-1) force pairs (Newtonian gravitation), [19]. This direct approach to the problem scales with $O(N^2)$ complexity, where N denotes the number of bodies, rendering the scheme restrictive for large values of particles.

Herewith, we propose a new scheme for computing the gravity potential and thus the forces in the bodies, using the Particle Mesh method and the GenAspI-MGV method to accelerate the solution of the linear system. The Particle Mesh method neglects the close interactions between particles and takes into account only the dynamics of superparticles consisting of a great number of particles. The density of these masses is then added according to weights computed from the distance of each node to the grid points surrounding each respective particle resulting in the rhs vector of the linear system (2). The assignment of the respective weights in each nodal point is performed using the Cloud in Cell (CIC) method, [19]. The linear system is solved using an iterative method and the potential is computed at the cell centers. Finally, by integrating the equations of motion through the Verlet integrator, [22], the new position of each particle is computed. Repeating the aforementioned process leads to the simulation of the system of particles. The Particle Mesh method is efficient due to the fact that the nodal points are, in general, less than the number of particles and thus a large number of pairing forces is not computed. Moreover, the low computational complexity of the mutligrid method utilized in each time step for the solution of the linear system renders the algorithm efficient especially for increasing numbers of particles.

The proposed N-body algorithm is parallelized using OpenMP. OpenMP is a collection of directives available for a variety of languages including C/C++/Fortran used to parallelize programs for Symmetric Multi-Processing Units.

Finally, numerical results on the performance and convergence behavior of the proposed GenAspI-MGV schemes are given for solving three-dimensional N-body simulation problems.

II. Multigrid Method In Conjunction With Generic Approximate Sparse Inverses

The multigrid method, especially in the last decades, is used extensively by the scientific community in the fields of computational physics, computational fluid dynamics and financial engineering due to its near-optimal complexity and its convergence behavior, [3, 4, 5, 9, 14, 15, 16, 17, 21, 23, 24].

Multigrid methods are composed by four distinct components: Smoothers, Prolongation and Restriction operators and Cycle strategy. The smoothers are an essential component of the multigrid algorithm and are used to obtain the respective corrections for the solution on each level. Smoothers are stationary iterative methods that can be described by the following relation, [4,5,21,23],

$$x_l^{(k+1)} = x_l^{(k)} + \omega M_l\left(b_l - A_l x_l^{(k)}\right),\ k = 0,1,2,.... \quad (3)$$

where l denotes the level in which the smoother is applied, M_l is the preconditioner to the Richardson's iterative method and ω is the damping parameter with values between 0 and 2. In case where the $M_l=D_l^{-1}$, the iterative scheme (3) results in the Damped Jacobi method, [4, 21, 23]. Substituting $M_l = \left(M_{lfill}^{drptol}\right)_l$, where $\left(M_{lfill}^{drptol}\right)_l$ is the l-th level Generic Approximate Sparse Inverse with lfill levels of fill and drptol drop tolerance results in the proposed parametric smoothing scheme. The Generic Approximate Sparse Inverse

(GenAspI) is inherently parallel and can be adjusted by modifying the drptol and lfill parameters.

A. Generic Approximate Sparse Inverse smoothing

Let us consider the ILU(0), [1], factorization of a matrix A,

$$\mathrm{A=LU+E} \tag{4}$$

where E is the error matrix and L and U are the lower and upper factors of the matrix A, retaining the same profile. In order to compute the GenAspI matrix a sparsity pattern must be known a priori. The approximate inverse sparsity pattern is computed through Powers of Sparsified Matrices (PSMs) of the filtered version of the coefficient matrix, [6,7]. Let the sparsified version of the matrix A be described as follows,

$$\tilde{A}_{\mathrm{ij}}=\begin{cases}1,\mathrm{i==j\ and}\ \left|\left(D^{-1/2}\mathrm{AD}^{-1/2}\right)_{\mathrm{i,j}}\right|>\mathrm{drptol}\\ 0,\mathrm{otherwise}\end{cases} \tag{5}$$

where D is a diagonal matrix such that

$$D_{\mathrm{ii}}=\begin{cases}|A_{\mathrm{ii}}|,|A_{\mathrm{ii}}|>0\\ 1,\mathrm{otherwise}\end{cases} \tag{6}$$

and drptol is the so called drop tolerance, [6,7]. The coefficient matrix A is sparsified with the process described by equation (5), which denotes the normalization of each element with the diagonal element, [6,7]. Then each element is compared in absolute value against a given drop tolerance in order to withhold or discard the element. The sparsification process is succeeded by the augmentation of the sparsity pattern by raising it to powers denoted by lfill, [6,7]. The remaining nonzero elements of the coefficient matrix represent the strong connections (neighbors) of each element, by considering the equivalent graph of the matrix, [6,7]. The k–th row of the ℓ –th level sparsity pattern $\tilde{\mathrm{A}}^{\ell}$ can be computed as $\tilde{\mathrm{A}}^{\ell}_{\mathrm{k,:}}=\tilde{\mathrm{A}}_{\mathrm{k,:}}\tilde{\mathrm{A}}^{\ell\text{-}1}$, which denotes that the k–th row of the $\tilde{\mathrm{A}}^{\ell}$ pattern is composed by "fusing" the non–zero indices of columns in the rows of $\tilde{\mathrm{A}}^{\ell-1}$ corresponding to the nonzero columns of the k–th row of the sparsified coefficient matrix $\tilde{A}$, [6,7]. By increasing the levels of fill the approximate inverse tends to the exact inverse of the linear system. The algorithm for the computation of an approximate inverse sparsity pattern is given in [6,7,8]. More information concerning the approximate inverse sparsity patterns can be found in [6,7].

The GenAspI matrix is computed, according to the sparsity pattern defined by the previous procedure, by solving recursively the following system, [8],

$$UM^{lfill}_{drptol}=I\ \ \text{and}\ \ M^{lfill}_{drptol}L=0 \tag{7}$$

where L and U are the lower and upper triangular factors computed by the ILU(0), [1], factorization, (4). The GenAspI algorithm has been presented in [8]. The complexity of the GenAspI algorithm in terms of its nonzero elements and its order can be shown to be $\approx(3/4)\left(nnz(M)^2/n\right)-(3/8)\left(nnz(M)\right)+(5/8)n$ multiplications and $\approx(3/4)\left(nnz(M)^2/n\right)-(3/2)\left(nnz(M)\right)+(3/4)n$ additions, [8].

The GenAspI matrix could be used in conjunction with the iterative scheme (3) in order to derive a parametric parallel smoother. The proposed smoother is inherently parallel because the smoothing procedure is limited to matrix – vector multiplication while complex orderings and parallel algorithmic schemes are avoided. In order for a smoother to be effective the smoothing property must be satisfied, [4, 15, 16, 17, 21, 23]. The smoothing property has been proven for the Optimized Banded Generalized Approximate Inverse (OBGAIM) matrix, [9]. Equivalently, the smoothing property can be proven for the GenAspI matrix. Moreover, sharp estimates for the convergence of multigrid algorithms, for generalized smoothing, have been proven by Bank and Douglas in [2]. Furthermore, Hackbusch has proven the optimal values for the damping parameter for various iterative schemes and various PDEs, [15, 16, 17]. It can be observed that the resulting scheme requires a damping parameter ω, which cannot be defined optimally for every problem. In order to tackle the problem of the value of the damping parameter, the Dynamic Over/Under Relaxation (DOUR) algorithm is used, [18]. The DOUR scheme is predictor – corrector scheme that dynamically determines the value of ω to ensure convergence of the smoothing scheme.

Let us consider the equivalent expression for the relaxation scheme (3),

$$x_l^{(\mathrm{k+1})}=\mathrm{x}_l^{(k)}+\omega\left(S\left(x_l^{(k)}\right)\text{-}\mathrm{x}_l^{(k)}\right) \tag{8}$$

where $S\left(x_l^{(k)}\right)=\mathrm{x}_l^{(k)}+\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)_l\left(f_l\text{-}\mathrm{A}_l x_l^{(k)}\right)$.

By applying the predictor–corrector scheme, [18], we have

$$\tilde{x}_l^{(k)}=\mathrm{x}_l^{(k)}+\omega\left(S\left(x_l^{(k)}\right)\text{-}\mathrm{x}_l^{(k)}\right) \tag{9}$$

$$x_l^{(\mathrm{k+1})}=\mathrm{x}_l^{(k)}+\kappa\left(\Delta x_l^{(k)}\right),\ \Delta x_l^{(k)}=\tilde{x}_l^{(k)}\text{-}\mathrm{x}_l^{(k)} \tag{10}$$

where

$$\kappa=\frac{\left\langle \Delta x_l^{(k)},\mathrm{b}_l\text{-}\mathrm{A}_l\tilde{x}_l^{(k)}\right\rangle}{\left\langle \Delta x_l^{(k)},\mathrm{A}_l\Delta x_l^{(k)}\right\rangle} \tag{11}$$

From (8), (9), (10) and (11) we obtain

$$x_l^{(\mathrm{k+1})}=\mathrm{x}_l^{(k)}+\omega_e\left(S\left(x_l^{(k)}\right)\text{-}\mathrm{x}_l^{(k)}\right),\ \omega_e=\omega(1+\kappa) \tag{12}$$

where ω_e is the effective relaxation parameter and equation (12) is the proposed iterative scheme, [18]. The equation (12) denotes a two stage non-stationary approximate inverse smoother. Further information and convergence analysis of the DOUR algorithm were given in [18].

B. Transfer Operators

The transfer operators are special operators that are used to transfer vectors from coarser to finer grids and finer to coarser grids. The transfer operators for the multigrid method are the restriction and prolongation operators. The

restriction operator is used to transfer vectors from finer to coarser grids. An effective choice for the restriction operator is the full-weighting, which for the three-dimensional case, [21], can be expressed by the following stencil,

$$R=\frac{1}{64}\left[\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}^{2h}\begin{bmatrix}2&4&2\\4&8&4\\2&4&2\end{bmatrix}^{2h}\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}^{2h}\right]_h$$

It can be observed that both for the two-dimensional and three-dimensional case the elements of the coarse vector are the weighted average of their neighbors in the finer grid. The prolongation operator is an interpolation procedure used to transfer vectors from coarser to finer grids. An effective choice for the prolongation operator is the tri-linear interpolation. For three-dimensional problems the tri-linear interpolation can be expressed by the following stencil, [21]

$$\mathrm{P}=\frac{1}{16}\left[\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}^{\mathrm{h}}\begin{bmatrix}2&4&2\\4&8&4\\2&4&2\end{bmatrix}^{\mathrm{h}}\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}^{\mathrm{h}}\right]_{2\mathrm{h}}$$

The full-weighting operator and the tri-linear interpolation operator are related through the Galerkin condition $[P]=c[R]^T$, thus simplifying the mapping on the data, [4, 21]. Transfer operators occupy about 30% of the total computational work of the multigrid algorithm, thus choosing higher order interpolation schemes may lead to excessive computational cost at no extra gain in the convergence behavior, [4, 21]. The prolongation and restriction operators for the proposed schemes are in sparse matrix representation and thus the transfer operation between the levels is limited to matrix "times" vector multiplication and their parallelization is covered by the parallelization of the matrix-vector multiplication. Further information concerning the transfer operators can be found in [3, 4, 21, 23].

C. Cycle Strategy

The cycle strategy is the last component of the multigrid method. The cycle strategy refers to the sequence in which the grids are visited and the respective corrections are obtained, [4,21,23]. The most commonly used cycle strategy is the V-Cycle where the method descends to coarser level executing ν_1 smoother iterations in each level and then the method ascends executing ν_2 iterations in each level. The V-Cycle strategy is depicted in Figure 2.

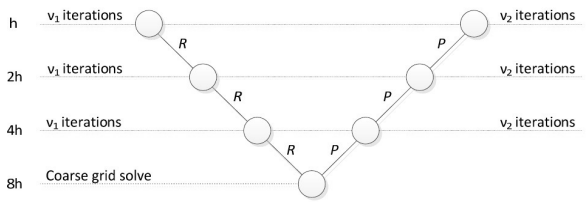


Fig. 2 The multigrid V – Cycle for four levels with ν_1 pre-smoothing iterations and ν_2 post-smoothing iterations.

The V-Cycle is parallelized only in the higher order levels to ensure the efficiency of the scheme, because lower order levels possess computational effort comparable to the parallelization overhead produced during the procedure of creating and detaching threads present in the parallel computation. Thus, levels with many nodes are computed in parallel using multiple threads, however levels with lower order are executed sequentially. As the number of levels is increased the sequential part is occupying lesser computational effort compared to higher order levels, thus increasing the speedup and efficiency of the proposed scheme.

The Parallel truncated V-Cycle multigrid algorithm with GenAspI parallel smoothing is the following, [4,21,23],

$$v^h \leftarrow \mathrm{MGV}\left(A^h,\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^h,\mathrm{v}^h,\mathrm{f}^h,\mathrm{l}\right) \quad (13)$$

if l is the coarsest level

$$\text{relax}\left(A^{\mathrm{lh}},\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^{\mathrm{lh}},\mathrm{v}^{\mathrm{lh}},\mathrm{f}^{\mathrm{lh}}\right)\ \nu_3\ \text{times} \quad (14)$$

else

if order (n) large enough

$$\text{Prelax}\left(A^h,\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^h,\mathrm{v}^h,\mathrm{f}^h\right)\ \nu_1\ \text{times} \quad (15)$$

$$\mathrm{f}^{2\mathrm{h}} \leftarrow \mathrm{Prestrict}\left(\mathrm{f}^{\mathrm{h}}-\mathrm{A}^{\mathrm{h}}\mathrm{v}^{\mathrm{h}}\right) \quad (16)$$

$$\mathrm{v}^{2\mathrm{h}} \leftarrow 0 \quad (17)$$

$$v^{2h} \leftarrow \mathrm{MGV}\left(A^{2h},\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^{2h},\mathrm{v}^{2h},\mathrm{f}^{2h},\mathrm{l}\text{-}1\right) \quad (18)$$

$$v^h \leftarrow v^h + \text{Pprolong}\left(v^h\right) \quad (19)$$

$$\text{Prelax}\left(A^h,\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^h,\mathrm{v}^h,\mathrm{f}^h\right)\ \nu_2\ \text{times} \quad (20)$$

else

$$\text{relax}\left(A^h,\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^h,\mathrm{v}^h,\mathrm{f}^h\right)\ \nu_1\ \text{times} \quad (21)$$

$$f^{2h} \leftarrow \text{restrict}\left(\mathrm{f}^{\mathrm{h}}-\mathrm{A}^{\mathrm{h}}\mathrm{v}^{\mathrm{h}}\right) \quad (22)$$

$$v^{2h} \leftarrow 0 \quad (23)$$

$$v^{2h} \leftarrow \mathrm{MGV}\left(A^{2h},\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^{2h},\mathrm{v}^{2h},\mathrm{f}^{2h},\mathrm{l}\text{-}1\right) \quad (24)$$

$$v^h \leftarrow v^h + \text{prolong}\left(v^h\right) \quad (25)$$

$$\text{relax}\left(A^h,\left(M^{\mathrm{lfill}}_{\mathrm{drptol}}\right)^h,\mathrm{v}^h,\mathrm{f}^h\right)\ \nu_2\ \text{times} \quad (26)$$

where ν_3 denotes the iterations for the inexact solution on the coarsest level. The prefix "P" denotes the parallel version of the method used. More information concerning the V-Cycle strategy as well as more Cycle schemes can be found in [3,4,21,23].

III. Particle Mesh Method Based On Multigrid Iterative Scheme

The Particle Mesh method (PM), introduced by Hockney in [19], significantly reduces the computational cost of N-body simulation algorithms in cosmological simulations of large scale structure formations. The PM method is based on the computation of the forces of the so-called "superparticles", which are composed by large formations of smaller particles and react together through their gravitational force as a whole.

In the PM method each particle contributes to the density of the concentrated mass along the grid points. Thus, each particle contributes to the neighboring points of the mass by a fraction analogous to the distance of the body from the surrounding nodes. The most commonly known method for adding the contribution of each body to the density of mass in the region is the Nearest Grid Point (NGP), [19]. The NGP method, however, increases the overall error of the PM scheme, because bodies inside a single cell share the same acceleration and force without considering the position of the particle. In order to decrease the computational error introduced during the discretization or the interpolation, the multilinear interpolation is used, namely Cloud in Cell (CIC), [19]. The Cloud In Cell method is used to interpolate the contribution of mass to every nearby grid point and raises significantly the accuracy of the computations. For the computation of the mass density of a particle with mass m_p located at (x_p, y_p, z_p) the Cloud In Cell (CIC) method is used and the respective mass densities for the eight surrounding grid points are given by equations (27), [19].

The grid points where each respectable mass contributes are depicted in Figure 3.

Let us consider the PDE (1) subjected to Dirichlet boundary conditions (1.a), discretized with the Finite Differences method and solved with the parallel GenAspI-MGV method. The solution of the resulting linear system provides the gravity potential in each point of the mesh.

$$\begin{aligned}
\rho_{i,j,k} &= \frac{m_p}{h^6}(h-\Delta_{pi})(h-\Delta_{pj})(h-\Delta_{pk}) \\
\rho_{i,j,k+1} &= \frac{m_p}{h^6}(h-\Delta_{pi})(h-\Delta_{pj})(\Delta_{pk}) \\
\rho_{i,j+1,k} &= \frac{m_p}{h^6}(h-\Delta_{pi})(\Delta_{pj})(h-\Delta_{pk}) \\
\rho_{i,j+1,k+1} &= \frac{m_p}{h^6}(h-\Delta_{pi})(\Delta_{pj})(\Delta_{pk}) \\
\rho_{i+1,j,k} &= \frac{m_p}{h^6}(\Delta_{pi})(h-\Delta_{pj})(h-\Delta_{pk}) \\
\rho_{i+1,j,k+1} &= \frac{m_p}{h^6}(\Delta_{pi})(h-\Delta_{pj})(\Delta_{pk}) \\
\rho_{i+1,j+1,k} &= \frac{m_p}{h^6}(\Delta_{pi})(\Delta_{pj})(h-\Delta_{pk}) \\
\rho_{i+1,j+1,k+1} &= \frac{m_p}{h^6}(\Delta_{pi})(\Delta_{pj})(\Delta_{pk})
\end{aligned} \tag{27}$$

The gravity potential is acquired from the solution of the linear system and the acceleration g is computed by the following equation, viz.

$$g = -\nabla\Phi = -\left(\frac{\partial\Phi}{\partial x}, \frac{\partial\Phi}{\partial y}, \frac{\partial\Phi}{\partial z}\right) \tag{28}$$

The computation of the acceleration in the center of the cells is commencing by using centered differences to retain the accuracy of the computed results. For the x-direction the acceleration can be computed by the following equation,

$$\frac{\partial\Phi}{\partial x} \approx \frac{\Phi_{i+1,j,k} - \Phi_{i-1,j,k}}{2\mathrm{h}} + O\left(h^2\right) \tag{29}$$

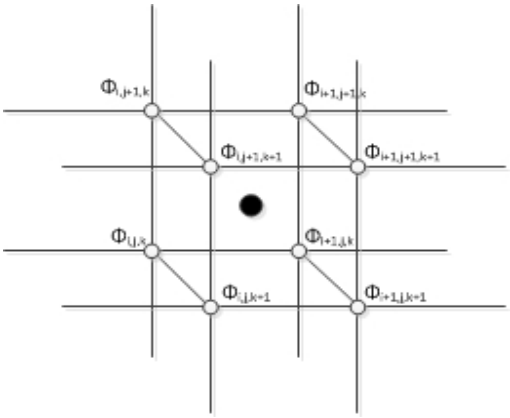


Fig. 3 Neighboring points in a three-dimensional cell of mesh size h.

The acceleration in the centers of the cells is interpolated back to the particles using the CIC method. Applying the CIC method for a particle in the three-dimensional space results in the following formula for the acceleration, [19],

$$\begin{aligned}
g_p = {} & \kappa_1 g_{i,j,k} + \kappa_2 g_{i,j,k+1} + \kappa_3 g_{i,j+1,k} + \kappa_4 g_{i,j+1,k+1} \\
& + \kappa_5 g_{i+1,j,k} + \kappa_6 g_{i+1,j,k+1} + \kappa_7 g_{i+1,j+1,k} \\
& + \kappa_8 g_{i+1,j+1,k+1}
\end{aligned} \tag{30}$$

where

$$\begin{aligned}
\kappa_1 &= (h-\Delta_{pi})(h-\Delta_{pj})(h-\Delta_{pk}) \\
\kappa_2 &= (h-\Delta_{pi})(h-\Delta_{pj})(\Delta_{pk}) \\
\kappa_3 &= (h-\Delta_{pi})(\Delta_{pj})(h-\Delta_{pk}) \\
\kappa_4 &= (h-\Delta_{pi})(\Delta_{pj})(\Delta_{pk}) \\
\kappa_5 &= (\Delta_{pi})(h-\Delta_{pj})(h-\Delta_{pk}) \\
\kappa_6 &= (\Delta_{pi})(h-\Delta_{pj})(\Delta_{pk}) \\
\kappa_7 &= (\Delta_{pi})(\Delta_{pj})(h-\Delta_{pk}) \\
\kappa_8 &= (\Delta_{pi})(\Delta_{pj})(\Delta_{pk})
\end{aligned} \tag{31}$$

while the Δ operator denotes the difference between the two points as denoted by the subscripts.

The acceleration of each mass is subject to weights computed by the distances from the nodal points of the grids. To compute the final displacement of a body, the equations of motion have to be integrated. In this article the velocity Verlet integrator is utilized due to its low computational cost and $O(h^2)$ accuracy, [22]. The Verlet integration of the equations of motion leads to the following equations,

$$\vec{x}(t+\Delta t)=\vec{x}(t)+\vec{u}(t)\Delta t+\frac{1}{2}\vec{a}(t)\Delta t^2 \qquad (32)$$

$$\vec{u}(t+\Delta t)=\vec{u}(t)+\frac{1}{2}\left(\vec{a}(t)+\vec{a}(t+\Delta t)\right)\Delta t \qquad (33)$$

where $\vec{u}, \vec{a}, \vec{x}$ are the three-dimensional vectors of the velocity, the acceleration and the position, respectively. This process is repeated according to the chosen time step and the maximum time of the simulation. It should be noted that particles cannot escape from the domain because the region is forced to be toroidal. The parallelization of the method is simplified and based on loop level parallelism using OpenMP, because the proposed schemes are mainly composed of matrix-vector multiplications and vector-vector computations. Further information concerning the Particle Mesh method can be found in [19].

IV. Numerical Results

In this section the applicability and performance of the proposed scheme is demonstrated by simulating the 3D spaces with different numbers of particles.

The numerical tests were performed on a Dual Socket AMD Opteron Processor 6128 HE, with 16GB RAM, running Ubuntu Linux 12.04.1.

The domain chosen for the simulations was the unit cube. The time step for the method was set to 0.001 and the simulation was executed for 10 consequent time steps.

The 3D region chosen was 10 parsec in each direction and the masses of the bodies were chosen to be 10^{32} kg. These variables were normalized in order to model the system on the unit cube.

The termination criterion was set to $\|r_i\|<1e\text{-}10\|r_0\|$. The pre-smoothing and post-smoothing steps were set to $\nu_1=2$ and $\nu_2=2$. The drop tolerance for the computation of the GenAspI matrices was set to drptol=0.0. The method for lfill=1 presented the fastest performance and required 7 iterations to converge to the desired tolerance. The levels below n=343 were executed sequentially because the computational overhead exceeds the computational effort required in order to obtain the coarse grid corrections.

In Table 1, the overall performance of the designed simulation algorithm based on the GenAspI-MGV method for various numbers of bodies, various numbers of threads and various resolutions, is presented. In Table 2, the speedups of the designed simulation algorithm based on the GenAspI-MGV method for various numbers of bodies, various numbers of threads and various resolutions, are presented. In Table 3, the efficiency of the designed simulation algorithm based on the GenAspI-MGV method for various numbers of bodies, various numbers of threads and various resolutions, is presented.

It should be noted that the preprocessing cost of the designed simulation algorithm concerning the computation of the Generic Approximate Sparse Inverses for resolutions h=1/16 and h=1/32 was t_{pre}=0.030093 seconds and t_{pre}=0.248670 seconds respectively, which is several orders less than the computational time needed for the model problems rendering the method efficient by slightly enlarging the computational cost.

V. Conclusion

The proposed schemes were proven experimentally effective for various choices of bodies and resolutions. Additionally, it should be stated that the proposed scheme presented satisfactory scalability as the number of particles increases and resolution is refined. Research efforts are under way to improve the parallel performance of the method with new hybrid schemes. Moreover, new computational schemes that will enhance convergence behavior and accuracy of the simulation are under further research.

Acknowledgment

The authors would like to express their thanks to Prof. K. G. Margaritis, Parallel Distributed Processing Laboratory, Department of Applied Informatics, University of Macedonia, for the provision of computational facilities.

References

[1] O. Axelsson, *Iterative solution methods.* Cambridge University Press, 1996.

[2] R.E. Bank and C.C. Douglas, "Sharp estimates for multigrid rates of convergence with general smoothing and acceleration", *SIAM Journal on Numerical Analysis*, vol. 22, pp. 617-633, 1985.

[3] A. Brandt, "Multi-level adaptive solutions to boundary-value problems", *Math. Comp.,* vol. 31, pp. 333-390, 1977.

[4] L.W. Briggs, V.E. Henson and F.S. McCormick, *A multigrid tutorial.* SIAM, 2000.

[5] O. Bröker, M.J. Grote, C. Mayer and A. Reusken, "Robust parallel smoothing for multigrid via sparse approximate inverses". *SIAM Journal on Scientific Computing,* vol. 23(4), pp. 1396–1417, 2001.

[6] E. Chow, "A priori sparsity patterns for parallel sparse approximate inverse preconditioners", *SIAM J. Sci. Comput.*, vol. 21, pp. 1804-1822, 2000.

[7] E. Chow, "Parallel implementation and practical use of sparse approximate inverses with a priori sparsity patterns", *Int. J. High Perf. Comput. Appl.*, vol. 15, pp. 56-74, 2001.

[8] C.K. Filelis-Papadopoulos and G.A. Gravvanis, "Generic Approximate Sparse Inverse Matrix Techniques", Report *TR/ECE/ASC-AMA/2012/14*, submitted.

[9] C.K. Filelis-Papadopoulos and G.A. Gravvanis, "On the multigrid method based on Finite Difference Approximate Inverses", *Computer Modeling in Engineering & Sciences*, vol. 90(3), pp. 233-253, 2013.

[10] P. Frederickson, "High performance parallel multigrid algorithms for unsctructured grids", *In 1996 Seventh Copper Mountain Conference on Multigrid Methods* CP 3339, NASA, pp. 317-326.

[11] G.A. Gravvanis, "High Performance Inverse Preconditioning", *Archives of Computational Methods in Engineering*, vol. 16(1), pp. 77-108, 2009

[12] G.A. Gravvanis, "Explicit Approximate Inverse Preconditioning Techniques", *Archives of Computational Methods in Engineering*, vol. 9(4), pp. 371-402, 2002.

[13] G. A. Gravvanis, "The rate of convergence of explicit approximate inverse preconditioning", *Inter. J. Comp. Math.*, vol. 60, pp. 77-89, 1996.

[14] G. A. Gravvanis, C. K. Filelis-Papadopoulos and P. I. Matskanidis, "Algebraic multigrid methods based on Generic Approximate Matrix Techniques", Report *TR/ECE/ASC-AMA/2012/13*, submitted.
[15] W. Hackbusch, "Multi-grid convergence theory". *In 1982 Lecture Notes in Mathematics 960*, Springer, pp. 177-219.
[16] W. Hackbusch, "On the convergence of multi-grid iterations", *Numer. Math,* vol. 9, pp. 213-239, 1981.
[17] W. Hackbusch, "Convergence of a multi-grid iteration applied to difference equations", *Math. Comp.*, vol. 34, pp. 425-440, 1980
[18] R. Haelterman, J. Viederndeels and D. Van Heule, "Non–stationary two–stage relaxation based on the principle of aggregation multi-grid", *In 2006 Computational Fluid Dynamics 2006*, Part 3, Springer, pp. 243–248.
[19] R.W. Hockney and J.W. Eastwood, *Computer Simulation Using Particles*, McGraw-Hill, 1981.
[20] E.A. Lipitakis and D.J. Evans, "Explicit semi-direct methods based on approximate inverse matrix techniques for solving boundary-value problems on parallel processors", *Math. and Computers in Simulation*, vol. 29, pp. 1-17, 1987.
[21] U. Trottenberg, C.W. Osterlee and A. Schuller, *Multigrain*. Academic Press, 2000.
[22] J. Verlet, "Computer Experiments on Classical Fluids", *In Physical Review,* vol. 159(1), pp. 98–103, 1967.
[23] P. Wesseling, "Theoretical and practical aspects of a multigrid method", *SIAM J. Sci. Stat. Comput.*, vol. 3, pp. 387-407, 1982.
[24] P. Wesseling, "The rate of convergence of a multiple grid method", Numerical analysis*, In 1980 Proceedings of the 8th bienn. Conference in Lect. Notes Math.,* 773, pp. 164-184.

TABLE I.
THE OVERALL PERFORMANCE OF THE DESIGNED SIMULATION ALGORITHM BASED ON THE GENASPI-MGV METHOD FOR VARIOUS NUMBERS OF BODIES, VARIOUS NUMBERS OF THREADS AND VARIOUS RESOLUTIONS.

N	Threads	h=1/16	h=1/32
10^5	**1**	1.260010	3.182530
	2	0.874061	2.036820
	4	0.534355	1.139530
	8	0.330208	0.722727
	16	0.269090	0.646396
10^6	**1**	10.821700	12.841400
	2	6.224200	7.336500
	4	3.630050	4.041050
	8	2.325110	2.350280
	16	1.488090	1.605920
10^7	**1**	110.123000	119.286000
	2	60.767680	65.667200
	4	34.057200	34.706500
	8	19.708600	19.180160
	16	12.293300	11.013100

TABLE II.
THE SPEEDUPS OF THE DESIGNED SIMULATION ALGORITHM BASED ON THE GENASPI-MGV METHOD FOR VARIOUS NUMBERS OF BODIES, VARIOUS NUMBERS OF THREADS AND VARIOUS RESOLUTIONS.

N	Threads	h=1/16	h=1/32
10^5	**2**	1.44	1.56
	4	2.36	2.79
	8	3.82	4.40
	16	4.68	4.92
10^6	**2**	1.74	1.75
	4	2.98	3.18
	8	4.65	5.46
	16	7.27	8.00
10^7	**2**	1.81	1.82
	4	3.23	3.44
	8	5.59	6.22
	16	8.96	10.83

TABLE III.
THE EFFICIENCY OF THE DESIGNED SIMULATION ALGORITHM BASED ON THE GENASPI-MGV METHOD FOR VARIOUS NUMBERS OF BODIES, VARIOUS NUMBERS OF THREADS AND VARIOUS RESOLUTIONS.

N	Threads	h=1/16	h=1/32
10^5	**2**	0.72	0.78
	4	0.59	0.70
	8	0.48	0.55
	16	0.29	0.31
10^6	**2**	0.87	0.88
	4	0.75	0.79
	8	0.58	0.68
	16	0.45	0.50
10^7	**2**	0.91	0.91
	4	0.81	0.86
	8	0.70	0.78
	16	0.56	0.68

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 479–486

Storing Sparse Matrices to Files in the Adaptive-Blocking Hierarchical Storage Format

Daniel Langr, Ivan Šimeček, Pavel Tvrdík
Czech Technical University in Prague
Faculty of Information Technology
Thákurova 9, 160 00, Praha, Czech Republic
Email: langrd@fit.cvut.cz

***Abstract*—When there is a need to store a sparse matrix into a file system, is it worth to convert it first into some space-efficient storage format? This paper tries to answer such question for the adaptive-blocking hierarchical storage format (ABHSF), provided that the matrix is present in memory either in the coordinate (COO) or in the compressed sparse row (CSR) storage format. The conversion algorithms from COO and CSR to ABHSF are introduced and the results of performed experiments are then presented and discussed.**

I. Introduction

SPARSE matrices are commonly present in computer memory in storage formats that provide high performance of the matrix-vector multiplication operation. Considering a generic sparse matrix without any particular pattern of its nonzero elements, such storage formats are usually not space-optimal [1]–[3]. If we need to store such a matrix in a file, we have two options:

1) either to store the matrix in its *in-memory storage format* (IMSF), in which is the matrix stored in a computer memory;
2) or to store it in some *space-efficient storage format* (SESF), which additionally requires to perform the conversion between these formats.

Question 1. *Which of these two options will take less time?*

The second option should result in a smaller file and hence its faster store operation. However, the price paid for that is the overhead of the conversion algorithm.

Let S_{IMSF} and S_{SESF} denote the amount of memory required to store a matrix in particular IMSF and SESF, respectively. The time that will be saved when storing the matrix in a file system in SESF instead of IMSF will be

$$t_{\text{saved}} = \frac{S_{\text{IMSF}} - S_{\text{SESF}}}{\text{file system I/O bandwidth}}. \quad (1)$$

Let further t_{overhead} denote the running time of the conversion algorithm between IMSF and SESF. Storing a matrix in SESF will pay off if $t_{\text{saved}} > t_{\text{overhead}}$.

The answer to Question 1 is especially important for high performance computing (HPC) applications where matrices are distributed among P processors of a massively parallel computer system (MPCS). Let L_{IMSF} and L_{SESF} denote the amount of memory required to store a local part of a matrix in IMSF and SESF, respectively, on a particular processor. If the distribution of matrix nonzero elements among processors is well-balanced, then $L_{\text{IMSF}} \approx S_{\text{IMSF}}/P$ and $L_{\text{SESF}} \approx S_{\text{SESF}}/P$, and we can rewrite (1) to

$$t_{\text{saved}} \approx \frac{L_{\text{IMSF}} - L_{\text{SESF}}}{\text{file system I/O bandwidth}} \times P. \quad (2)$$

As the size of a computational problem, and therefore the size of a given sparse matrix, varies, then:

- L_{IMSF} (and hence L_{SESF} as well) is more or less constant, since it is limited by the amount of physical memory available to a single processor on a given MPCS.
- $\boldsymbol{t_{\text{overhead}}}$ **is approximately constant**, since the IMSF-to-SESF conversion algorithm is executed independently by all processors on their local parts of the matrix (of size L_{IMSF}).
- The file system I/O bandwidth—at least its listed maximum value—is constant.
- The number of processors P varies.
- $\boldsymbol{t_{\text{saved}}}$ **varies**, according to (2), **proportionally to** $\boldsymbol{P}$.

Thus, we may expect that as the size of a computational problem grows, beyond some point it will be faster to store a sparse matrix to a file system in SESF instead of IMSF.

In this paper, we focus on situations where the IMSF is either the *coordinate* (COO) or the *compressed sparse row* (CSR) storage format [4, Section 3.4], [5, Section 4.3.1] and the SESF is the *adaptive-blocking hierarchical storage format* (ABHSF) [2]. These formats are introduced in more details in Section II. We have developed conversion algorithms from COO and CSR to ABHSF, which are presented in Section III. The experiments performed with these algorithms are then described and the results discussed in Section IV.

Note that within the context of this paper, by a *storage format* we mean the way how sparse matrices are stored in computer memory (physical/disk), by a *file format* we mean the way how sparse matrices are stored in files (in a particular storage format), and by a *storage scheme* we mean a storage format at a block level for ABHSF.

This work was supported by the Czech Science Foundation under Grant No. P202/12/2011. We acknowledge the Aerospace Research and Test Establishment in Prague, Czech Republic, for providing HPC resources.

II. Data Structures

Let A be an $m \times n$ sparse matrix with z nonzero elements. The COO storage format consist of 3 arrays of size z that contain row indexes, column indexes, and values of the nonzero elements of A. We can thus define a data structure `COO` that stores A in the COO storage format as follows:

structure `COO` := {
- m, n: matrix size;
- z: number of nonzero elements;
- $rows[]$: row indexes of nonzero elements;
- $cols[]$: column indexes of nonzero elements;
- $vals[]$: values of nonzero elements;

}.

Note that we accompany a data name with [] if the data is meant to be an array.

The advantages of the COO storage format are its clear concept, simple usage, and no requirement for the order of matrix nonzero elements. Its drawback is relatively high amount of memory needed for storing A, i.e., high S_{COO}.

If we order matrix nonzero elements according to the increasing row index, we can modify the COO storage format such that we substitute the array of row indexes by the array of positions of each row data in the remaining two arrays. Such approach results in the CSR storage format, which can be defined by the following data structure:

structure `CSR` := {
- m, n: matrix size;
- z: number of nonzero elements;
- $colinds[]$: column indexes of nonzero elements;
- $vals[]$: values of nonzero elements;
- $rowptrs[]$: offsets of places where data of each row in the $colinds$ and $vals$ arrays start;

}.

Since the array $rows[]$ of `COO` is of size z and the array $rowptrs[]$ of `CSR` is of size m (usually $m+1$ for the sake of simpler implementation), and since $z \gg m$ usually holds for real-world sparse matrices, the CSR storage format typically requires considerably less amount of memory for storing A when compared with COO, i.e., $S_{\text{CSR}} < S_{\text{COO}}$.

ABHSF is a two-level hierarchical storage format (see Figure 1) based on partitioning a matrix into $\lceil m/s \rceil \times \lceil n/s \rceil$ blocks of size $s \times s$ and storing each nonzero block in its space-optimal storage scheme. This approach can considerably reduce the memory requirements for storing sparse matrices when compared not only with COO and CSR but also with other fixed-scheme hierarchical storage formats.

We consider the following storage schemes within this text:

1) **dense:** all block elements are stored including zeros,
2) **bitmap:** only nonzero block elements are stored and their structure is defined by a bit map,
3) **COO:** equivalent of the COO storage format at a block level,
4) **CSR:** equivalent of the CSR storage format at a block level.

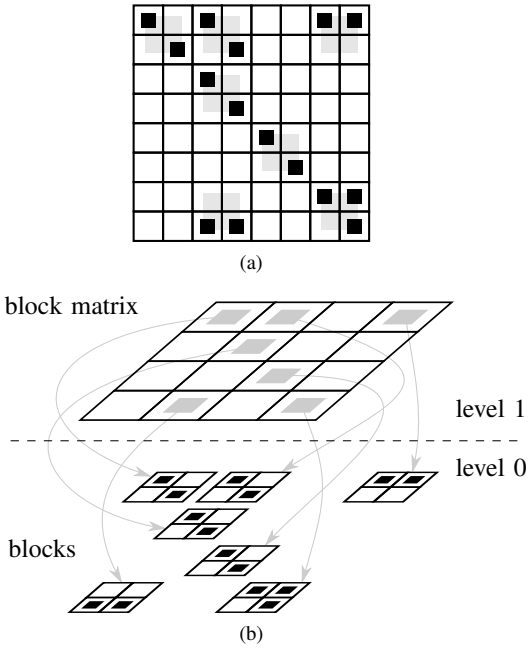


Fig. 1: An 8×8 matrix (a) represented as a hierarchical data structure with 2×2 blocks (b).

The optimal block size s for a particular matrix is application dependent, however, block sizes between 64 and 256 provide best results in general.

We can define a data structure `ABHSF` that stores A in the ABHSF format as follows:

structure `ABHSF` := {
- m, n: matrix size;
- z: number of nonzero elements;
- s: block size;
- $brows[]$: row indexes of nonzero blocks;
- $bcols[]$: column indexes of nonzero blocks;
- $zetas[]$: the number of nonzero elements of nonzero blocks;
- $schemes[]$: optimal storage scheme tag of nonzero blocks;
- $bitmap[]$: a bit map;
- $lrows[]$: row indexes local to a block;
- $lcols[]$: column indexes local to a block;
- $lrowptrs[]$: offsets of places where data of each row of a block start;
- $vals[]$: values of the elements of nonzero blocks;

}.

More details about ABHSF are beyond the scope of this paper, however, they were presented by Langr et at. [2].

Let `BLOCK` be an auxiliary data structure used for storing data of a single nonzero block, defined as follows:

structure `BLOCK` := {
- $brow$: row index of a block within A;
- $bcol$: column index of a block within A;

$zeta$: a number of block nonzero elements;
$lrows[]$: local row indexes of nonzero elements;
$lcols[]$: local column indexes of nonzero elements;
$lvals[]$: local values of nonzero elements;
}.

III. Algorithms

We further suppose that all indexes are 0-based (as used in the C/C++ programming languages).

A. Conversion from COO to ABHSF

The pseudocode of the conversion process from COO to ABHSF is presented as Algorithm 1. It is based on the successive gathering and processing of data for each nonzero block. To optimize this process, the nonzero elements of A are first sorted, at line 5, according to the key represented by the following quadruple with left-to-right significance of its elements:

$$\Big(\lfloor coo.rows[i]/s \rfloor, \lfloor coo.cols[i]/s \rfloor, coo.rows[i] \bmod s, coo.cols[i] \bmod s\Big). \quad (3)$$

Such ordering puts the nonzero elements of each nonzero block into a continuous chunks within the arrays of the `COO` data structure. Consequently, the conversion can be performed within a single iteration over matrix nonzero elements (lines 7–21).

The ProcessBlock procedure, which stores data of a single nonzero block into the `ABHSF` data structure at line 20, is defined in Section III-C.

B. Conversion from CSR to ABHSF

The pseudocode of the conversion process from CSR to ABHSF is presented as Algorithm 2. It is based on the same principle as Algorithm 1, i.e., on the successive gathering and processing of the data of individual nonzero blocks. However, this process is here more complicated, since we cannot simply reorder the matrix nonzero elements according to (3), as in Algorithm 1. Therefore, instead of iterating over matrix nonzero elements, Algorithm 2 iterates over all blocks (loops at lines 5 and 15) and for each block it tries to obtain its nonzero elements. If there are any, they are then processed by the ProcessBlock procedure as well.

C. Processing Blocks

The ProcessBlock procedure stores data of a single nonzero block into the `ABHSF` data structure. Its pseudocode is shown as Algorithms 3.

We assume that the `ABHSF` data structure represents an open file and that all updates into this data structure will be directly translated into corresponding updates of its file representation. Within the pseudocode, we regard all `ABHSF` arrays as file output streams/virtual dynamic arrays to which the elements are successively *appended*.

The space-optimal storage schemes for blocks are selected at line 1 line by comparing their memory requirements, which were defined by Langr et al. [2] as functions (1a)–(1d). According to the optimal storage scheme, the block data are then stored into the `ABHSF` data structure as follows:

- **dense** (lines 7–17): The procedure iterates over all elements of a block. If the corresponding nonzero element is found, then its value is appended to the $vals[]$ array of the `ABHSF` data structure. Otherwise, 0 is appended instead.
- **bitmap** (lines 19–30): The procedure iterates over all elements of a block. If the corresponding nonzero element is found, then its value is appended to the $vals[]$ array of the `ABHSF` data structure and 1 is appended to the $bitmap[]$ array. Otherwise, 0 is appended to the to the $bitmap[]$ array instead.
- **COO** (lines 32–36): The procedure iterates over nonzero block elements and appends their row/column indexes and values to the corresponding arrays of the `ABHSF` data structure.
- **CSR** (lines 38–50): The procedure iterates over nonzero block elements and appends their column indexes and values to the corresponding arrays of the `ABHSF` data structure, while it also constructs the array $lrowptrs[]$ of positions of data for each row of a block.

Note that to ProcessBlock work properly, the nonzero elements in the input `BLOCK` data structure need to be ordered according to growing row index and for each row according to the growing column index. Both Algorithm 1 and Algorithm 2 conform to this requirement.

IV. Experiments and Discussion

We have designed and performed experiments to evaluate the suitability of storing sparse matrices in files in ABHSF. Within these experiments, same matrices were stored in the COO, CSR, and ABHSF storage formats into files based on the HDF5 file format [6] so that particular data from the `COO`, `CSR`, and `ABHSF` data structures were stored as HDF5 attributes and data sets.

Within our implementation, the `ABHSF` data structure represented an open file, i.e., all updates to its data were directly translated into updates of corresponding file attributes and data sets. The data types for data sets containing indexes were always chosen to be unsigned integer data types of minimal possible bit width. All floating-point numbers were stored in files in single precision.

For all the experiments, we used the block size $s = 256$, which generally provides reasonable results for a wide range of matrices, as shown by Langr et al. [2]. To achieve maximum performance, we implemented experimental programs so that:

- All HDF5 data sets were chosen to be *fixed-size*. The conversion algorithms hence needed to be executed twice. Within the first *dry run*, the sizes of data sets were computed. Within the second run, data were actually written into them. (Sorting of elements in Algorithm 1 was performed only once within the dry run.)
- All updates of HDF5 data sets were buffered. We used buffers of size 1024 elements for each data set.

Algorithm 1: Conversion from COO to ABHSF

Input: coo: `COO`; s: integer
Output: $abhsf$: `ABHSF`
Data: $block$: `BLOCK`; $k, brow, bcol$: integer

```
1  abhsf.m ← coo.m
2  abhsf.n ← coo.n
3  abhsf.z ← coo.z
4  abhsf.s ← s
5  sort the coo.rows, coo.cols, and coo.vals arrays all at once according to (3)
6  k ← 0
7  while k < coo.z do                                          // iterate over nonzero elements
8      block.brow ← ⌊coo.rows[k]/s⌋
9      block.bcol ← ⌊coo.cols[k]/s⌋
10     block.zeta ← 0
11     while ⌊coo.rows[k]/s⌋ = block.brow and ⌊coo.cols[k]/s⌋ = block.bcol do   // while element is in the actual block
12         block.lrows[block.zeta] ← coo.rows[k] mod s
13         block.lcols[block.zeta] ← coo.cols[k] mod s
14         block.lvals[block.zeta] ← coo.vals[k]
15         block.zeta ← block.zeta + 1
16         k ← k + 1                                           // go to next nonzero element
17         if k ≥ coo.z then break
18
19     end
20     ProcessBlock(block, abhsf)                              // store block data into the ABHSF structure
21 end
```

- All the *bitmap*, *lrows*, *lcols*, and *lrowptrs* arrays from the `ABHSF` data structure were implemented in files as a single data set.

Reasoning for the listed decisions is beyond the scope of this paper. However, they all originated from results of our complementary tests and measurements.

A. File Sizes for Benchmark Matrices

First, we compared sizes of files for different matrices that emerged in real-world scientific and engineering applications. All used *benchmark matrices* were taken from *The University of Florida Sparse Matrix Collection* (UFSMC) [7]. We tried to choose matrices from different computational domains and therefore with different structural properties. Their list together with their characteristics is presented in Table I, where z' denotes the relative number of nonzero elements in percents, i.e., the inverse measure of sparsity of a matrix.

We stored matrices in HDF5-based files in the COO, CSR, and ABHSF storage formats. We further refer to these options as HDF5-COO, HDF5-CSR, and HDF5-ABHSF, respectively. The results are presented in Figure 2 where the file sizes are relative (in percents) to the HDF5-ABHSF option. For comparison, we also included the sizes of compressed (`.mtx.gz`) and uncompressed (`.mtx`) files in the Matrix Market file format [8], in which matrices are originally published in UFSMC.

The main conclusion from these results is that for all benchmark matrices, HDF5-COO resulted in files about twice as big as HDF5-ABHSF and HDF5-CSR resulted in files about 1.4 times bigger than HDF5-ABHSF. Therefore, **if we convert matrices to ABHSF, we can save a considerable amount of file system capacity.**

Unfortunately, we cannot simply compare the results for the text-based Matrix Market file format and the binary-based HDF5 file format, since in the text-based file formats, the floating-point values are generally represented in various precisions. However, note that for some matrices the smallest files were achieved for the compressed Matrix Market file format. This effect was caused by the fact that in these special cases, many matrix elements had identical floating-point values, which led to high efficiency of text compression. HDF5 also allows to compress data, which should reduce the sizes of data sets containing repeated floating-point values. We have, however, not tested this possibility.

B. Parallel Experiments

The matrices in UFSMC are of smaller sizes suitable for sequential rather than parallel processing. Since we did not find any suitable scalable HPC application able to generate very large sparse matrices, we simulated such matrices by parallel generation of random block matrices. The developed generating algorithm works as follows:

1) an *imaginary* global matrix is partitioned into P submatrices,
2) each submatrix is further treated by a single processor,
3) each submatrix is partitioned into blocks,

Algorithm 2: Conversion from CSR to ABHSF

Input: csr: CSR; s: integer
Output: $abhsf$: ABHSF
Data: $block$: BLOCK; $k, brow, bcol, firstrow, lastrow, nrows, row, lrow$: integer; $from$, $remains$: integer array

```
1  abhsf.m ← csr.m
2  abhsf.n ← csr.n
3  abhsf.z ← csr.z
4  abhsf.s ← s
5  for brow ← 0 to ⌈csr.m/s⌉ − 1 do                                   // iterate over block rows
6      firstrow ← brow · s
7      if firstrow + s ≤ csr.m then nrows ← s
8      else nrows ← csr.m − firstrow
9
10     lastrow ← firstrow + nrows − 1
11     for row ← firstrow to lastrow do                               // for each row of a block row find out:
12         from[row − firstrow] ← csr.rowptrs[row]                    // position of data to be processed
13         remains[row − firstrow] ← csr.rowptrs[row + 1] − csr.rowptrs[row]   // number of elements to be processed
14     end
15     for bcol ← 0 to ⌈csr.n/s⌉ − 1 do                               // iterate over block columns
16         block.brow ← brow
17         block.bcol ← bcol
18         block.zeta ← 0
19         for row ← firstrow to lastrow do                           // for each row of a block row
20             lrow ← row − firstrow
21             while remains[lrow] > 0 and csr.colinds[from[lrow]] < (bcol + 1) · s do
                                                                      // while elements belong to the actual block
22                 block.lrows[block.zeta] ← lrow
23                 block.lcols[block.zeta] ← csr.colinds[from[lrow]] − bcol · s
24                 block.lvals[block.zeta] ← csr.vals[from[lrow]]
25                 block.zeta ← block.zeta + 1
26                 from[lrow] ← from[lrow] + 1
27                 remains[lrow] ← remains[lrow] − 1
28             end
29         end
30         if block.zeta > 0 then ProcessBlock(block, abhsf)
                                                                      // store block data into the ABHSF data structure
31     end
32 end
```

4) each block becomes nonzero with some probability,
5) each nonzero block contains a random number of nonzero elements,
6) each nonzero element is assigned a random row/column index and a random value.

We further set up the generator so that:

- $L_{\text{COO}} \approx 600$ MB, therefore submatrices took about 600 MB each when stored in COO (the typical amount of physical memory per processor is 1–2 GB on contemporary MPCSs).
- Resulting matrices contained nonzero blocks of various properties, thus various storage schemes were generally space-optimal for them.
- Processors used different seeds for their instances of a pseudorandom number generator to produce different submatrices. However, these seeds were preserved in time, thus each processor generated the very same submatrices through all experiments.

The parallel experiments were carried out in the following steps:

1) Each processor generated a random submatrix and stored it in memory either in COO or in CSR.
2) All processors stored their submatrices to a file system in the original IMSF, which resulted in HDF5-COO or HDF5-CSR files.
3) All processors stored their submatrices to a file system in the ABHSF storage format utilizing either Algorithm 1 or Algorithm 2, which resulted in HDF5-ABHSF files.

Algorithm 3: PROCESSBLOCK(b, a)

Input: $block$: BLOCK; $abhsf$: ABHSF
Output: $abhsf$: ABHSF
Data: $scheme$: scheme tag; k, row, col: integer

```
scheme ← space-optimal storage scheme for block block   // functions (1a)–(1d) defined by Langr et al. [2]
append block.brow to abhsf.brows
append block.bcol to abhsf.bcols
append scheme to abhsf.schemes
append block.zeta to abhsf.zetas
if scheme = dense then                                  // optimal scheme is dense
    k ← 0
    for row ← 0 to abhsf.s − 1 do                       // iterate over all block elements
        for col ← 0 to abhsf.s − 1 do
            if k < block.zeta and block.lrows[k] = row and block.lcols[k] = col then   // if element exists
                append block.lvals[k] to abhsf.vals     // store its nonzero value
                k ← k + 1
            else
                append 0 to abhsf.vals                  // otherwise store 0
            end
        end
    end
else if scheme = bitmap then                            // optimal scheme is bitmap
    k ← 0
    for row ← 0 to abhsf.s − 1 do                       // iterate over all block elements
        for col ← 0 to abhsf.s − 1 do
            if k < block.zeta and block.lrows[k] = row and block.lcols[k] = col then   // if element exists
                append block.lvals[k] to abhsf.vals     // store its nonzero value
                append 1 to abhsf.bitmap                // and 1 to bit map
                k ← k + 1
            else
                append 0 to abhsf.bitmap                // otherwise store 0 to bitmap
            end
        end
    end
else if scheme = COO then                               // optimal scheme is COO
    for k ← 0 to block.zeta − 1 do                      // iterate over block nonzero elements
        append block.lrows[k] to abhsf.lrows            // and store them into COO storage scheme
        append block.lcols[k] to abhsf.lcols
        append block.lvals[k] to abhsf.vals
    end
else if scheme = CSR then                               // optimal scheme is CSR
    row ← 0
    for k ← 0 to block.zeta − 1 do                      // iterate over block nonzero elements
        while row ≤ block.lrows[k] do                   // and store them in the CSR storage scheme
            append k to abhsf.lrowptrs
            row ← row + 1
        end
        append block.lcols[k] to abhsf.lcols
        append block.lvals[k] to abhsf.vals
    end
    while row ≤ abhsf.s do                              // align final rows if needed
        append block.zeta to abhsf.lrowptrs
        row ← row + 1
    end
end
```

Matrix	Domain	m	n	z' [%]	Symmetric
ldoor	structural problem	$9.5 \cdot 10^5$	$9.5 \cdot 10^5$	$2.6 \cdot 10^{-3}$	yes
Freescale1	circuit simulation	$3.4 \cdot 10^6$	$3.4 \cdot 10^6$	$1.6 \cdot 10^{-4}$	no
atmosmodj	computational fluid dynamics	$1.3 \cdot 10^6$	$1.3 \cdot 10^6$	$5.5 \cdot 10^{-4}$	no
cage12	directed weighted graph	$1.3 \cdot 10^5$	$1.3 \cdot 10^5$	$1.2 \cdot 10^{-2}$	no
ohne2	semiconductor device	$1.8 \cdot 10^5$	$1.8 \cdot 10^5$	$3.3 \cdot 10^{-2}$	no
FEM_3D_thermal2	thermal problem	$1.5 \cdot 10^5$	$1.5 \cdot 10^5$	$1.6 \cdot 10^{-2}$	no
bmw7st_1	structural problem	$1.4 \cdot 10^5$	$1.4 \cdot 10^5$	$1.8 \cdot 10^{-2}$	yes
nlpkkt120	optimization problem	$3.5 \cdot 10^6$	$3.5 \cdot 10^6$	$4.0 \cdot 10^{-4}$	yes

TABLE I: The list of the benchmark matrices used for the performed experiments.

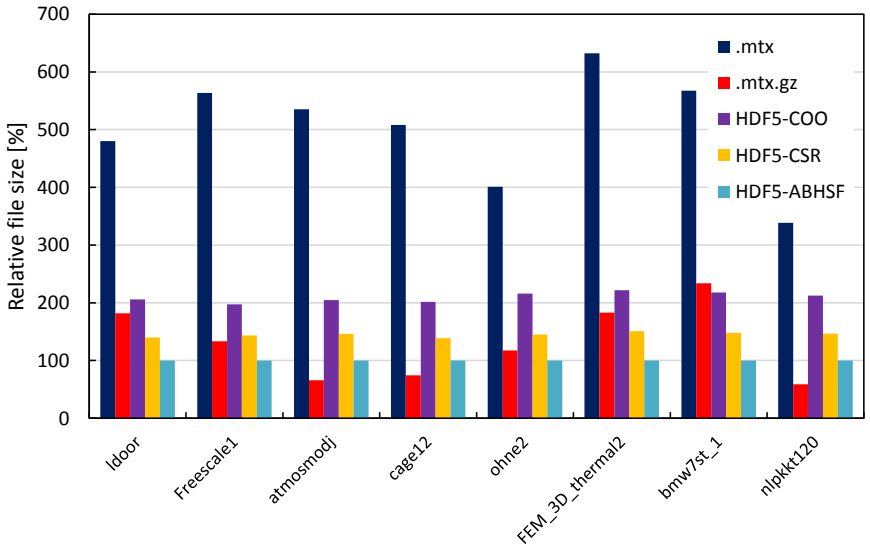


Fig. 2: File sizes in percents relative to HDF5-ABHSF for benchmark matrices and different file/storage formats.

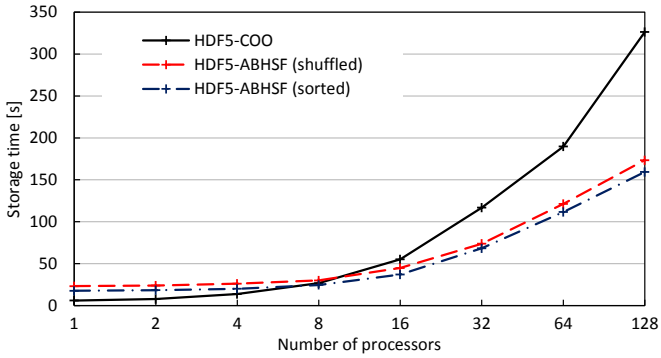


Fig. 3: Storage times for cases when COO was used as IMSF.

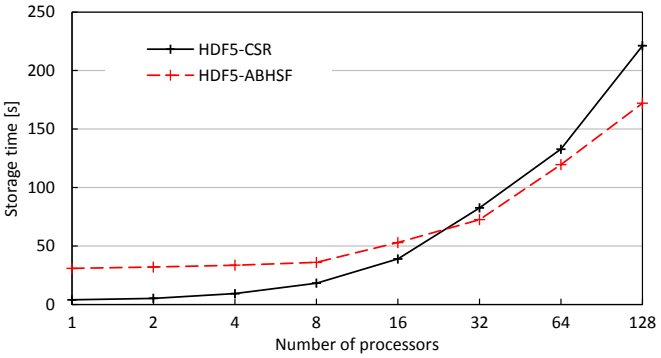


Fig. 4: Storage times for cases when CSR was used as IMSF.

We measured the storage times of steps 2 and 3 for different numbers of processors. Results for the case of using COO and CSR as IMSF are in Figure 3 and Figure 4, respectively. All measurements were performed 3 times and the average values are presented. Since the conversion algorithm from COO to ABHSF contains sorting of elements, we kept nonzero elements in memory in COO in two different orderings to evaluate the influence of the sorting algorithm. In the first case the elements were randomly *shuffled* and in the second case they were *sorted* by their (row index, column index) keys.

The obtained results clearly correspond to our assumption introduced in Section I. In case of ABHSF, there was some constant computational overhead imposed by the conversion algorithms (note that this overhead was considerably higher for the conversion from CSR, which was caused by the higher complexity of the conversion algorithm compared with the

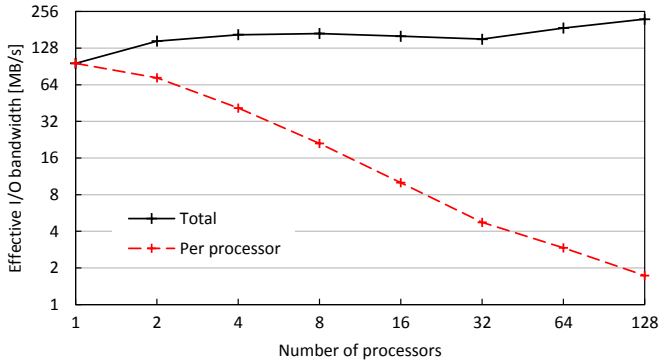


Fig. 5: Measured I/O bandwidth of the file system used for experiments.

COO case). For smaller numbers of processors, this overhead dominated the overall storage time ($t_{\text{saved}} < t_{\text{overhead}}$) and therefore it was faster to store matrices to a file system in their original IMSFs. However, **as the matrix size increased** (proportionally to P), **the amount of saved data** ($S_{\text{IMSF}} - S_{\text{SESF}}$) **increased as well, and from some point, it paid off to store matrices in ABHSF.**

From the measured storage times, we have also computed the *total I/O bandwidth* of the used file system. Provided that this total bandwidth is shared by all processors evenly, we can also express the *I/O bandwidth per processor*. These values are shown in Figure 5.

C. Generalization of Results

If we assume that the total I/O file system bandwidth is shared evenly among processors, we may rewrite (2) as

$$t_{\text{saved}} \approx \frac{L_{\text{IMSF}} - L_{\text{SESF}}}{\text{file system I/O bandwidth per processor}}.$$

This implies that **the amount of saved time generally grows inversely proportionally to the file system I/O bandwidth per processor**.

Within our experiments, we utilized a small parallel computer system with the GPFS-based storage subsystem [9]. The total I/O bandwidth of this file system varied, according to Figure 5, approximately from 100 to 200 MB/s. On this system, the point where ABHSF started to provide faster storage of matrices emerged around 16–32 processors, which corresponded to the I/O bandwidth of 4–8 MB/s per processor.

On todays biggest MPCSs, the per-processor I/O bandwidth would be typically much lower for large-scale computations. For instance, the Hopper/NERSC MPCS consists of over 153 thousands processors (CPU cores) and the listed maximum I/O bandwidth of its fastest file system is 35 GB/s. Therefore, we cannot get the I/O bandwidth per processor higher than 0.23 MB/s when utilizing the whole system. In general, for such low I/O rates, we may expect that the ABHSF storage format would be much more superior to the original IMSF when storing matrices to a file systems.

V. Conclusions

The contribution of this paper are new conversion algorithms for sparse matrices from the COO and CSR to the ABHSF storage formats and the evaluation of suitability of storing sparse matrices into file systems in ABHSF using these algorithms, with the focus on the HPC application domain. We showed that as the size of a computational problem grows, and so does the number of processors, there is some point from which it pays off to store matrices to a file system in ABHSF instead of their original IMSF.

Unfortunately, we cannot simply predict this point, since it depends on many factors, such as the I/O bandwidth of the file system, the actual workload of the file system, the clock rate of processors, the bandwidth of memory units, the available amount of physical memory per processor, the quality of the compiler, the quality of the program code, structural properties of the matrix, etc. However, provided that we use a particular MPCS and a particular HPC application that generates matrices with similar properties, many of these factors becomes fixed. Moreover, computational power and compiler capabilities that influence the overhead imposed by the conversion algorithms generally do not differ much across contemporary MPCSs. Then, the suitability of storing matrices to a file system in ABHSF (generally in any SESF) would be determined primarily by the I/O bandwidth of the file system per processor.

References

[1] Ivan Šimeček, Daniel Langr, and Pavel Tvrdík. Space-efficient sparse matrix storage formats for massively parallel systems. In *Proceedings of the 14th IEEE International Conference of High Performance Computing and Communications (HPCC 2012)*, pages 54–60. IEEE Computer Society, 2012.

[2] D. Langr, I. Šimeček, P. Tvrdík, T. Dytrych, and J. P. Draayer. Adaptive-blocking hierarchical storage format for sparse matrices. In *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS 2012)*, pages 545–551. IEEE Xplore Digital Library, September 2012.

[3] I. Šimeček, D. Langr, and P. Tvrdík. Minimal quadtree format for compression of sparse matrices storage. In *Proceedings of the 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012)*. IEEE Computer Society, September 2012. Accepted for publication.

[4] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.

[5] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, PA, 2nd edition, 1994.

[6] The HDF Group. Hierarchical data format version 5, 2000-2013. http://www.hdfgroup.org/HDF5/ (accessed June 3, 2013).

[7] T. A. Davis and Y. F. Hu. The University of Florida Sparse Matrix Collection. *ACM Transactions on Mathematical Software*, 38(1), 2011.

[8] Ronald F. Boisvert, Roldan Pozo, and Karin Remington. The Matrix Market Exchange Formats: Initial Design. Technical Report NISTIR 5935, National Institute of Standards and Technology, Dec. 1996.

[9] Frank Schmuck and Roger Haskin. GPFS: A shared-disk file system for large computing clusters. In *Proceedings of the 1st USENIX Conference on File and Storage Technologies*, FAST '02, Berkeley, CA, USA, 2002. USENIX Association.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 487–493

Schur Complement Domain Decomposition in conjunction with Algebraic Multigrid methods based on Generic Approximate Inverses

P.I. Matskanidis
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Kimmeria, GR 67100 Xanthi, Greece
Email: pascmats@ee.duth.gr

G.A. Gravvanis
Department of Electrical and Computer Engineering, School of Engineering, Democritus University of Thrace, University Campus, Kimmeria, GR 67100 Xanthi, Greece
Email: ggravvan@ee.duth.gr

***Abstract*—For decades, Domain Decomposition (DD) techniques have been used for the numerical solution of boundary value problems. In recent years, the Algebraic Multigrid (AMG) method has also seen significant rise in popularity as well as rapid evolution. In this article, a Domain Decomposition method is presented, based on the Schur complement system and an AMG solver, using generic approximate banded inverses based on incomplete LU factorization. Finally, the applicability and effectiveness of the proposed method on characteristic two dimensional boundary value problems is demonstrated and numerical results on the convergence behavior are given.**

I. Introduction

DOMAIN decomposition includes a significant range of computing techniques for the numerical solution of Partial Differential Equations (PDEs). Domain decomposition techniques are based on splitting the computational domain into smaller subdomains, with or without overlap. The problems in the subdomains are independent, thus rendering domain decomposition methods suitable for parallelization. Domain decomposition techniques can themselves be used as stationary iterative schemes, as well as preconditioners in order to accelerate the convergence of other iterative methods, specifically Krylov subspace methods [14].

Domain decomposition methods can be split into two categories: overlapping and non-overlapping methods. In overlapping DD methods, often referred to as Schwarz methods due to Schwarz's work in 1870 [22], the subdomains overlap by more than the interface. The overlapping methods have a simple algorithmic structure, since there is no need to solve special interface problems between neighbouring subdomains. This feature differentiates overlapping from non-overlapping DD methods [5],[23]. Overlapping methods operate by an iterative procedure, where the PDE is repeatedly solved within every subdomain. For each subdomain, the artificial internal boundary condition is provided by its neighbouring subdomains. The convergence of the solution on these internal boundaries ensures the convergence of the solution in the entire solution domain.

In non-overlapping DD methods, also referred to as iterative substructuring methods, the subdomains intersect only on their interface. Non-overlapping methods can furthermore be distinguished in primal and dual methods. Primal methods, such as BDDC [7], enforce the continuity of the solution across the subdomain interface by representing the value of the solution on all neighbouring subdomains by the same unknown. In dual methods, such as FETI [10], the continuity is further enforced by the use of Lagrange multipliers. Hybrid methods, such as FETI-DP [8],[9],[16], have also been introduced.

In the past decades, the development of multigrid methods has also been critical for the numerical solution of PDEs. An essential component of the multigrid method is the relaxation scheme, which efficiently reduces high frequency components of the error, however is inefficient at reducing the lower frequency ones [2]. Transferring the problem to a coarser grid, those low frequency errors become more oscillatory and can be effectively damped by a stationary iterative method. Recursive application of this process produces the multigrid methods [18].

The algebraic multigrid algorithm (AMG) was first introduced over twenty years ago [1],[20]. Unlike geometric multigrid, the algebraic multigrid method does not require knowledge of the geometry of the problem to define its components. This is the reason AMG is perfectly suited for unstructured grids, both in two and three dimensions, and complicated domains. Specifically, by considering a linear system Au=f, the AMG method requires only the coefficient matrix A and the right-hand side vector f. As a result, AMG solvers can easily be integrated into existing problem solving environments as standard solvers or preconditioners [18].

Consider a linear system Au=f, where $A=\left(a_{i,j}\right), i,j\in[1,n]$ is an (n×n) coefficient matrix. A "grid" is a set of indices of the variables, thus the first grid is $\Omega=\{1,2,\ldots,n\}$. Since AMG is independent of the geometry of the problem, the coarser grids, where the successive corrections to the solu-

978-83-60810-53-8/$25.00

tion will be obtained, have to be constructed by the coarsening process, which is an essential component of the AMG algorithm.

The components needed for AMG, where superscripts indicate the level with 1 being the finest level [4],[27], are the following:

- Grids $\Omega^1 \supset \Omega^2 \supset \ldots \supset \Omega^N$ ($\Omega^1=\Omega$) containing the following two disjoint subsets:
  Coarse points set (C-points): $C^k, \mathrm{k}=1,\ldots,\mathrm{N}-1$.
  Fine points set (f-points): $F^k, \mathrm{k}=1,\ldots,\mathrm{N}-1$.
- Grid operators: $A^1, \mathrm{A}^2, \ldots, \mathrm{A}^N$, where $A^1 = A$.
- Interpolation and restriction operators:
  $I_{k+1}^k, \mathrm{k}=1,\ldots,\mathrm{N}-1$; $I_k^{k+1}, \mathrm{k}=1,\ldots,\mathrm{N}-1$.
- A smoother (relaxation scheme) for each level.

AMG consists of two main phases: the setup phase, where the above components are created, and the solution phase that utilizes the components in the recursively defined multigrid cycle.

In this article, a domain decomposition method is presented, based on the Schur complement system. An Algebraic Multigrid method is used to solve the resulting linear systems for each domain, based on the use of generic approximate banded inverses, derived from the ILU(0) factorization [13],[18]. In section II, the Schur complement method is showcased, while in section III, the AMG method is presented.

Finally in section IV, the applicability of the new proposed scheme on two dimensional boundary value problems is demonstrated and numerical results on the convergence behavior and performance are given.

II. The Schur Complement Method

In this section, the Schur complement domain decomposition method is presented.

The Schur complement method is the earliest version of non-overlapping DD methods. Methods such as Dirichlet-Neumann and Neumann-Neumann are essentially the Schur complement method with the use of particular preconditioners.

Let us consider the Poisson equation on a region Ω, with zero Dirichlet data given on $\partial\Omega$, the boundary of Ω. Let us also suppose that Ω is partitioned into two non-overlapping subdomains Ω_i:

$$\overline{\Omega} = \overline{\Omega_1 \cup \Omega_2}, \Omega_1 \cap \Omega_2 = \emptyset, \Gamma = \partial\Omega_1 \cap \partial\Omega_2$$

as shown in Fig. 1 [25].

Assuming the boundaries of the subdomains are Lipschitz continuous, we consider the problem:

$$-\Delta \mathrm{u}(\mathrm{x},\mathrm{y}) = f \quad (\mathrm{x},\mathrm{y}) \in \Omega \tag{1}$$

$$u(\mathrm{x},\mathrm{y}) = 0 \quad (\mathrm{x},\mathrm{y}) \in \partial\Omega \tag{1.a}$$

Considering a triangulation of the domain Ω and a finite element approximation of the problem (1), assuming that subdomains consist of unions of elements, leads to a linear system

$$\mathrm{Au}=\mathrm{f} \tag{2}$$

with a symmetric, positive definite matrix A. Partitioning the degrees of freedom to those internal to Ω_1, Ω_2 and those interior of Γ, the matrix A and vectors u, f can be expressed as:

$$A = \begin{bmatrix} A_{\mathrm{II}}^{(1)} & 0 & A_{\mathrm{I\Gamma}}^{(1)} \\ 0 & A_{\mathrm{II}}^{(2)} & A_{\mathrm{I\Gamma}}^{(2)} \\ A_{\Gamma\mathrm{I}}^{(1)} & A_{\Gamma\mathrm{I}}^{(2)} & A_{\Gamma\Gamma} \end{bmatrix}, \mathrm{u} = \begin{bmatrix} u_I^{(1)} \\ u_I^{(2)} \\ u_\Gamma \end{bmatrix}, \mathrm{f} = \begin{bmatrix} f_I^{(1)} \\ f_I^{(2)} \\ f_\Gamma \end{bmatrix} \tag{3}$$

The first step of many iterative domain decompositions methods eliminates the unknowns in the interior of the subdomains $u_I^{(i)}$. This leads to a block factorization of the matrix A (3) [25]:

$$A = \mathrm{LR} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A_{\Gamma\mathrm{I}}^{(1)} A_{\mathrm{II}}^{(1)^{-1}} & A_{\Gamma\mathrm{I}}^{(2)} A_{\mathrm{II}}^{(2)^{-1}} & I \end{bmatrix} \begin{bmatrix} A_{\mathrm{II}}^{(1)} & 0 & A_{\mathrm{I\Gamma}}^{(1)} \\ 0 & A_{\mathrm{II}}^{(2)} & A_{\mathrm{I\Gamma}}^{(2)} \\ 0 & 0 & S \end{bmatrix} \tag{4}$$

and the resulting linear system:

$$\begin{bmatrix} A_{\mathrm{II}}^{(1)} & 0 & A_{\mathrm{I\Gamma}}^{(1)} \\ 0 & A_{\mathrm{II}}^{(2)} & A_{\mathrm{I\Gamma}}^{(2)} \\ 0 & 0 & S \end{bmatrix} u = \begin{bmatrix} f_I^{(1)} \\ f_I^{(2)} \\ g_\Gamma \end{bmatrix} \tag{5}$$

where I is the identity matrix and $S = A_{\Gamma\Gamma} - A_{\Gamma\mathrm{I}}^{(1)} A_{\mathrm{II}}^{(1)^{-1}} A_{\mathrm{I\Gamma}}^{(1)} - A_{\Gamma\mathrm{I}}^{(2)} A_{\mathrm{II}}^{(2)^{-1}} A_{\mathrm{I\Gamma}}^{(2)}$ is the Schur complement matrix relative to the unknowns on Γ.

Defining the local Schur complements by

$$S^{(i)} := A_{\Gamma\Gamma}^{(i)} - \sum_{i=1}^{2} A_{\Gamma\mathrm{I}}^{(i)} A_{\mathrm{II}}^{(i)^{-1}} A_{\mathrm{I\Gamma}}^{(i)} \tag{6}$$

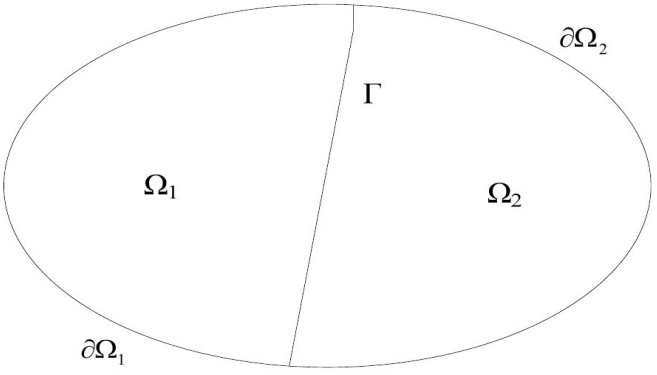


Fig. 1 Ω partitioned into two non-overlapping subdomains.

we find the Schur complement system for u_Γ to be [25]:

$$\mathrm{Su}_\Gamma = g_\Gamma \tag{7}$$

with

$$S = S^{(1)} + S^{(2)} \tag{8}$$

$$\begin{aligned} g_\Gamma := g_\Gamma^{(1)} + g_\Gamma^{(2)} = & \left(f_\Gamma^{(1)} - A_{\Gamma\mathrm{I}}^{(1)} A_{\mathrm{II}}^{(1)^{-1}} f_I^{(1)} \right) \\ & + \left(f_\Gamma^{(2)} - A_{\Gamma\mathrm{I}}^{(2)} A_{\mathrm{II}}^{(2)^{-1}} f_I^{(2)} \right) \end{aligned} \tag{9}$$

Once u_Γ is found by solving (7), the internal components can be found by using (5):

$$u_I^{(i)} = A_{\mathrm{II}}^{(i)^{-1}} \left(f_I^{(1)} - A_{\mathrm{I}\Gamma}^{(i)} u_\Gamma \right) \tag{10}$$

Equations (8) and (9) can be extended to a generic case, where there are more than two domains.

In the method presented in this paper, equation (7) is solved by using a typical direct solver, such as Gaussian Elimination. An AMG solver, based on the Generic Approximate Banded Inverse (GenAbI) algorithm [13], is used in order to solve the equations (10) for each domain.

III. The AMG Solver

In this section, we present the AMG method [13],[18] used for the solution of the systems that arise from the Schur complement method described in the previous section.

The key part of AMG's setup phase is the coarse-grid selection, which is the process of creating the degrees of freedom of a coarse-level problem. The goal of coarse-grid selection is to determine the sets C and F of coarse-grid and fine-grid points respectively, as well as a small set $C_i \subset C$ of interpolating points for each fine-grid point [20],[24]. This is called a C/F splitting, where C-points are variables that exist on both the fine and coarse levels and F-points are variables only on the fine level. Interpolation can then be defined as follows:

$$\left(I_{k+1}^k u^{k+1} \right)_i = \begin{cases} u_i^{k+1}, & i \in C \\ \sum_{j \in C_i} w_{ij} u_j^{k+1}, & i \in F \end{cases} \tag{11}$$

An important concept in the coarse grid selection is that of strong influence and strong dependence. It is highly likely that not all matrix coefficients are equally important to the selection of the coarse grids and thus only those that are "large enough" should be considered [18],[27].

It should be stated that point i depends on point j if a_{ij} is sufficiently large, denoting that in order to satisfy the i-th equation of the system, the node u_i is affected more from node u_j than other neighbouring nodes. We can then define the set of dependencies for point i as:

$$S_i = \left\{ j \neq i, -a_{ij} \geq \theta \max_{k \neq i} \left(-a_{ik} \right) \right\} \tag{12}$$

where θ is called strength threshold and is important for its influence on stencil size and convergence [27]. A typical value for θ is 0.25. The set of influences for point i can be defined as the transpose of the dependencies set.

The concept of strong influence/dependence in conjunction with the following two heuristics is vital to creating a valid coarse grid [4],[20],[27]:

- ***H1***: For each point j that strongly influences a fine-grid point i, j is either a coarse-grid point or strongly depends on a coarse-grid point that also strongly influences i.
- ***H2***: C is a maximal set with the property that no C-point influences another C-point.

Condition H1 ensures the quality of the interpolation and condition H2 restricts the size of coarser grids.

The coarsening schemes of early AMG methods are based on the Ruge-Stüben (RS) coarsening method [20]. The Ruge-Stüben coarsening is the classical coarse-grid selection algorithm, based on enforcing heuristic H1, while implicitly using heuristic H2 as a guideline. It is a two-pass process, where the first pass selects a maximal independent set guaranteeing that every fine-grid point strongly depends on at least one coarse-grid point. Further details on the two-pass RS coarsening are given in [20].

Since RS coarsening selects only a single C-point in each iteration, its main drawback is its sequential nature. However, RS exhibits optimal scalability and convergence behavior for a variety of problems.

In recent years, there has been significant progress in the development of parallel coarse grid selection schemes, most of which are based on their sequential predecessors, such as CLJP [4] and PMIS [6]. An overview of such coarsening schemes and their performance is presented in [27].

The interpolation formula used for the purposes of this article is direct interpolation, in which the weights w_{ij} are defined as follows:

$$w_{ij} = -\left(\frac{\sum_{k \in N_i} a_{ik}}{\sum_{l \in C_i} a_{il}} \right) \frac{a_{ij}}{a_{ii}} \tag{13}$$

This formula is easy to implement and only requires immediate neighbours of i, however, it generally leads to worse convergence rates compared to interpolation formulas that use extended neighbourhoods [27]. The interpolation operator can now be computed according to (11).

The restriction operator is defined as the transpose of the interpolation operator through the Galerkin condition [19]. Thus the next coarser level matrix is defined as the triple matrix product of the restriction operator, the finer grid matrix and the interpolation operator. When the current grid is considered "coarse enough", then the setup phase terminates and AMG proceeds to the solution phase.

The cycle strategy is an essential component of any multigrid algorithm and refers to the sequence in which the various grids are visited and the respective coarse grid corrections are obtained. The common cycle strategy is the V–cycle algorithm and is shown in Fig. 2 [18].

The solution can be achieved by successive applications of the cycle according to arbitrary termination criterion. The proposed multigrid scheme descends to the coarsest level and then, the multigrid method ascends to the finer levels and corrects the respective solution [18].

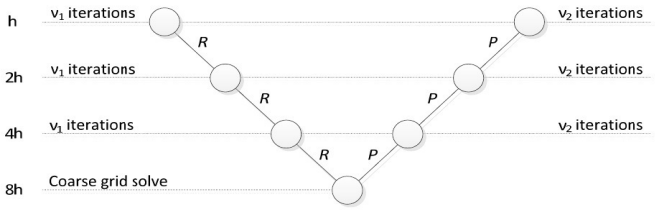


Fig. 2 The V–Cycle with finest grid Ω^h, coarsest grid Ω^{8h}, and ν_1,ν_2 pre–smoothing and post–smoothing steps, respectively.

An important component in multigrid methodology is the stationary iterative solver, namely smoother, that can be described by the following recurrence relation [13],[18]:

$$x_{(i+1)}^{k}=x_{(i)}^{k}+M^{k}r^{k},\mathrm{r}^{k}=f^{k}-A^{k}x_{(i)}^{k} \qquad (14)$$

where f^k,A^k are the right-hand side and coefficient matrix (at the k-th coarse level) and $x_{(i)}^{k}$ is the solution vector at the i-th iterative step. Equation (14) describes a family of stationary iterative methods, according to the choice of the M^k matrix.

Generic approximate banded inverses in conjunction with the general iterative method (14) can be used as smoothers for multigrid schemes, by choosing $M^{k}=\left(M^{k}\right)^{\delta \mathrm{l}}$, where $\left(M^{k}\right)^{\delta \mathrm{l}}$ is a class of approximate inverses. The class of smoothing methods proposed can be described as follows:

$$x_{(i+1)}^{k}=x_{(i)}^{k}+\omega\left(M^{k}\right)_{r}^{\delta \mathrm{l}}\left(f^{k}-A^{k}x_{(i)}^{k}\right) \qquad (15)$$

where ω is the damping parameter with $0<\omega\leq 1$ [13].

Let us assume the incomplete LU factorization, such that

$$A\approx \mathrm{LU}+R \qquad (16)$$

where L and U are upper and lower matrices of the same nonzero structure as the lower and upper parts of coefficient matrix A respectively and R is some error matrix. This is the so-called incomplete LU factorization with zero fill-in, or more commonly ILU(0) factorization [21].

Let $M^{\delta \mathrm{l}}=\left(\mu_{\mathrm{i,j}}\right),\mathrm{i}\in[1,\mathrm{n}],\mathrm{j}\in[1-\delta \mathrm{l}+1,\mathrm{i}+\delta \mathrm{l}-1]$ be the generic approximate banded inverse of the coefficient matrix A. The elements of a class of banded forms of the generic approximate inverse, by retaining δl and δl-1 elements in the lower and upper parts, can be computed by solving recursively the following systems [11],[13],[17]:

$$M^{\delta \mathrm{l}}L=U^{-1} \text{ and } \mathrm{UM}^{\delta \mathrm{l}}=L^{-1} \qquad (17)$$

Then, the elements of the approximate inverse are computed by the Generic Approximate Banded Inverse (GenAbI) algorithm [13].

Specific information on the smoothing and approximation property for the GenAbI algorithm, as well as the use of the DOUR scheme [15] in order to dynamically compute the relaxation parameter ω for the smoothing scheme can be found in [12],[13],[18].

IV. Numerical Results

In this section we examine the effectiveness of the new proposed scheme, namely Domain Decomposition-Algebraic Multigrid method in conjunction with the Generic Approximate Banded Inverse matrix.

The convergence factor depends on the required number of iterations for convergence [2],[3],[26]. The convergence factor with respect to the 2-norm is defined as:

$$q=\sqrt[m]{\|r_m\|_2/\|r_0\|_2} \qquad (18)$$

where r_m is the residual vector at the m-th iteration. The termination criterion for the AMG solver is $\|r_m\|_2<10^{-10}\|r_0\|_2$ and the numbering of the grid is lexicographical. The maximum number of iterations was set to 200 iterations.

The strength threshold θ was set to 0.25 for the AMG solver. The values for the pre-smoothing and post-smoothing steps were set to ν_1, ν_2=2. The coarsest level, with its maximum amount of variables allowed set to 15, was solved using the BiCGSTAB method.

The coarsening scheme used in the first problem was the CLJP algorithm, while for the second problem the PMIS coarsening technique was utilized. It should also be mentioned that for the second problem, the AMG solver was modified to use the V-cycle as a preconditioner for the BiCGSTAB method in order to accelerate convergence.

Model Problem I: The model problem to be solved with the proposed scheme is the Poisson equation:

$$-\Delta \mathrm{u}(\mathrm{x},\mathrm{y})=f \qquad (19)$$

$$u(\mathrm{x},\mathrm{y})=0 \quad (\mathrm{x},\mathrm{y})\in\partial\Omega \qquad (19.a)$$

discretized with the finite element method, where f(x,y)=1, Ω is $[0,1]\times[0,2]$ and $\partial\Omega$ denotes the boundary of Ω. The domain Ω was split into 8 subdomains, as shown in Fig. 3.

In Table I, the convergence factors and convergence behavior are presented for various values of the order of the linear system n and "retention" parameter δl of the generic approximate banded inverse. Additionally, the convergence factors and convergence behavior for the same values of n and δl using AMG as a standalone solver without domain decomposition techniques are given. In Table II, the performance, in seconds, of both the DD/AMG and standalone AMG methods is showcased for various values of the order of the linear system n and "retention" parameter δl of the generic approximate banded inverse.

Model Problem II: Let us consider the following elliptic PDE:

$$-\Delta \mathrm{u}(\mathrm{x},\mathrm{y})+\alpha(\mathrm{x},\mathrm{y})u(\mathrm{x},\mathrm{y})=f(\mathrm{x},\mathrm{y}) \qquad (20)$$

$$u(\mathrm{x},\mathrm{y})=0 \quad (\mathrm{x},\mathrm{y})\in\partial\Omega \qquad (20.a)$$

discretized with the finite element method, where α(x,y)=-40x+70y, f(x,y)=19x-44y^2, Ω is the unit square and $\partial\Omega$ denotes the boundary of Ω. The domain Ω was split into 16 subdomains, as shown in Fig. 4.

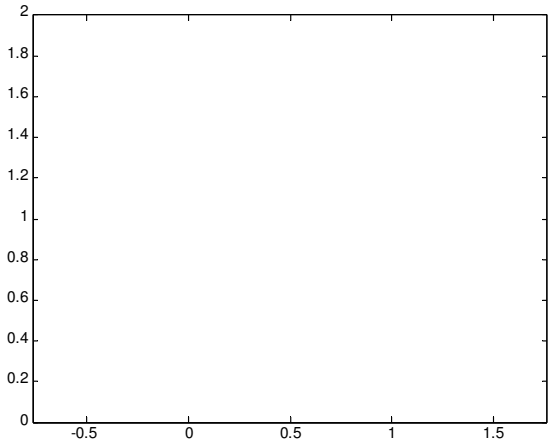


Fig. 3 Domain Ω split into 8 subdomains for model problem I.

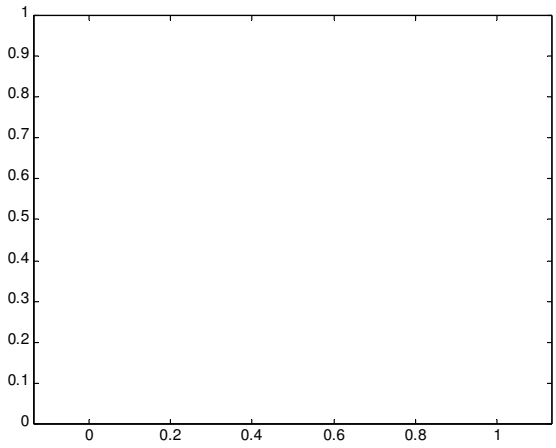


Fig. 4 Domain Ω split into 16 subdomains for model problem II.

In Table III, convergence behavior for various values of the order of the linear system n and "retention" parameter δl of the generic approximate banded inverse is presented. Additionally, convergence behavior for the same values of n and δl using AMG-BiCGSTAB as a standalone solver without domain decomposition techniques is given. In Table IV, the performance, in seconds, of both the DD/AMG-BiCGSTAB and standalone AMG-BiCGSTAB methods is showcased for various values of the order of the linear system n and "retention" parameter δl of the generic approximate banded inverse.

It should be noted that, for both problems, the convergence and performance results for the DD/AMG method are the average of the results taken from all domains, since the domains were not identical.

As already expected considering past results [18], increasing the value of the "retention" parameter δl leads to improved convergence behavior. Additionally, we notice that solving the linear systems arising from the subdomains is significantly more efficient, both in performance and convergence.

One of the drawbacks of AMG methods is the computational work added by the setup phase in addition to the solution phase. The resulting smaller linear systems from the domain decomposition method significantly reduce the workload for both phases. Considering that these systems can be solved in parallel since domain decomposition methods are well suited for parallel computing, the combination of domain decomposition with AMG can be very efficient.

Finally, it should be stated that the effectiveness and applicability of the new proposed method will be shown when applied to more general problems, such as quasilinear boundary-value problems or convection-diffusion problems.

V. Conclusions

A Schur complement domain decomposition method, utilizing an AMG solver, based on generic approximate banded inverse matrices, for solving the resulting subdomain systems was presented. The use of domain decomposition techniques was proven to be effective by improving the performance and convergence behavior of Algebraic Multigrid method. Since domain decomposition methods lead to smaller linear systems, arising from each subdomain, the load for both the setup and solution phase of the AMG solver is significantly reduced.

References

[1] A. Brandt, S.F. McCormick and J.W. Ruge, "Algebraic multigrid (AMG) for sparse matrix equations", in *Sparsity and Its Applications*, D.J. Evans (ed), Cambridge University Press, 1984.

[2] L.W. Briggs, V. Henson and S.F. McCormick, *A multigrid tutorial*, SIAM, 2000.

[3] O. Bröker, M.J. Grote, C. Mayer and A. Reusken, "Robust parallel smoothing for multigrid via sparse approximate inverses", *SIAM Journal on Scientific Computing*, vol. 23(4), pp. 1396–1417, 2001.

[4] A. Cleary, R. Falgout, V. Henson and J. Jones, "Coarse-grid selection for parallel algebraic multigrid", in *Proc. Fifth International Symposium on Solving Irregularly Structured Problems in Parallel*, Vol. 1457, *Lecture Notes in Computer Science*, Springer-Verlag, pp. 104–115, 1998.

[5] T.F. Chan and T.P. Mathew, "Domain decomposition algorithms", In *Acta Numerica 1994*, Cambridge University Press, pp. 61-143,1994.

[6] H. De Sterck, U. Yang and J. Heys, "Reducing complexity in parallel algebraic multigrid preconditioners", *SIAM J. Mat. Anal. and Appl.*, vol. 27, pp. 1019-1039, 2006.

[7] C.R. Dohrmann, "A preconditioner for substructuring based on constrained energy minimization", *SIAM J. Sci. Comput*., vol. 25, pp. 246-258, 2003.

[8] C. Farhat, M. Lesoinne, P. LeTallec, K. Pierson, and D. Rixen, "FETI-DP: a dual-primal unified FETI method – part I: A faster alternative to the two-level FETI method", *Int. J. Numer. Meth. Engng,* vol. 50, pp. 1523-1544, 2001.

[9] C. Farhat, M. Lesoinne, and K. Pierson, "A scalable dual-primal domain decomposition method", *Numer. Linear Algebra Appl*., vol. 7, pp. 687-714, 2000.

[10] C. Farhat and F.X. Roux, "A method of finite element tearing and interconnecting and its parallel solution algorithm", *Int. J. Numer. Meth. Engng,* vol. 32, pp. 1205-1227, 1991.

[11] G.A. Gravvanis, Explicit Approximate Inverse Preconditioning Techniques, *Archives of Computational Methods in Engineering,* vol. 9(4), pp. 371-402, 2002.

[12] G.A. Gravvanis, C.K. Filelis-Papadopoulos and P.I. Matskanidis, "A survey on the parallelization issues of geometric and algebraic multigrid methods based on generic banded approximate inverses", in Computational Technology Reviews vol. 7, B.H.V. Topping and P. Ivanyi (eds), pp. 65-98, Saxe-Coburg Publications, 2013.

[13] G.A. Gravvanis, C. K. Filelis-Papadopoulos and P. I. Matskanidis, "Algebraic multigrid methods based on Generic Approximate Matrix Techniques", Report TR/ECE/ASC-AMA/2012/13, submitted.
[14] A. Greenbaum, *Iterative methods for solving linear systems*, SIAM, Philadelphia, PA, USA, 1997.
[15] R. Haelterman, J. Viederndeels and D. Van Heule, "Non–stationary two–stage relaxation based on the principle of aggregation multi-grid", *Computational Fluid Dynamics 2006*, Part 3, Springer, pp. 243–248, 2009.
[16] A. Klawonn, O.B. Widlund, and M. Dryja, "Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients", *SIAM J. Numer. Anal.*, vol. 40, pp.159-179, 2002.
[17] E.A. Lipitakis and D.J. Evans, "Explicit semi-direct methods based on approximate inverse matrix techniques for solving boundary-value problems on parallel processors", *Math. and Computers in Simulation*, vol. 29, pp. 1-17, 1987.
[18] P.I. Matskanidis and G.A. Gravvanis, "On the algebraic multigrid method based on generic approximate banded inverses", in *Proc. 2012 Panhellenic Conference on Informatics (PCI 2012)*, pp. 211-216, IEEE Computer Society, 2012.
[19] S.F. McCormick, "Multigrid methods for variational problems: general theory for the V-cycle", *SIAM J. Numer. Anal.*, vol. 22, pp. 634-643, 1985.
[20] J.W. Ruge and K. Stuben, "Algebraic multigrid (AMG)", in *Multigrid Methods*, S.F. McCormick (ed), vol. 3, Frontiers in Applied Mathematics, SIAM, pp. 73–130, 1987.
[21] Y. Saad, *Iterative methods for sparse linear systems*, PWS Publishing, 1996.
[22] H.A. Schwarz, "Über eine grenzübergang durch alternirendes verfahren, Gesammelete Mathematische Abhandlungen", Springer-Verlag, 2, pp. 133-143, 1980. First published in *Viertel-jaahrsschift der Naturforschenden Gesellschaft* in Zürich, vol.15, pp. 272-286, 1870.
[23] B.F. Smith, P.E. Bjørstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, 1996.
[24] K. Stuben, "Algebraic multigrid (AMG): An introduction with applications", in *Multigrid*, U. Trottenberg, C. Oosterlee, and A. Schuller (eds), Academic Press, 2001.
[25] A. Toselli and O.B. Widlund, *Domain Decomposition methods – Algorithms and Theory*, Springer, 2005.
[26] U. Trottenberg, C.W. Osterlee and A. Schuller, *Multigrid*, Academic Press, 2000.
[27] U.M. Yang, "Parallel Algebraic Multigrid Methods - High Performance Preconditioners", in *Numerical Solution of Partial Differential Equations on Parallel Computers*, A.M. Bruaset and A. Tveito (eds), vol. 51, Springer-Verlag, pp. 209-236, 2006.

TABLE I.
CONVERGENCE BEHAVIOUR OF THE DD/AMG METHOD FOR MODEL PROBLEM I

		n=24353		**n=96833**		**n=386177**	
Method	**δl**	**q**	**its**	**q**	**its**	**q**	**its**
DD/AMG	**1**	0.2152	15	0.2556	17	0.3074	20
	2	0.2102	15	0.2467	17	0.3024	20
	50	0.1946	14	0.2374	16	0.2873	19
AMG	**1**	0.3483	22	0.4137	27	0.4881	33
	2	0.3427	22	0.4090	26	0.4839	32
	50	0.3142	20	0.3794	24	0.4667	30

TABLE II.
PERFORMANCE, IN SECONDS, OF THE DD/AMG METHOD FOR MODEL PROBLEM I

		n=24353	**n=96833**	**n=386177**
Method	**δl**	**time**	**time**	**time**
DD/AMG	**1**	0.0971	0.4499	2.4231
	2	0.0990	0.4653	2.4991
	50	0.2994	1.3975	7.1642
AMG	**1**	1.1680	6.025	33.2329
	2	1.1954	6.2288	33.7139
	50	3.339	15.4747	87.9342

TABLE III.
CONVERGENCE BEHAVIOUR OF THE DD/AMG-BICGSTAB METHOD FOR MODEL PROBLEM II

		n=11457	**n=45441**	**n=180993**
Method	**δl**	**its**	**its**	**its**
DD/AMG-BiCGSTAB	**1**	10	11	14
	2	9	10	13
	50	7	9	11
AMG-BiCGSTAB	**1**	15	21	34
	2	13	17	27
	50	12	15	23

TABLE IV.
PERFORMANCE, IN SECONDS, OF THE DD/AMG-BICGSTAB METHOD FOR MODEL PROBLEM II

		n=11457	**n=45441**	**n=180993**
Method	**δl**	**time**	**time**	**time**
DD/AMG-BiCGSTAB	**1**	0.0174	0.0704	0.3441
	2	0.0182	0.0706	0.3618
	50	0.0399	0.1962	1.0097
AMG-BiCGSTAB	**1**	0.3936	2.2724	17.0591
	2	0.3854	2.003	14.9678
	50	1.2891	5.5477	32.5779

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 495–498

3D Non-Local Means denoising via multi-GPU

Giuseppe Palma,
Marco Comerci
and Bruno Alfano
Inst. of Biostructures and Bioimaging
National Research Council of Italy
Via De Amicis 95, 80145 Naples, Italy
Email: giuseppe.palma@ibb.cnr.it

Salvatore Cuomo,
Pasquale De Michele
and Francesco Piccialli
Dept. of Mathematics and Applications
University of Naples Federico II
Via Cinthia, 80126 Naples, Italy
Email: salvatore.cuomo@unina.it

Pasquale Borrelli
Dept. of Advanced Biomedical Sciences
University of Naples Federico II
Via Pansini 5, 80131 Naples, Italy
Email: pasquale.borrelli@unina.it

***Abstract*—Non-Local Means (NLM) algorithm is widely considered as a state-of-the-art denoising filter in many research fields. High computational complexity led to implementations on Graphic Processor Unit (GPU) architectures, which achieve reasonable running times by filtering, slice-by-slice, 3D datasets with a 2D NLM approach. Here we present a fully 3D NLM implementation on a multi-GPU architecture and suggest its high scalability. The performance results we discuss encourage the coding of further filter improvements and the investigation of a large spectrum of applicative scenarios.**

I. Introduction

IMAGE denoising represents one of the most common tasks of image processing. Several techniques have been developed in the last decades to face the problem of removing noise from images, still preserving the small structures from an excessive blurring [2]. All those schemes share the belief that an improved value of a given image point can be expressed as a function of the image itself; each of them diverges in how the function is defined.

One of the most performing and robust denoising approaches is the non-local means (NLM) filter, introduced in [1]. Since its first appearance, the family of NLM algorithm and implementation variants has enormously grown (just to mention some of the most relevant improvements, see [5], [6], [7], [14], [10], [4], [11]); nevertheless, all of them assume that the restoring function for a given point is a mean of all the image values, largely weighted according to the radiometric similarity between values and only weakly tied to a spatial proximity criterion.

The result is a general-purpose denoising scheme, whose performances are widely accepted to be better with respect to the previous state-of-the-art algorithms, such as the total variation, the wavelet thresholding or the anisotropic filtering [13]. In particular, it has been shown that NLM filter guarantees the homogeneity of flat zones, preserves edges and fine structures, and transforms white noise into white noise, thus avoid the introduction of artifacts and spurious correlated signal [2].

Unsurprisingly, the NLM algorithm is computationally very heavy, and even some fast versions of the scheme are quite demanding on 2D images and almost daunting on 3D datasets. The huge amount of computational demand has been recently addressed by using accelerated hardware, the Graphic Processor Units (GPUs) in particular.

In 3D datasets, *e.g.* in the context of the Magnetic Resonance Imaging (MRI), the use of fully 3D filters is more appropriate than a 2D-based slice-by-slice filtering approach to exploit all the information contained in the image.

To the best of our knowledge, although there are several 2D GPU-based NLM versions ([9], [3], [8]), the 3D version of NLM filter has been poorly investigated in terms of both implementation and performance on GPUs.

In this paper, we present GPU and Multi-GPU versions of the 3D NLM filter based on Compute Unified Device Architecture (CUDA) [12]. We report the performance of the implementation for different 3D synthetic and real datasets. The parallelization of the filter via GPUs gives clinically-feasible MRI denoising execution times.

The plan of the paper is as follows. In §II we briefly describe the NLM algorithm. To follow, in §III we provide the implementation details. In §IV we present and discuss the results. Finally, in §V we draw conclusions and future works.

II. Theory

A. General description

An N-D image X can be considered as a real function $X : \mathbb{R}^N \to \mathbb{R}$ with a bounded support $\Omega \subset \mathbb{R}^N$. The NLM filter [1] is a class of endomorphisms of the image space, identified by 2 parameters (a and h), that acts as follows:

$$[\mathrm{NLM}_{a,h}(X)](\vec{x}) = Y(\vec{x}) = \frac{\int_\Omega \exp\left[-\frac{d_a^2(\vec{x},\vec{y})}{h^2}\right] X(\vec{y}) d\vec{y}}{\int_\Omega \exp\left[-\frac{d_a^2(\vec{x},\vec{y})}{h^2}\right] d\vec{y}}\,, \quad (1)$$

where

$$d_a^2(\vec{x},\vec{y}) \equiv \int_{\mathbb{R}^N} \left|X(\vec{x}+\vec{t}) - X(\vec{y}+\vec{t})\right|^2 \cdot \frac{\exp -\frac{\|\vec{t}\|^2}{2a^2}}{(2\pi)^{n/2}\cdot a} d\vec{t}. \quad (2)$$

The intensity of a given point of the new image is a mean of the intensities of the original image, according to a weight function that disregards any explicit criterion of spatial proximity and only considers a measure (ruled by h) of self-similarity between windows of radius a centered on each point (radiometric proximity).

If the image is defined on a discrete, regular grid $\left\{\vec{x}_i \,\middle|\, 1 \le i \le \prod_{l=1}^N L_l\right\}$,

$$X(\vec{x}) = \sum_i X_i \delta(\vec{x} - \vec{x}_i)\,, \quad (3)$$

978-83-60810-53-8/$25.00

from Eqns. 1–2 it follows that the filtered dataset is given by

$$Y_i = \frac{\sum_j \exp\left[-\frac{d_a^2(\vec{x}_i,\vec{x}_j)}{h^2}\right] X_j}{\sum_j \exp\left[-\frac{d_a^2(\vec{x}_i,\vec{x}_j)}{h^2}\right]}, \tag{4}$$

$$d_a^2(\vec{x}_i,\vec{x}_j) = \sum_k \left| X(\vec{x}_i+\vec{\Delta}_k) - X(\vec{x}_j+\vec{\Delta}_k)\right|^2 \cdot \frac{\exp -\frac{\|\vec{\Delta}_k\|^2}{2a^2}}{(2\pi)^{n/2}\cdot a}. \tag{5}$$

Moreover, from Eqns. 4–5 it follows that the complexity of the filter is $O\left(\prod_{l=1}^{N} L_l^3\right)$.

B. Actual algorithm

Both computational issues and the convenience to introduce a geometric proximity criterion in addition to the pure radiometric distance measure led to a change in the original version of the NLM filter [7].

Therefore, given a search radius M, for each voxel i located at $\vec{x}_i$ we define a search box V_i as

$$V_i \equiv \left\{ \vec{x}_j \in \Omega \,|\, \|\vec{x}_j - \vec{x_i}\|_\infty < M \right\}. \tag{6}$$

The search box associated with the i-th voxel defines the ensemble of voxels whose intensities will be available in the following for restoring (denoising) of the intensity $X(\vec{x}_i)$, thus reducing the search freedom of Eqn. 4 (in that case, $V_i \equiv \Omega$). The authors of ([7], [10]) suggest that a good choice for M should guarantee the cardinality of the search box, $|V_i|$, to be of the order of 10^3.

Analogously, given a similarity radius d, for each voxel $\vec{x}_j$ within a given search box V_i, we can define a similarity box

$${}_jB_i \equiv \left\{ \vec{x}_k \in \Omega \,|\, \|\vec{x}_k - \vec{x_j}\|_\infty < d \right\}. \tag{7}$$

In this case, d plays the role of a in Eqn. 4, provided that the original smooth Gaussian kernel is replaced by a binary cut-off; a good choice for d should guarantee $|{}_jB_i| \sim 30$ ([7], [10]).

Finally, the denoised image is

$$Y_i = \frac{\sum_{\vec{x}_j \in V_i} \exp\left[-\frac{\|{}_jB_i - {}_iB_i\|_2^2}{h^2}\right] X_j}{\sum_{\vec{x}_j \in V_i} \exp\left[-\frac{\|{}_jB_i - {}_iB_i\|_2^2}{h^2}\right]}, \tag{8}$$

whence it results that the algorithm complexity is $O\left(|V_i|\cdot|{}_jB_i|\cdot\prod_{l=1}^{N} L_l\right)$.

The filter strength, which is determined by h, can be automatically tuned to obtain an optimized denoising, independently from the search radius M and the standard deviation of noise σ:

$$h^2 = 2\beta\sigma^2 |V_i| \tag{9}$$

($\beta \sim 1$ is an adimensional constant to be manually tuned).

III. Implementation

General Purpose computation on Graphics Processing Units (GPGPU) is the use of GPUs to perform highly parallelizable computations that would normally be handled by CPU devices. Programming with GPUs requires both a deep understanding of the underlying computing architecture and a massive rethinking of existing CPU based algorithms.

A. Architecture

We implement the 3D NLM filter on the NVIDIA parallel computing architecture, which consists in a set of cores, or Scalar Processors (SPs), performing simple mathematical operations.

In the NVIDIA Fermi architecture, each SM has scheduler and dispatch units, execution units and a configurable memory of 64KB, which consists of a register file, an internal shared memory and an L1 cache. This memory is configurable in 16KB (or 48KB) for shared memory and 48KB (or 16KB) for L1 cache.

B. Mapping the algoritm on GPU

The Algorithm 1 is the pseudo-code of the NLM filter.

Algorithm 1 Pseudo-code of the NLM algorithm

```
1: for each voxel (i1, i2, i3) of the 3D image to be filtered do
2:     Initialize the cumulative sum of weights and the restored value to 0;
3:     for each voxel (j1, j2, j3) of the search window V_(i1,i2,i3) do
4:         for each voxel (k1, k2, k3) of the similarity window
           (j1,j2,j3)B_(i1,i2,i3) do
5:             Cumulate squared Euclidean distance;
6:         end for
7:         Calculate and cumulate the weight of the voxel in search window;
8:         Cumulate the restored value;
9:     end for
10:    Normalize restored value to the sum of the weights;
11: end for
```

In details, the statement at line 1 represents a nested iteration structure. In our GPU version, the loops on line 1 and line 3 in the Algorithm 1 are logically mapped onto the grid of thread blocks defined by means of the CUDA framework.

A first implementation in CUDA is presented in the Algorithm 2. Moreover, in order to make this algorithm compatible

Algorithm 2 CUDA code of NLM algorithm

```
1: int const i_1 = threadIdx.x +
     blockDim.x*blockIdx.x;
2: int const i_2 = threadIdx.y +
     blockDim.y*blockIdx.y;
3:   /* local statements */
4:   if ((i_1 ≥ 0) && (i_1 < X_Dim) &&     (i_2 ≥ 0)
   && (i_2 < Y_Dim)) {
5:     for(i_3=0; i_3 < Dim_Z; i_3++) {
6:       /* do something on img[i_1 +
           i_2*X_Dim + i_3*X_Dim*Y_Dim] */ }}
```

with multi-GPU architectures, we introduce some improvements. The number of GPU devices is returned by means of a CUDA library function and stored in the variable `n_gpus`. Then, the third dimension of the image is "splitted" between the available GPUs, setting the first (`start_k`) and the last (`end_k`) slices that each GPU has to manage. We report a sketch of the GPU implementation in the Algorithm 3.

In order to explore different types of data access, we test several configurations, both mono- and bi-dimensional, for the thread block size in which each slice is divided. Each thread processes sequentially the voxels along the third dimension. The workload is divided along the third dimension for multi-GPU configurations. Inside each GPU the workload

Algorithm 3 CUDA MULTI-GPU code of NLM algorithm

```
1: int const i_1 = threadIdx.x +
     blockDim.x*blockIdx.x;
2: int const i_2 = threadIdx.y +
     blockDim.y*blockIdx.y;
3:   /* split the image ``img'' between the
   ``n_gpus'' GPUs: each GPU works on the section
   of the image ``my_img'' */
4:   /* local statements */
5:   for(i_3 = 0; i_3 < Z_Dim/n_gpus;     i_3++) {
6:       /* do something on my_img[i_1 + i_2*X_Dim +
   i_3*X_Dim*Y_Dim/n_gpus] */
7:     } }
```

is divided along the first and second dimensions, in strips (mono-dimensional configurations) and tiles (bi-dimensional configurations) of threads. Strip or tile is allowed to cover entirely or only partially the slice grid.

We test also the impact of L1-cache on performance, using the binary L1-prefer setting, which allows to choose between two possible configurations: 48KB of shared memory and 16KB of L1-cache (no L1-prefer), or 16KB of shared memory and 48KB of L1-cache (L1-prefer).

The computing system is equipped with 2 Intel Xeon CPU E5620 (2.4 GHz) and an NVIDIA TESLA S2050 card. This device consists of 4 GPGPU units, each of which with 3GB of RAM memory and 448 cores at 1.15 GHz. The numerical code is implemented by using the single precision arithmetic. The CPU system is equipped with an Intel core i5-2500S (2.7-3.7 GHz).

IV. Results and Discussions

A. Consistency

As we aim to produce a strictly equivalent GPU implementation of the sequential NLM algorithm, we check the implementation consistency by comparing voxel-by-voxel the images obtained by one-core-CPU and GPU denoising. In Fig. 1 we show the 3D NLM filtering result on a real 3D knee MRI dataset. The difference between the GPU and CPU restored images falls within machine precision order of magnitude which are likely to be due to the arithmetic logic unit precision.

B. Performance

In order to investigate cache size impact on the execution time, we perform several test runs varying L1-prefer switch. Results are shown in Table I. L1-prefer choice gives a benefit on lager dataset, with a performance improvement ranging from fraction of percent in the smallest dataset to some 5% in the largest ones. These results suggest that the L1 miss rate, is low enough to have high performance processing even with old generation cards having small amount of cache.

The strip or tile thread division influences the performance of the filter in terms of computing time due to the different type of data access. Experimental results prove that optimal configuration is given by the strip subdivision. In Table I we report running times of (128,1,1) configuration on 2-GPU.

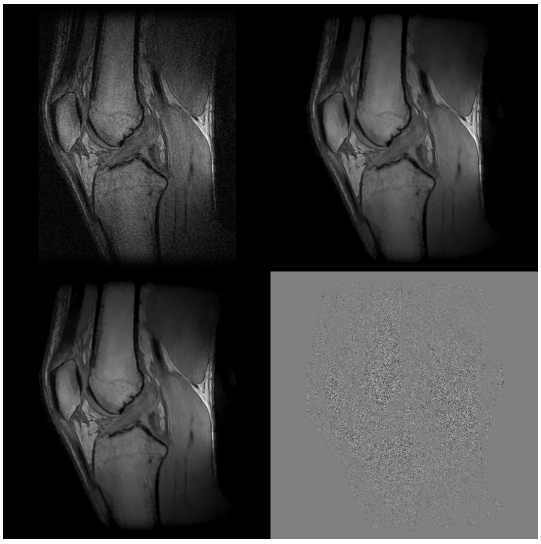


Fig. 1. From left to right and from top to bottom, the frames show a central slice of the original dataset, the GPU restored image, the CPU restored image and the difference between CPU and GPU filtered images (enhanced by a scaling factor of 10^6), respectively.

TABLE I
L1-prefer switch influence on execution times for (128,1,1) block size configuration and 3D random datasets.

	Cache configuration	
	L1-prefer	no L1-prefer
64x64x64	8.04	8.55
128x128x64	9.43	9.42
128x128x128	11.2	11.2
256x256x128	19.9	20.2
256x256x256	28.3	29.0
512x512x128	39.5	40.4
512x512x256	71.7	75.0
512x512x512	138	148

In Table II we report a comparison between running times of CPU, single GPU and multi-GPU implementation of 3D NLM filter with various thread block size. Reported running times include the overall data transfer between CPU and GPU and viceversa, which even for the biggest datasets appears negligible. Speed-up values suggest that the bigger the dataset to be filtered, the better the scalability of the implementation, which, for datasets size typical of MRI clinical practice, is close to be ideal. Moreover, the optimal thread size seems to be strips of thread between 128 and 256 elements. This result is consistent with NVIDIA guidelines [12]. Finally, on large datasets strip configuration should be prefered to tile configuration of the same size because more sequential memory access of the former. In Table III, we investigate the behavior of running times against the search ($|V_i|$) and similarity ($|_j B_i|$) window cardinalities. We note an high and almost constant speed-up among the various experiments, which makes feasible large window filter testing in a reasonable time.

Finally, in Figure 2 we outline the CPU, single GPU and multi-GPU GFlops for variable dataset sizes. It should be remarked that we are able to exploit up to 43.5% of single precision floating point peak performance of the GPU.

TABLE II
EXECUTION TIMES AND SPEED-UP VALUES FOR SEVERAL BLOCK SIZE CONFIGURATIONS AND 3D RANDOM DATASETS. SEARCH AND SIMILARITY WINDOWS HAVE BEEN SET ACCORDING TO $|V_i| = 11^3$ AND $|_jB_i| = 3^3$.

	Execution time/**Speed-up**								
	Single GPU				Multi-GPU				CPU
Dataset size	(16,16,1)	(128,1,1)	(256,1,1)	(512,1,1)	(16,16,1)	(128,1,1)	(256,1,1)	(512,1,1)	
64^3	5.08/**4.47**	5.73/**3.96**	5.63/**4.03**	6.94/**3.27**	11.7/**1.94**	8.04/**2.82**	8.53/**2.66**	12.6/**1.80**	22.7
128^2x64	6.48/**13.6**	7.30/**12.1**	7.08/**12.5**	10.2/**8.61**	8.97/**9.83**	9.43/**9.35**	9.31/**9.47**	10.8/**8.13**	88.2
128^3	9.06/**19.3**	10.8/**16.2**	10.4/**16.9**	16.7/**10.5**	10.3/**17.0**	11.2/**15.7**	10.9/**16.0**	14.0/**12.5**	175
256^2x128	22.1/**31.8**	24.3/**28.9**	25.0/**28.0**	31.0/**22.6**	16.9/**41.6**	19.9/**35.2**	18.6/**37.8**	21.1/**33.2**	702
256^3	40.5/**34.6**	44.7/**31.3**	47.2/**29.6**	59.4/**23.6**	26.0/**53.8**	28.3/**49.4**	29.5/**47.5**	35.1/**39.9**	1400
512^2x128	68.8/**41.0**	67.0/**42.1**	67.1/**42.0**	72.7/**38.8**	40.5/**69.6**	39.5/**71.4**	40.0/**70.5**	42.4/**66.5**	2820
512^2x256	136/**41.2**	132/**42.6**	131/**42.8**	142/**39.5**	73.8/**76.3**	71.7/**78.5**	72.0/**78.2**	77.5/**72.6**	5630
512^3	277/**40.7**	268/**42.2**	264/**42.7**	285/**39.6**	142/**79.3**	138/**82.0**	137/**82.4**	148/**76.2**	11300

TABLE III
EXECUTION TIMES AND SPEED-UP VALUES FOR A 3D RANDOM DATASET (SIZE = $512 \times 512 \times 128$) FOR SEVERAL WINDOW CONFIGURATIONS.

	Execution time / **Speed-up**								
	Single GPU				Multi-GPU				CPU
$(\|V_i\|, \|_jB_i\|)$	(16,16,1)	(128,1,1)	(256,1,1)	(512,1,1)	(16,16,1)	(128,1,1)	(256,1,1)	(512,1,1)	
$(11^3, 3^3)$	68.8/**41.0**	67.0/**42.1**	67.1/**42.0**	72.7/**38.8**	40.5/**69.6**	39.5/**71.4**	40.0/**70.5**	42.4/**66.5**	2820
$(21^3, 3^3)$	447/**44.4**	434/**45.7**	434/**45.7**	467/**42.4**	228/**87.0**	222/**89.4**	221/**89.6**	239/**82.8**	19800
$(11^3, 5^3)$	235/**34.6**	221/**36.7**	223/**36.6**	255/**31.9**	123/**61.1**	116/**69.9**	117/**69.3**	130/**62.4**	8140
$(21^3, 5^3)$	1650/**35.5**	1510/**38.8**	1520/**38.7**	1820/**32.2**	817/**72.0**	757/**77.6**	764/**77.0**	906/**64.8**	58800

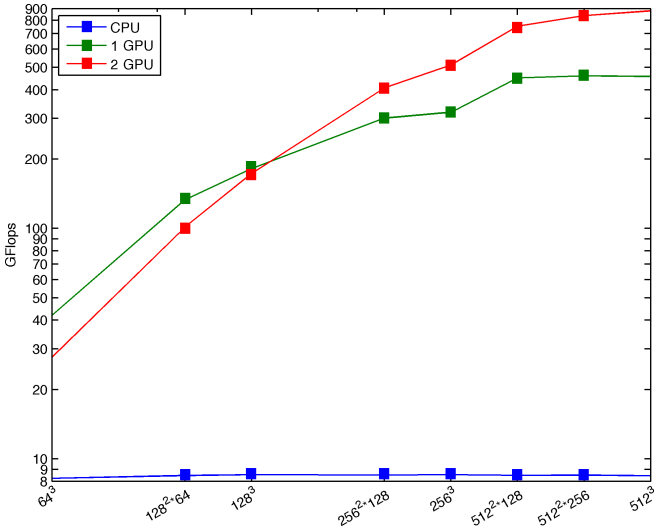


Fig. 2. Outline of the CPU, single GPU and multi GPU GFlops values for different dataset sizes. Please note the logarithmic scale of the axes.

V. CONCLUSIONS

NLM filter is a state-of-the-art denoising algorithm. However, the huge amount of computational load prevents the large-scale diffusion of its most common implementations.

To the best of our knowledge, in this paper we presented the first multi-GPU implementation of a fully 3D NLM filter. We analyzed several configurations of thread block organization and data access, thus identifying a set of optimal settings that guarantee high performance results for a wide spectrum of application scenarios. The reduction of running times shows that scalability is close to ideal one for most common dataset sizes, *e.g.* those typical of MRI clinical pratice. Speed-up high values encourage the exploration of more sophisticated algorithm variants, and reduce the gap between the previous execution times and acceptable performance for real-time scenarios.

REFERENCES

[1] A. Buades, B. Coll, J.M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul. 4, 490-530 (2005).
[2] A. Buades, B. Coll, J.M. Morel, *Image Denoising Methods. A New Nonlocal Principle*, SIAM Review. 52, 113-147 (2010).
[3] F.P.X. De Fontes, G.A. Barroso, P. Coupé, P. Hellier, *Real time ultrasound image denoising*, J. Real-Time Image Process 6, 15-22 (2011).
[4] K. Dabov, A. Foi, V. Katkovnik K. Egiazarian, *Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering*, Image Processing, IEEE Transactions on, 2080-2095 (2007).
[5] M. Ebrahimi, E. Vrscay, *Examining the role of scale in the context of the non-local-means filter*, in Image Analysis and Recognition, LNCS. vol. 5112, pp. 170-181, Springer-Verlag, Berlin (2008).
[6] H. Xu, J. Xu, F. Wu, *On the biased estimation of nonlocal means filter*, Proceedings of IEEE IC on Multimedia and Expo, pp. 1149-1152 (2008).
[7] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, C. Barillot, *An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images*, IEEE Trans. Med. Imag. 27, 425-441 (2008).
[8] B. Goossens, Q. Luong, J. Aelterman, A. Pizurica and W. Philips, *A GPU-accelerated real-time NLMeans algorithm for denoising color video sequences*, Proceedings of 12th IC on Advanced Concepts for Intelligent Vision System pp. 46-57, 6475 of LNCS, Springer (2010).
[9] K. Huang, D. Zhang, K. Wang *Non-local means denoising algorithm accelerated by GPU*, SPIE Conference Series (2009).
[10] J.V. Manjón, N.A. Thacker, J.J. Lull, G. Garcia-Martí, L. Martí-Bonmatí, M. Robles, *Multicomponent MR Image Denoising*, IJBI, 2009.
[11] M. Maggioni, V. Katkovnik, K. Egiazarian, A. Foi, *Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction*, IEEE Transactions on Image Processing, 119-113 (2013).
[12] *NVIDIA CUDA programming guide* http://developer.download.nvidia.com, Nvidia Technical Report, 2012.
[13] H. Seo, P. Chatterjee, H. Takeda, P. Milanfar, *A comparison of some state of the art image denoising methods*, in Proceedings of the 41st Asilomar Conference on Signals, Systems, and Computers (2007).
[14] N. Wiest-Daesslé, S. Prima, P. Coupé, S. Morrissey, C. Barillot, *Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI*, MICCAI, LNCS. vol. 5242, pp. 171-179, Springer-Verlag, Berlin (2008).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 499–505

Examples of Ramanujan and expander graphs for practical applications

Monika Polak
Institute of Mathematics,
Maria Curie-Sklodowska University,
pl. M. Curie-Sklodowskiej 5,
20-031 Lublin, Poland
Email: monika.katarzyna.polak@gmail.com

Vasyl Ustimenko
Institute of Mathematics,
Maria Curie-Sklodowska University,
pl. M. Curie-Sklodowskiej 5,
20-031 Lublin, Poland
Email: ustymenko_vasyl@yahoo.com

***Abstract*—Expander graphs are highly connected sparse finite graphs. The property of being an expander seems significant in many of these mathematical, computational and physical contexts. Even more, expanders are surprizingly applicably applicable in other computational aspects: in the theory of error corecting codes and the theory of pseudorandomness, which are used in probabilistic algorithms. In this article we present a method to obtain a new examples of families of expanders graphs and some examples of Ramanujan graphs which are the best expanders. We describe properties of obtained graphs in comparison to previously known results. Numerical computations of eigenvalues presented in this paper have been computed with MATLAB.**

I. Introduction

THERE are many different algorithms in everyday life where graphs are used. The development of information technology has allowed various representations of graphs in the memory of a computer. Graph based algorithms are used, in particular, in cryptography, coding theory, car navigation systems, sociology, mobile robotics and even in computer games. Graphs used for different purposes often must have some special properties.

One of the most interesting features of the new graphs is their expansion property. Expander graphs are highly connected sparse finite graphs. This property seems to be very significant. From a practical viewpoint, these graphs resolve an extremal problem in communication network theory. Second, they fuse diverse branches of pure mathematics: number theory, representation theory and algebraic geometry.

Expander graphs are used to efficient error reduction in probabilistic algorithms. A randomized algorithm uses a source of pseudorandom bits. During execution, it takes random choices depending on those random data. However, to collect a reasonable collection of random bits is not an easy task. Algorithms that use the random input to reduce the expected running time or memory usage have a chance of producing an incorrect result. Using expander walks allows to achieve the same error probability, with much fewer random bits. The exact form of the exponential decay in error using expander walks and its dependence on the spectral gap was found by Gillman [6].

Constructions of the best expander graphs with a given regularity and order is not easy and in many cases, it is an open problem. In this article we present a method to obtain a new examples of families of expanders graphs and some examples of Ramanujan graphs which are the best expanders. We describe properties of obtained graphs in comparison to previously known results.

Throughout this paper, only undirected simple graphs without loops or multiple edges are considered. A distance between vertices v_1 and v_2 in the graph is the length of minimal path from v_1 to v_2. A graph is connected if for arbitrary pair of vertices v_1, v_2 there is a path from v_1 to v_2. The length g of the shortest cycle in a graph is called a *girth*, [3]. Bipartite graph is a graph whose vertices set V can be divided into two disjoint subsets V_1 and V_2 such that every edge connects a vertex in V_1 to one in V_2. We refer to bipartite graph $\Gamma(V_1 \cup V_2, E)$ as biregular one if the number of neighbors for vertices from each partition sets are constants s and t (bidegrees). We call a graph regular in the case $s = t$.

By the theorem of Alon and Boppana, large enough members of an infinite family of d-regular graphs with constant d satisfy the inequality $\lambda \geq 2\sqrt{d-1} - o(1)$, where λ is the second largest eigenvalue in absolute value. Ramanujan graphs are d-regular graphs for which the inequality $\lambda \leq 2\sqrt{d-1}$ holds.

We say that a family of regular graphs of bounded degree q of increasing order n *has an expansion constant* c, $c > 0$ if for each subset A of the vertex set X, $|X| = n$ with $|A| \leq n/2$ the inequality $|\partial A| \geq c|A|$ holds. The expansion constant of the family of q-regular graphs can be estimated via upper limit $q - \lambda_n$, $n \to \infty$, where λ_n is the second largest eigenvalue of family representative of order n. It is clear that a family of Ramanujan graphs of bounded degree q has the best expansion constant.

The first explicit expander graph family was constructed by Gregory Margulis in the 1970's via studies of Cayley graphs of large girth [13].

A family of graphs G_n is a family of graphs of increasing girth if $g(G_n)$ goes to infinity with the growth of n.

The *family of graphs of large girth* is an infinite family of

978-83-60810-53-8/$25.00

simple regular graphs Γ_i of degree k_i and order v_i such that

$$g(\Gamma_i) \geq \gamma \log_{k_i} v_i, \tag{1}$$

where c is an independent of i constant (see [1], [2]).

A sparse graph has a small number of edges in comparison to the number of vertices. A simple relationship describing the density of the graph $\Gamma(V, E)$ is

$$D = \frac{2|E|}{|V|(|V|-1)}, \tag{2}$$

where $|E|$ is the number of edges of graph Γ and $|V|$ is the number of vertices. The maximal density is $D = 1$ when a graph is complete and the minimal density is 0 (Coleman & Moré 1983).

One of the very important classes of small world bipartite graphs with additional geometric properties important in this context, is a class of regular generalized m-gon, i.e. regular tactical configurations of diameter m and girth $2m$. For each parameter m, a regular generalized m-gon has degree $q+1$ and order $2(1+q+...+q^{m-1})$, [15].

According to the famous Feit-Higman theorem the regular thick (i.e. degree ≥ 3) generalized m-gons exist only for $m = 3, 4$ and 6, [5]. Thus Generalized Pentagon does not exist, in particular. We have the following properties of generalized polygons:

- the incidence graph of a projective plane $PG(2, q)$ has $|V| = \nu(q+1, 6) = 2(1+q+q^2)$ and $g = 6$,
- the incidence graph of a generalized quadrangle $GQ(q, q)$ has $|V| = \nu(q+1, 8) = 2(1+q+q^2+q^3)$ and $g = 8$,
- the incidence graph of a generalized hexagon $GH(q, q)$ has $|V| = \nu(q+1, 10) = 2(1+q+q^2+q^3+q^4+q^5)$ and $g = 12$.

II. Construction of the families

Described below families of graphs $D(n, \mathbb{F}_q)$ and $W(n, \mathbb{F}_q)$ can be used to obtain the new construction of expander graphs or even Ramanujan graphs.

Let F_q, where q is prime power, be a finite field. $CD(n, q)$ (connected components of $D(n, \mathbb{F}_q)$) and $W(n, \mathbb{F}_q)$ are connected, regular, bipartite families of graphs.

Traditionally in graph theory one subset of vertices in bipartite graphs is denoted by $V_1 = P$ and called a set of points and another one $V_2 = L$ is called a set of lines. Let P and L be two copies of Cartesian power ${\mathbb{F}_q}^n$, where $n \geq 2$ is an integer. Brackets and parenthesis will allow the reader to distinguish points and lines. In this note we concentrate on finite bipartite graphs on the vertex set $P \cup L$, where P and L are two copies of ${\mathbb{F}_q}^n$. If $z \in \mathbb{F}_q^n$, then $(z) \in P$ and $[z] \in L$.

First, we introduce the bipartite graph $D(\mathbb{F}_q)$, [9], with the following points and lines, which are infinite dimensional vectors over $\mathbb{F}_q$ written in the following way

$$(p) =$$

$$(p_{0,1}, p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}, p'_{2,2}, p_{2,3}, ..., p_{i,i}, p'_{i,i}, p_{i,i+1}, p_{i+1,1}...),$$

$$[l] = [l_{1,0}, l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}, l'_{2,2}, l_{2,3}, ..., l_{i,i}, l'_{i,i}, l_{i,i+1}, l_{i+1,1}...].$$

The point (p) is incident with the line $[l]$, which is written by the formula: $(p)I[l]$, if the following relations between their coordinates hold:

$$\begin{cases} l_{1,1} - p_{1,1} = l_{1,0}p_{0,1} \\ l_{1,2} - p_{1,2} = l_{1,1}p_{0,1} \\ l_{2,1} - p_{2,1} = l_{0,1}p_{1,1} \\ l_{i,i} - p_{i,i} = l_{0,1}p_{i-1,i} \\ l'_{i,i} - p'_{i,i} = l_{i,i-1}p_{0,1} \\ l_{i,i+1} - p_{i,i+1} = l_{i,i}p_{0,1} \\ l_{i+1,i} - p_{i+1,i} = l_{0,1}p'_{i,i} \end{cases} \tag{3}$$

where $i \geq 2$. The set of vertices of the graph $D(\mathbb{F}_q)$ of this infinite structure is $V = P \cup L$ and the set of edges consisting of all pairs $\{(p), [l]\}$ for which $(p)I[l]$.

For each positive integer $n > 2$ we obtain a finite incidence structure $(P_n, L_n, I_n)_D$ as follows. Firstly, P_n and L_n are obtained from P and L, respectively, by projecting each vector onto its n initial coordinates with respect to the natural order. The incidence I_n is then defined by imposing the first $n-1$ incidence equations and ignoring all others. The graph corresponding to the finite incidence structure (P_n, L_n, I_n) is denoted by $D(n, \mathbb{F}_q)$. $D(n, \mathbb{F}_q)$ becomes disconnected for $n \geq 6$. Graphs $D(n, \mathbb{F}_q)$ are edge transitive. It means that their connected components are isomorphic. A connected component of $D(n, \mathbb{F}_q)$ is denoted by $CD(n, \mathbb{F}_q)$. Notice that all connected components of infinite graph $D(\mathbb{F}_q)$ are q-regular trees.

The family of graphs $D(n, \mathbb{F}_q)$ is a family of q-regular, bipartite graphs of large girth (1). Graphs $D(n, \mathbb{F}_q)$, $n \geq 2$ of fixed degree q form a family of expanders with the second largest eigenvalue bounded from above by $2\sqrt{q}$, [9]. So, family $D(n, \mathbb{F}_q)$ consist of "almost Ramanujan graphs". A graph $D(n, \mathbb{F}_q)$ has practical application in the construction of error correcting codes. Firstly LDPC codes based on graphs $CD(n, \mathbb{F}_q)$ were described in [7]. They are still in practical use.

Let us consider an alternative way of presentation of q-regular infinite graph via equations over finite field F_q We consider an infinite graph $W(\mathbb{F}_q)$ with the points and lines:

$$(p) = (p_{0,1}, p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4}, ..., p_{1,i}, ...),$$

$$[l] = [l_{1,0}, l_{1,1}, l_{1,2}, l_{1,3}, l_{1,4}, ..., l_{1,i}, ...].$$

$W(\mathbb{F}_q)$ is a graph of infinite incidence structure $(P, L, I)_W$ such that a point (p) is incident with the line $[l]$ ($(p)I[l]$), if the following relations between their coordinates hold:

$$l_{1,i} - p_{1,i} = l_{1,i-1}p_{0,1} \tag{4}$$

Like in the case of $D(\mathbb{F}_q)$ for each positive integer $n > 2$ we obtain an finite incidence structure $(P_n, L_n, I_n)_W$ where P_n and L_n are obtained from P and L, respectively, by projecting each vector onto its n initial coordinates with respect to the natural order. The incidence I_n is then defined by imposing the first $n-1$ incidence equations and ignoring all others.

The graph corresponding to the finite incidence structure (P_n, L_n, I_n) is denoted by $W(n, \mathbb{F}_q)$.

The family $W(n, \mathbb{F}_q)$ is a family of q-regular, bipartite graphs with $g = 8$, given by a nonlinear system of equations.

By theorem 4.2 in [14] Wenger graph $W(n, \mathbb{F}_q)$ graph is an edge transitive one.

In fact, $W(n, \mathbb{F}_q)$ form a family of small world graphs. There is a conjecture that $CD(n, \mathbb{F}_q)$ is another family of small world graphs.

Firstly, let us consider an ordinary $n+1$-gon as a bipartite graph with vertex set $V = \{(1), (2), \ldots, (n+1)\} \cup \{[1,2], [2,3], \ldots, [n, n+1], [n+1, 1]\}$. We can write the incidence relation I in $n+1$-gon as follows:

$$(A)I[a,b] \Longleftrightarrow A = a \vee A = b.$$

A line is incident with point if this point belong to this line.

Graphs $G(n+1, \Gamma(n, \mathbb{F}_q))$ correspond to incidence structure with the point set P, the line set L and symmetric incidence relation I_G. Γ is a q-regular bipartite family of graphs defined by systems of equations. Then the number of vertices in graph G is $|V| = 2(1 + q + q^2 + \ldots + q^n)$. The graph is bipartite $V = P \cup L$ and a set V consists of:

2 elements of type t_0–$((1), \emptyset)$ and $[[1,2], \emptyset]$,

$2q$ elements of type t_1–$((2), *)$ and $[[1,2], *]$,

$2q^2$ elements of type t_2–$((n+1), *, *)$ and $[[2,3], *, *]$,

$\vdots$

$2q^n$ elements of type t_n–$((\lceil \frac{n+3}{2} \rceil), \underbrace{*, \ldots, *}_{n})$ and

$[[\lfloor \frac{n+3}{2} \rfloor, \lfloor \frac{n+5}{2} \rfloor], \underbrace{*, \ldots, *}_{n}]$.

Each $*$ represents an arbitrary element from $\mathbb{F}_q$. Brackets and parenthesis will allow the reader to distinguish points $(\cdot)$ and lines $[\cdot]$. The set of edges consisting of all pairs $\{(p), [l]\}$ for which $(p)I_G[l]$.

The incidence relation I_G in graphs $G(n+1, \Gamma(n, \mathbb{F}_q))$ is described as follows. A point of type t_0– $((1), \emptyset)$ is connected by an edge with a line of type t_0– $[[1,2], \emptyset]$ and lines of type t_1. A line of type t_0– $[[1,2], \emptyset]$ is connected by an edge with a point of type t_0– $((1), \emptyset)$ and points of type t_1. For $n \geq x, y \geq 1$, the point $(p) = ((A), \alpha_1, \alpha_2, \ldots, \alpha_x)$ of type t_x is incident $(p)I_G[l]$ with the line $[l] = [[a,b], \beta_1\beta_2, \ldots, \beta_y$ of type t_y if $A = a \vee A = b$ and the following hold:

$$\begin{cases} \alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_x = \beta_{y-1}, \text{ for } x+1 = y \\ \alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_{x-1} = \beta_y, \text{ for } x = y+1 \\ (\alpha_1, \alpha_2, \ldots, \alpha_n)I[\beta_1, \beta_2, \ldots, \beta_n] \text{ in } \Gamma, \text{ for } x = y = n \end{cases} \tag{5}$$

If we rewrite incidence relation for a graph $D(n, \mathbb{F}_q)$ with the notation for points as lines as for graph $W(n, \mathbb{F}_q)$:

$$(p) = (p_{0,1}, p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4}, ..., p_{1,i}, ...),$$

$$[l] = [l_{1,0}, l_{1,1}, l_{1,2}, l_{1,3}, l_{1,4}, ..., l_{1,i}, ...],$$

($p_{2,1} = p_{1,3}$, $p_{2,2} = p_{1,4}$, $p'_{2,2} = p_{1,5}$ and $l_{2,1} = l_{1,3}$, $l_{2,2} = l_{1,4}$, $l'_{2,2} = l_{1,5}$) then the first 5 equations describing incidence relations for graph $D(n, \mathbb{F}_q)$ can be written as follows:

$$\begin{cases} l_{1,1} - p_{1,1} = l_{1,0}p_{0,1} \\ l_{1,2} - p_{1,2} = l_{1,1}p_{0,1} \\ l_{1,3} - p_{1,3} = l_{0,1}p_{1,1} \\ l_{1,4} - p_{1,4} = l_{0,1}p_{1,2} \\ l_{1,5} - p_{1,5} = l_{1,3}p_{0,1} \end{cases} \tag{6}$$

and tables II, III,IV, V describe incidence relations I_G for "small" representatives of the family.

TABLE I
REGULARITY AND ORDER FOR SOME REPRESENTATIVES OF THE FAMILY

Construction	Regularity	|V|
$G(3, D(2, \mathbb{F}_q))$ $\cong G(3, W(2, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2)$
$G(4, D(3, \mathbb{F}_q))$ $\cong G(4, W(3, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2+q^3)$
$G(5, D(4, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2+q^3+q^4)$
$G(5, W(4, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2+q^3+q^4)$
$G(6, W(5, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2+q^3+q^4+q^5)$
$G(6, D(5, \mathbb{F}_q))$	$q+1$	$2(1+q+q^2+q^3+q^4+q^5)$

III. COMPARISON WITH PREVIOUSLY KNOWN RESULTS

The graphs $G(n+1, \Gamma(n, \mathbb{F}_q))$ have a structure which is some aspects similar to generalized polygons. They are $q+1$ regular graphs and have the same number of vertices for fixed $n+1 = 3, 4, 6$ as generalized polygons. According to the famous Feit-Higman theorem regular thick polygons exist only for $n+1 = 3, 4, 6$ (see [5]). For $n+1 = 2$ the described construction yields classical projective plane which is a generalized 3-gon and has the second eigenvalue $\lambda_1 = \sqrt{q}$. To show that the constructed graphs for $n+1 = 4, 6$ are not isomorphic to generalized quadrangles and hexagons we prove the following theorem.

Theorem 1. *Family of graphs $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ are families of graphs of girth 6.*

Proof. Graphs $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ are bipartite so there is no cycle C_3 and C_5. Because of the structure of this families there are two possibilities of appearence C_4:

1) There is a cycle C_4 consisting of two points of type t_n and two lines of type t_n. But it means that $D(n, \mathbb{F}_q)$ or $W(n, \mathbb{F}_q)$ have cycles of length 4. and we know from [14], [9] that $g(D(n, \mathbb{F}_q)) \geq 6$ and $g(W(n, \mathbb{F}_q)) \geq 6$.
2) There exists two vertices v_1 and v_2 of type t_n in the same branch which are separated by a path of length 2 ($[l_1]I(p_2)I[l_2]$), where p_2 is of type t_{n-1} and have a common neighbor of type t_n.

   Suppose that the graph has C_4. v_1 and v_2 are from the same branch so they have equal coordinates except the last one. Assume without loss of generality that these are lines and denote them as follows:
   $[l_1] = [[\lfloor \frac{n+3}{2} \rfloor, \lfloor \frac{n+5}{2} \rfloor], *_1, *_2, \ldots, *_{n-1}, Y_1]$,

TABLE II
INCIDENCE RELATIONS FOR GRAPH $G(3, W(2, \mathbb{F}_q)) \cong G(3, D(2, \mathbb{F}_q))$

	$((1), \emptyset)$	$((2), p_{0,1})$	$((3), p_{0,1}, p_{1,1})$
$[[1,2], \emptyset]$	$+$	$+$	$-$
$[[3,1], l_{1,0}]$	$+$	$-$	$+ : p_{0,1} = l_{1,0}$
$[[2,3], l_{1,0}, l_{1,1}]$	$-$	$+ : p_{0,1} = l_{1,0}$	$+ : l_{1,1} - p_{1,1} = l_{1,0}p_{1,0}$ the first incydence equation for used graph

TABLE III
INCIDENCE RELATIONS FOR GRAPH $G(4, W(3, \mathbb{F}_q)) \cong G(4, D(3, \mathbb{F}_q))$

	$((1), \emptyset)$	$((2), p_{0,1})$	$((4), p_{0,1}, p_{1,1})$	$((3), p_{0,1}, p_{1,1}, p_{1,2})$
$[[1,2], \emptyset]$	$+$	$+$	$-$	$-$
$[[4,1], l_{1,0}]$	$+$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$
$[[2,3], l_{1,0}, l_{1,1}]$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$	$+ : p_{0,1} = l_{1,0}$ $p_{1,1} = l_{1,1}$
$[[3,4], l_{1,0}, l_{1,1}, l_{1,2}]$	$-$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$	$+ : l_{1,1} - p_{1,1} = l_{1,0}p_{0,1}$, $l_{1,2} - p_{1,2} = l_{1,1}p_{0,1}$

TABLE IV
INCIDENCE RELATIONS FOR GRAPH $G(5, W(4, \mathbb{F}_q))$ AND $G(5, D(4, \mathbb{F}_q))$

	$((1), \emptyset)$	$((2), p_{0,1})$	$((5), p_{0,1}, p_{1,1})$	$((3), p_{0,1}, p_{1,1}, p_{1,2})$	$((4), p_{0,1}, p_{1,1}, p_{1,2}, p_{1,3})$
$[[1,2], \emptyset]$	$+$	$+$	$-$	$-$	$-$
$[[1,5], l_{1,0}]$	$+$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$	$-$
$[[2,3], l_{1,0}, l_{1,1}]$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$	$+ : p_{0,1} = l_{1,0}$ $p_{1,1} = l_{1,1}$	$-$
$[[4,5], l_{1,0}, l_{1,1}, l_{1,2}]$	$-$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$
$[[3,4], l_{1,0}, l_{1,1}, l_{1,2}, l_{1,3}]$	$-$	$-$	$-$ $-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$	$+ : l_{1,1} - p_{1,1} = l_{1,0}p_{0,1}$, $l_{1,2} - p_{1,2} = l_{1,1}p_{0,1}$ $l_{1,3} - p_{1,3} = l_{1,2}p_{0,1}$

TABLE V
INCIDENCE RELATIONS FOR GRAPH $G(6, W(5, \mathbb{F}_q))$ AND $G(6, D(5, \mathbb{F}_q))$

	$((1), \emptyset)$	$((2), p_{0,1})$	$((6), p_{0,1}, p_{1,1})$	$((3), p_{0,1}, p_{1,1}, p_{1,2})$	$((5), p_{0,1}, \ldots, p_{1,3})$	$((4), p_{0,1}, \ldots, p_{1,4})$
$[[1,2], \emptyset]$	$+$	$+$	$-$	$-$	$-$	$-$
$[[1,6], l_{1,0}]$	$+$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$	$-$	$-$
$[[2,3], l_{1,0}, l_{1,1}]$	$-$	$+ : p_{0,1} = l_{1,0}$	$-$	$+ : p_{0,1} = l_{1,0}$ $p_{1,1} = l_{1,1}$	$-$	$-$
$[[5,6], l_{1,0}, l_{1,1}, l_{1,2}]$	$-$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$	$-$
$[[3,4], l_{1,0}, \ldots, l_{1,3}]$	$-$	$-$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$ $p_{1,3} = l_{1,3}$
$[[4,5], l_{1,0}, \ldots, l_{1,4}]$	$-$	$-$	$-$	$-$	$+ : p_{0,1} = l_{1,0}$, $p_{1,1} = l_{1,1}$ $p_{1,2} = l_{1,2}$ $p_{1,3} = l_{1,3}$	$+ : l_{1,1} - p_{1,1} = l_{1,0}p_{0,1}$, $l_{1,2} - p_{1,2} = l_{1,1}p_{0,1}$ $l_{1,3} - p_{1,3} = l_{1,2}p_{0,1}$ $l_{1,4} - p_{1,4} = l_{1,3}p_{0,1}$

$[l_2] = [[\lfloor \frac{n+3}{2} \rfloor, \lfloor \frac{n+5}{2} \rfloor], *_1, *_2, \ldots, *_{n-1}, Y_2]$.
Denote their neighbor of type t_n as:
$(p_1) = ((\lceil \frac{n+3}{2} \rceil), \alpha_1, \alpha_2, \ldots, \alpha_n)$
and their neighbor (p_2) of type t_{n-1} from the same branch as:
$(p_2) = ((\lfloor \frac{n+3}{2} \rfloor), *_1, *_2, \ldots, *_{n-1})$. If in the graph $G(n+1, W(n, \mathbb{F}_q))$: $(p_1)I[l_1]$ and $(p_1)I[l_2]$ accordingly to (4) the following relations hold:

$$\begin{array}{cc} *_2 - \alpha_2 = *_1\alpha_1 & *_2 - \alpha_2 = *_1\alpha_1 \\ *_3 - \alpha_3 = *_2\alpha_1 & *_3 - \alpha_3 = *_2\alpha_1 \\ *_4 - \alpha_4 = *_3\alpha_1 & *_4 - \alpha_4 = *_3\alpha_1 \\ \vdots & \vdots \\ *_{n-1} - \alpha_{n-1} = *_{n-2}\alpha_1 & *_{n-1} - \alpha_{n-1} = *_{n-2}\alpha_1 \\ Y_1 - \alpha_n = *_{n-1}\alpha_1 & Y_2 - \alpha_n = *_{n-1}\alpha_1 \end{array}$$

From the above equality one can see that Y_1 and Y_2 are uniquely determined by the remaining coordinates and $Y_1 = Y_2$. So we got a contradiction.

Analogous procedure can be performed for the graph $G(n+1, D(n, \mathbb{F}_q))$. Any graph $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ without vertices of type t_n is a tree and does not have any cycle.

For an arbitrary $n \geq 2$ in $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ there is a cycle of length 6:

$$[[\lfloor \tfrac{n+3}{2} \rfloor, \lfloor \tfrac{n+5}{2} \rfloor], \underbrace{0, 0, \ldots, 0}_{n}] I((\lfloor \tfrac{n+3}{2} \rfloor), \underbrace{0, 0, \ldots, 0}_{n-1}) \; I$$
$$[[\lfloor \tfrac{n+3}{2} \rfloor, \lfloor \tfrac{n+5}{2} \rfloor], \underbrace{0, 0, \ldots, 0}_{n-1}, 1] I((\lceil \tfrac{n+3}{2} \rceil), \underbrace{0, 0, \ldots, 0}_{n-1}, 1) \; I$$
$$[[\lceil \tfrac{n+3}{2} \rceil, \lceil \tfrac{n+5}{2} \rceil], \underbrace{0, 0, \ldots, 0}_{n-1}] I((\lceil \tfrac{n+3}{2} \rceil), \underbrace{0, 0, \ldots, 0}_{n}) \; I$$
$$[[\lfloor \tfrac{n+3}{2} \rfloor, \lfloor \tfrac{n+5}{2} \rfloor], \underbrace{0, 0, \ldots, 0}_{n}].$$

□

The above theorem leads to the following conclusion.

Corollary 2. *For $n \geq 3$ graphs $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ are not isomorphic to generalized polygons.*

IV. Expanding and other properties

The families $G(n+1, \Gamma(n, \mathbb{F}_q))$ consist of bipartite graphs with $|V| = 2(1 + q + q^2 + \ldots + q^n)$ vertices and $(q+1)(1 + q + q^2 + \ldots + q^n)$ edges. $G(n+1, D(n, \mathbb{F}_q))$ and $G(n+1, W(n, \mathbb{F}_q))$ are $q+1$-regular sparse graphs and the density according to (2) is

$$\frac{q+1}{2(q + \ldots + q^n) + 1}.$$

Fig. 1. shows the graph $G(3, \Gamma(2, \mathbb{F}_2))$ with 14 vertices $V = \{((1), \emptyset), ((2), 0), ((2), 1), ((3), 0, 0), ((3), 0, 1), ((3), 1, 0), ((3), 1, 1)\} \cup \{[[1, 2], \emptyset], [[1, 3], 0], [[1, 3], 1], [[2, 3], 0, 0], [[2, 3], 0, 1], [[2, 3], 1, 0], [[2, 3], 1, 1]\}$ and density $\frac{3}{13}$. The red vertices correspond to points and the blue vertices correspond to lines.

Each of the representatives of the presented family is $q+1$-regular graph so the first eigenvalue of the adjacency matrix, corresponding to this graph, is $\lambda_0 = q + 1$. Let us denote the second eigenvalue by $\lambda_1 = \max_{\lambda_i \neq q+1} |\lambda_i|$. On the basis of numerical calculations included in tables (VI), (VII), (IX), (X), (XI) we state the following conclusions:

- For $q = 2, 3, 4, 5, 7, 9, 11, 13, 17, 19, 23$ the constructed graphs $G(4, D(3, \mathbb{F}_q))$=$G(4, W(3, \mathbb{F}_q))$ are Ramanujan graphs. The spectral gap increases with the value of q. Basing on on this observation we have included Conjecture 1.
- For $q = 2, 3, 4, 5, 7, 11$ the constructed graphs $G(5, D(4, \mathbb{F}_q))$ and $G(5, W(4, \mathbb{F}_q))$ are Ramanujan graphs. The spectral gap $|\lambda_0 - \lambda_1| = |q + 1 - 2\sqrt{q}|$ increases with the value of q and basing on this observation we have included Conjecture 2.
- For $q = 2, 3, 4, 5, 7$ the constructed graphs $G(6, D(5, \mathbb{F}_q))$ and $G(6, W(5, \mathbb{F}_q))$ are expander graphs. The spectral gap for graph $G(6, W(5, \mathbb{F}_q))$ increases with the value of q and for graph $G(6, D(5, \mathbb{F}_p))$ increases with the value of p. This observation allows us to formulate Conjecture 3.

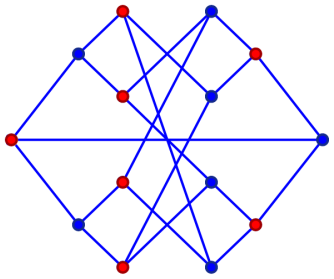


Fig. 1. $G(3, \Gamma(2, \mathbb{F}_2))$ with $|V| = 2(1 + 2 + 2^2) = 14$.

TABLE VI
Expanding properties of $G(4, D(3, \mathbb{F}_q))$=$G(4, W(3, \mathbb{F}_q))$

Number field	regularity $q+1$ first eigenvalue	second eigenvalue	$2\sqrt{q}$	$\lvert V \rvert$
$\mathbb{F}_2$	3	2.2882	2.8284	30
$\mathbb{F}_3$	4	2.8025	3.4641	80
$\mathbb{F}_4$	5	3.2361	4	170
$\mathbb{F}_5$	6	3.6180	4.4721	312
$\mathbb{F}_7$	8	4.2809	5.2915	800
$\mathbb{F}_{11}$	12	5.3664	6.6332	2928
$\mathbb{F}_{13}$	14	5.8339	7.2111	4760
$\mathbb{F}_{17}$	18	6.6713	8.2462	10440
$\mathbb{F}_{19}$	20	7.0528	8.7178	14480
$\mathbb{F}_{23}$	24	7.7598	9.5917	25440

We can use a finite ring $\mathbb{Z}_s$ and modulo operation instead of $\mathbb{F}_q$. The incidence relation for graph $G(n+1, \Gamma(n, \mathbb{Z}_s))$ can be described the same as for graph $G(n+1, D(n, \mathbb{F}_q))$. When we choose $s = 2r$ then the graphs $G(4, D(3, \mathbb{Z}_{2r}))$=$G(4, W(3, \mathbb{Z}_{2r}))$ have interesting constant value of spectral gap: $|\lambda_0 - \lambda_1| = 1$, for $2 \leq r \leq 13$. The

results of such calculations (Tab. VII) allow us to formulate Conjecture 4. When we choose $s = 3r$, where $r = 3, 5, 7, 9$, then the graphs $G(4, D(3, \mathbb{Z}_{3r}))$=$G(4, W(3, \mathbb{Z}_{3r}))$ have an interesting value of the second largest eigenvalue

$$\lambda_1 = 2\sqrt{3r\left\lfloor \frac{r}{2} \right\rfloor + \frac{3r}{2}}$$

and the spectral gap increases with the value of q (Tab. VIII). Basing on on this observation we can not say whether for arbitrarily large s above formula is true. If yes, there is a question about r: if it should be a prime power ($\neq 2^l$) or an odd number? To answer this question we must calculate λ_1 for $r = 15$ but this case can not be investigated with MATLAB in computer with 8GB RAM. The adjacency matrix in this case has 186392×186392 elements.

TABLE VII
EXPANDING PROPERTIES OF $G(4, D(3, \mathbb{Z}_{2r}))$=$G(4, W(3, \mathbb{Z}_{2r}))$

Finite ring	**regularity** $q+1$ **first eigenvalue**	**second eigenvalue**	$2\sqrt{q}$	$\lvert V\rvert$
$\mathbb{Z}_4$	5	4	4	170
$\mathbb{Z}_6$	7	6	4.899	518
$\mathbb{Z}_8$	9	8	5.6569	1170
$\mathbb{Z}_{10}$	11	10	6.3246	2222
$\mathbb{Z}_{12}$	13	12	6.9282	3770
$\mathbb{Z}_{14}$	15	14	7.4833	5910
$\mathbb{Z}_{16}$	17	16	8	8738
$\mathbb{Z}_{18}$	19	18	8.4853	12350
$\mathbb{Z}_{20}$	21	20	8.9443	16842
$\mathbb{Z}_{22}$	23	22	9.3808	22310
$\mathbb{Z}_{24}$	25	24	9.798	28850
$\mathbb{Z}_{26}$	27	26	10.198	36558

TABLE VIII
EXPANDING PROPERTIES OF $G(4, D(3, \mathbb{Z}_{3r}))$=$G(4, W(3, \mathbb{Z}_{3r}))$

Finite ring	**regularity** $q+1$ **first eigenvalue**	**second eigenvalue**	$2\sqrt{q}$	$\lvert V\rvert$
$\mathbb{Z}_9$	10	7.3485	6	1640
$\mathbb{Z}_{15}$	16	12.2474	7.746	7232
$\mathbb{Z}_{21}$	22	17.1464	9.1652	19448
$\mathbb{Z}_{27}$	28	22.0454	10.3923	40880

TABLE IX
EXPANDING PROPERTIES OF $G(5, D(4, \mathbb{F}_q))$ AND $G(5, W(4, \mathbb{F}_q))$

Number field	**regularity** $q+1$ **first eigenvalue**	**second eigenvalue**	$2\sqrt{q}$	$\lvert V\rvert$
$\mathbb{F}_2$	3	2.7855	2.8284	62
$\mathbb{F}_3$	4	3.4641	3.4641	242
$\mathbb{F}_4$	5	4	4	682
$\mathbb{F}_5$	6	4.4721	4.4721	1562
$\mathbb{F}_7$	8	5.2915	5.2915	5602
$\mathbb{F}_{11}$	12	6.6332	6.6332	32210

Conjecture 1. *The graphs $G(4, D(3, \mathbb{F}_q))$ and $G(4, W(3, \mathbb{F}_q))$ for arbitrary large q are $q+1$-regular Ramanujan graphs and $\lambda_1 \leq 2\sqrt{q}$.*

Conjecture 2. *The graphs $G(5, D(4, \mathbb{F}_q))$ and $G(5, W(4, \mathbb{F}_q))$ for arbitrary large q are $q+1$-regular Ramanujan graphs and $\lambda_1 = 2\sqrt{q}$.*

TABLE X
EXPANDING PROPERTIES OF $G(6, W(5, \mathbb{F}_q))$

Number field	**regularity** $q+1$ **first eigenvalue**	**second eigenvalue**	$2\sqrt{q}$	$\lvert V\rvert$
$\mathbb{F}_2$	3	2.8688	2.8284	126
$\mathbb{F}_3$	4	3.8979	3.4641	728
$\mathbb{F}_4$	5	4.4721	4	2730
$\mathbb{F}_5$	6	5.0321	4.4721	7812
$\mathbb{F}_7$	8	5.9541	5.2915	39216

TABLE XI
EXPANDING PROPERTIES OF $G(6, D(5, \mathbb{F}_q))$

Number field	**regularity** $q+1$ **first eigenvalue**	**second eigenvalue**	$2\sqrt{q}$	$\lvert V\rvert$
$\mathbb{F}_2$	3	2.9032	2.8284	126
$\mathbb{F}_3$	4	3.3557	3.4641	728
$\mathbb{F}_4$	5	4.8284	4	2730
$\mathbb{F}_5$	6	4.6852	4.4721	7812
$\mathbb{F}_7$	8	5.9228	5.2915	39216

Conjecture 3. *The graphs $G(6, D(5, \mathbb{F}_p))$ and $G(6, W(5, \mathbb{F}_q))$ for arbitrary large q (primr power) and p (prime number) are expanders.*

Conjecture 4. *The graphs $G(4, D(3, \mathbb{Z}_{2r}))$ and $G(4, W(3, \mathbb{Z}_{2r}))$ for arbitrary large r are $2r+1$-regular expander graphs with constant spectral gap $|2r+1-\lambda_1| = 1$.*

The graphs $G(n+1, \Gamma(n, \mathbb{F}_q))$ for arbitrary n, q and any bipartite graph Γ are connected even if Γ is disconnected. What more we have conjecture that the family $G(n+1, \Gamma(n, \mathbb{F}_q))$ is $q+1$-connected, namely highly connected. A graph is said to be k-connected when there does not exist a set of $k-1$ vertices whose removal disconnects the graph.

The connectivity of graphs is important property used in many practical and theoretical aspects.

REFERENCES

[1] N. L. Biggs, *Algebraic Graph Theory,* (2nd ed), Cambridge, University Press, 1993.

[2] N. L. Biggs, "Graphs with large girth," *Ars Combinatoria,* 1988, pp. 73–80.

[3] B. Bollobas, *Extremal Graph Theory,* Academic Press, 1978.

[4] A. Brouwer, A. Cohen, A. Neumaier, *Distance-Regular Graphs,* Springer-Verlag, 1989.

[5] W. Feit, D. Higman, "The nonexistence of certain generalised polygons," *J. of Algebra 1,* 1964, pp. 114–131.

[6] D. Gillman, "A Chernoff bound for random walks on expander graphs," *SIAM J. Comput.,* 27(4), pp. 1203–1220 (electronic), 1998.

[7] P. Guinand, J. Lodge, "Tanner type codes arising from large girth graphs," *in Canadian Workshop on Information Theory CWIT,* Toronto, Ontario, Canada, 1997, pp. 5–7.

[8] S. Hoory, N. Linial, A. Wigderson, "Expander graphs and their applications," *Bulletin (New Series) of the American Mathematical Society,* vol. 43, 2006, pp. 439–561.

[9] F. Lazebnik, V. A. Ustimenko, A. J. Woldar, "A characterization of the components of the graphs $D(k, q)$," *Discrete Mathematics,* vol. 157, 1996, pp. 271–283.

[10] F. Lazebnik, V. A. Ustimenko, A. J. Woldar, "A new series of dense graphs of high girth," *Bulletin (New Series) of the AMS,* vol. 32, 1995, pp. 73–79.

[11] A. Lubotsky, R. Philips, P. Sarnak, "Ramanujan graphs," *Combinatorica,* vol. 9, 1988, pp. 261–277.

[12] A. Lubotzky and T. Nagnibeda, "Not every uniform tree covers Ramanujan graphs," *J. Combin. Theory Ser. B,* 1998, pp. 202–212.

[13] G. A. Margulis, "Explicit constructions of expanders," *Problemy Peredači Informacii,* 1973, pp. 71–80.
[14] V. Futorny, V. Ustimenko, "On small world semiplanes with generalised Schubert cells," *Acta Appl Math,* 2007, pp. 47–61.
[15] R. Weiss, "Distance transitive graphs and generalised polygons," *Arch. Math,* vol. 45, 1985, pp.186–192.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 507–510

Performance Impact of Reconfigurable L1 Cache on GPU Devices

Sasko Ristov, Marjan Gusev
Ss. Cyril and Methodius University
Rugjer Boshkovik 16, PO Box 393, 1000 Skopje, Macedonia
Email: {sashko.ristov, marjan.gushev}@finki.ukim.mk

Leonid Djinevski, Sime Arsenovski
FON University
Av. Vojvodina, 1000 Skopje, Macedonia,
Email: {leonid.djinevski, sime.aresnovski}@fon.edu.mk

***Abstract*—The newest GPU Kepler architecture offers a reconfigurable L1 cache per Streaming Multiprocessor with different cache size and cache associativity. Both these cache parameters affect the overall performance of cache intensive algorithms, i.e. the algorithms which intensively reuse the data. In this paper, we analyze the impact of different configurations of L1 cache on execution of matrix multiplication algorithm for different problem sizes. The basis of our research is the existing theoretical analysis of performance drawbacks which appear for matrix multiplication while executed on multicore CPU. We perform series of experiments to analyze the matrix multiplication execution behavior on GPU and its set associative L1 and L2 cache memory with three different configurations: cache size of 16KB, 32KB and 48KB with appropriate set associativity of 4 and 6, respectively. The results show that only L2 cache impacts the algorithm's overall performance, particularly the L2 capacity and set associativity. However, the configuration of the L1 cache with 48KB and 6-way set associativity slightly reduces these performance drawbacks, compared to other configurations of L1 with 32KB and 16KB using 4-way cache set associativity, due to greater set associativity.**

***Index Terms*—Cache Memory, Set Associativity, GPGPU.**

I. Introduction

CACHE memory is a very important part of memory hierarchy since it reduces the performance gap between the main memory and the CPU [1]. The algorithm performance with a certain problem size depends on several cache parameters: cache size, cache replacement policy, cache levels, cache-line size, cache inclusivity, cache associativity, etc.

Today's GPU (Graphics Processing Unit) devices are more appropriate for applications with regular data access patterns [2]. They have multilevel set associative cache memory expressed with L1 and L2 level. The former is private per SM (Streaming Processor), while the latter is shared among all SMs on a single GPU device. The NVIDIA's Fermi architecture introduced size configuration (and automatically the appropriate cache set associativity) of L1 cache memory, while the newest Kepler architecture allows the programmer even further configuration.

In this paper, we configure the GPU device with three different cache sizes and two different set associativity sizes in order to determine how this new feature impacts the most common cache intensive algorithm, i.e. dense matrix multiplication (DMM). Our intention is neither to speedup the algorithm execution using the power of many core GPU, nor to speedup the algorithm using some existing transformations, but to use the DMM algorithm as a benchmark and evaluate the impact of cache sizes and associativity on the overall performance. We use only one processing unit of only one SM and realize a micro-benchmark to avoid the impact of many cores and potential additionally generated cache misses.

Since the cache set associativity can provide huge performance drawbacks for cache intensive algorithms, such as DMM, we perform additional analysis on the performance of those matrix sizes where the drawbacks are expected due to L1 and L2 cache set associativity. A performance drawback is a phenomenon where the performance does not follow the existing trend and has smaller value than the performance obtained in the neighboring points. Usually this is reflected as a negative performance peak, i.e. the performance in analyzed point x is lower than the performance in the points left or right of x, which follow a trend in performance behavior.

The goal in this research is to determine which configuration of L1 cache memory provides the best cost - performance ratio.

The rest of the paper is organized as follows. In Section II, we give an overview of related work in the area of the research problem. Analysis of possible performance drawbacks and a description of methodology used in the experiments is presented in Section III. The results of the experiments are elaborated in Section IV. Finally, we conclude our work followed by our plans for future work in Section V.

II. Related work

The latest GPUs have two level cache hierarchy organized with set cache associativity. The impact of cache associativity on GPU performance was analyzed by several authors. Performance drawbacks are likely expected for DMM execution on GPU for particular matrix sizes, due to the usage of only small subset of the cache due to the matrix storage pattern, similar to the effect on multicore architectures reported by Gusev and Ristov [3]. An example of huge performance drawbacks of DGEMM (Double precision General Matrix Multiply) for matrix size that are multiples of 1024 are reported by Matsumoto et. al [4] without deeper explanation. This problem was also analyzed by Batson and Vijakumar [5]. They propose reactive mechanisms (selective displacement and feedback) as a solution. Calder et al. propose that way prediction [6] can improve set-associative cache access times.

978-83-60810-53-8/$25.00

TABLE I
CONDITIONS FOR PERFORMANCE DRAWBACKS

L1 (16KB)				L1 (32KB)				L1 (48KB)				L2 (512KB)			
N	d	N/d	n	N	d	N/d	n	N	d	N/d	n	N	d	N/d	n
64	16	4	4	64	32	2	4	64	32	2	6	64	128	0.5	16
128	**8**	16	4	**128**	**16**	8	4	**128**	**16**	8	6	128	64	2	16
256	**4**	64	4	**256**	**8**	32	4	**256**	**8**	32	6	256	32	8	16
512	**2**	256	4	**512**	**4**	128	4	**512**	**4**	128	6	**512**	**16**	32	16
1024	**1**	1024	4	**1024**	**2**	512	4	**1024**	**2**	512	6	**1024**	**8**	128	16
2048	**/**	/	4	**2048**	**1**	2048	4	**2048**	**1**	2048	6	**2048**	**4**	512	16
4096	**/**	/	4	4096	**/**	/	4	4096	**/**	/	6	**4096**	**2**	2048	16
8192	**/**	/	4	8192	**/**	/	4	8192	**/**	/	6	**8192**	**1**	8192	16

Greater set associativity will reduce the cache misses, but will still not improve the performance since this will increase the cache hit access time. Padding the first element of the second matrix will amortize the performance drawback due to cache associativity [7]. Hongil [8] dynamically selects an optimized replacement policy for each cache set via workload speculation mechanism to improve the cache performance. Ding et al. [9] designed a software runtime library to include intelligence in the cache allowing the programmers to manage and optimize last level cache usage by allocating proper cache space for different data sets of different threads.

Gusev and Ristov [3] proved both theoretically and experimentally that CPU cache memory storage pattern can significantly reduce the performance of DMM execution by increasing the generation of last level cache misses due to the usage of set associative cache. By using their theorems one can determine the matrix sizes where maximum cache performance drawback in the matrix multiplication algorithm will appear due to matrix storage pattern in a n-way associative memory. Our recent research proved those theoretical results for GPU's L2 set associativity cache [10]. In this paper, we set a research problem to check validity of theoretical results and experimentally test if they hold for different configurations of L1 set associative cache in GPU architectures.

Two problems are exposed with usage of the caches, cache capacity problem refers to the lack of the resources, while the cache associativity problem refers to inefficient usage of the cache. In this paper, we are focused on performance analysis of the cache associativity problem.

III. TESTING METHODOLOGY

We use the classical DMM algorithm, where the operations are performed column-wise in order to exploit the effect of cache reuse, assuming that the matrix elements are stored in row-major order, usually used in C programming language.

This research is focused on GPUs analyzing both the cache capacity and cache associativity problems defined for multicore architectures by Gusev and Ristov [3].

Table I presents the cache parameters of the GPU model GeForce GTX 680 for each configuration of L1 cache and for L2 cache, using the theoretical analysis described in [3]. According to this analysis, we expect the performance drawbacks for bold values of d ($n < N/d$), as presented in Table I. Further on we calculate that maximum N for which the performance drawback will appear is determined as:

- $N = 1024$ for L1 cache configured with 16KB cache and 4 way set associative;
- $N = 2048$ for L1 cache configured with 32KB (4 way set) or 48KB (6 way set); and
- $N = 8192$ for 512KB L2 cache 16 way set associative.

This paper aims to confirm these theoretical results for GPUs by experimental research.

The Ubuntu 12.04 LTS operating system runs on Intel i7-3770 CPU@3.40GHz, 32GB of Kingston RAM @ 1.60GHz and NVIDIA GeForce GTX 680 GPU. The implementations of all of the experiments are compiled with the Nvidia's nvcc compiler from the CUDA 5.0 toolkit.

We conducted experiments for three different configurations of 16/32/48KB of L1 cache memory. Since the cache-line size (cbs) does not influence the equations in our theoretical analysis, there isn't any particular reason to choose a value of cbs. In our case we have chosen 128B. We also assume that the cache memories are set-associative.

Six experiments of the sequential matrix multiplication algorithm were performed in the regions around the critical matrix sizes: $N = 64, 128, 256, 512, 1024$ and 2048. The sequential implementation of the DMM runs one thread per only one active SM [11], thus the whole L1 cache is dedicated to the thread. Each experiment consists of twenty test cases for problems in the area around the critical points.

Average execution time is measured from 10 iterations, excluding the first iteration.

IV. RESULTS OF THE EXPERIMENTS

The obtained results of the experiments on the GPU to analyze the impact of different L1 cache size and set associativity configuration are presented in this section. Our focus is to observe the areas around the problem size for the points where maximum drawbacks are expected from the theoretical analysis performed in Section III. All 6 experiments are performed for three different configurations of L1 cache size, i.e., 16KB, 32KB or 48KB.

The results on performance around the area of $N = 64$ (**Experiment 1**) are depicted in the left part of Figure 1. For this example, both matrices can be stored completely in the L1 cache.

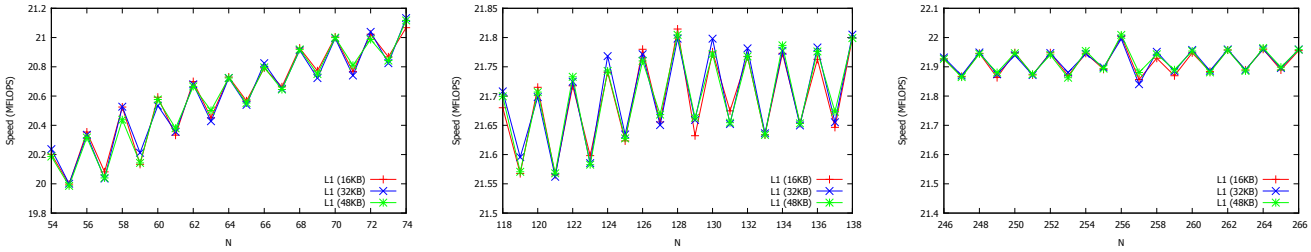


Fig. 1. Speed in the area around $N = 64$ (left), $N = 128$ (middle) and 256 (right)

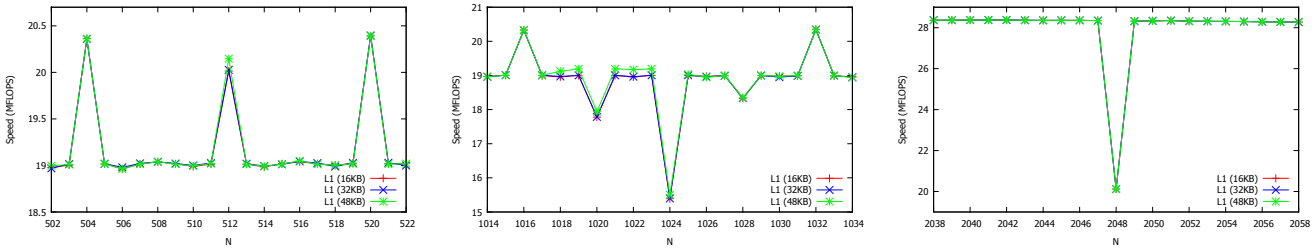


Fig. 2. Speed in the area around $N = 512$ (left), $N = 1024$ (middle) and 2048 (right)

This experiment proves our theoretical analysis since the performance drawbacks do not appear in this region. The elements of a matrix B column can be stored in a particular set and no cache misses will be generated. All three L1 cache size configuration comply with the theoretical analysis.

We observe a very strange speed curve, which increases for even matrix size N and decreases for odd matrix size N. This phenomenon appears due to the average load time for a matrix element, which is smaller for an even matrix size N. Therefore, the matrix elements fulfill the cache line more efficiently within a given cache line.

The speed in this region has a positive increasing trend due to increased amount of data reuse without repeated generation of cache misses.

Experiment 2 covers the area around matrix size $N = 128$. The middle part of Figure 1 depicts the results. The second matrix cannot be stored completely in the L1 cache for these matrix sizes and therefore drawbacks appear due to insufficient L1 cache. However, comparing it with the previous experiment, we can conclude that despite the increased number of generated L1 cache misses, the speed in the Experiment 2 is greater than the speed achieved in Experiment 1.

The results show that performance drawbacks due to L1 cache set associativity are seemingly small due to the unsatisfied condition for L2 cache set associativity, as presented in Table I. We observe a slight speed discrepancy for different L1 cache configurations using the same matrix size. Similar to the Experiment 1, performance discrepancy is observed for even and odd matrix sizes. The speed holds the positive trend as in the Experiment 1, but with smaller intensity.

The **Experiment 3** covers the area of $N = 256$. The second matrix cannot be stored completely in the L1 cache as in the Experiment 2. The speed in this region has even lighter positive trend than experiments 1 and 2, as depicted in the right part of Figure 1.

Similar to the previous case, performance drawbacks are not observed in this region, due to the smaller impact of L1 cache associativity in comparison to L2, where the set associativity problem does not appear in this region. Additionally, we observe that the performance for $N = 256$ is even greater than the values near N in that region, for each L1 cache size configuration. We explain this observation with the fact that despite the L1 cache associativity problem, the whole matrix row can be stored in the exact number of cache blocks and no L2 cache misses are generated neither due to L2 capacity nor L2 associativity problem. Therefore, the average access time is smaller for a matrix element stored in a particular cache block for $N = 256$. We observe a slightly higher speed while L1 cache is configured with 48KB.

Similar performance discrepancy is observed for even and odd matrix sizes, as in previous experiments.

The **Experiment 4** covers the area around matrix size $N = 512$ and the results of the experiment are depicted in Figure 2 (left). Matrix B cannot be stored completely neither in L1 nor L2 cache, and thus the drawback exists mainly due to their size and associativity.

Although one might think the results are strange in this region, there is an explanation. The speed for matrix size $N = 512$ is much greater than the other problem sizes in the region, except for $N = 504$ and $N = 520$. We have also tested the other close positioned points $N \in \{488, 496, 528, 536\}$ and achieved the same positive peaks. The conclusion is that the execution for $N = 512$ has performance drawback compared to these points. The observation consists of two parts: greater speed and performance drawbacks. The former appears for the same reason as explained for $N = 256$. The latter appears

compared due to cache associativity and condition of Table I, compared to points $N = 504$ and $N = 520$ (and for the other points that we measured additionally).

We also observe a slightly better speed while L1 cache is configured with 48KB for $N = 512$.

Experiment 5 analyzes the area around $N = 1024$. Matrix B cannot be stored completely in L2 cache and drawbacks appear due to L2 cache size and associativity. The speed drawbacks are clearly detected and they are depicted in Figure 2 (middle).

Significant performance drawback appears as stated in Table I, but also smaller performance drawbacks appeared in points $N - 4$ and $N + 4$ (as well as for $N + 12$ and $N - 12$).

The same positive peaks are observed in the points $N + 8$ and $N - 8$ as in the region around $N = 512$.

We also observe a slightly better speed while L1 cache is configured with 48KB in the point $N = 1024$, similar to the result for $N = 512$.

Experiment 6 analyzes the area around the matrix size $N = 2048$ where matrix B also cannot be stored completely in the L2 cache. Performance drawback is clearly observed for $N = 2048$ as depicted in Figure 2 (right).

Neither additional positive nor negative peaks are observed in this area since the number of generated L2 cache misses is huge. The impact of L2 cache capacity problem is greater than the positive impact of cache memory to data locality in this region, i.e., loading the elements of the whole cache line while reading one element of that cache line.

V. Conclusion and Future Work

The performance of GPU general purpose application can be seriously degraded by the set associative L1/L2 caches. In this paper, we present the performance drawbacks for specific problem sizes of the DMM algorithm for different L1 cache configurations and fixed L2 cache size. We have performed series of experiments in the areas of critical problem sizes, which prove the analysis.

It is shown that the side effects of the associative cache on the CPU, as discussed in our earlier paper are also present in the GPU environment. However, this paper shows also some other interesting conclusions, due to a specific organization of caches in GPU, which is quite different from CPU (very small 1st level cache and no third level cache in comparison to CPU).

A total of six experiments were evaluated in points where theoretical results expect negative performance peaks with analysis of speed diagrams. The results show that the configuration of L1 cache size does not influence significantly on performance drawbacks, which appear for $N = 1024$ and 2048 due to L2 cache set associativity. Because L2 cache set size is enough to fit the cache storage requirements for problem sizes $N = 64, 128$ and 256, performance drawbacks are not observed, i.e., the algorithm performance depends mostly on L2 cache size, rather than L1's. The performance for $N = 256$ is even greater than the matrix size values near N in that region for all L1 cache size configuration.

An interesting phenomenon appears in the region around $N = 512$. The speed for $N = 512$ is greater than the other problem sizes in the region, except for $N = 504$ and $N = 520$. Although higher values are obtained than the neighboring points, still there is a performance drawback compared to analyzed points $N = 504$ and $N = 520$. More interestingly, we have found smaller negative peaks in the region around $N = 1024$ in points $N + 4$ and $N - 4$, as well as positive peaks in the points $N + 8$ and $N - 8$.

Another phenomenon was observed in the regions around $N = 64, 128$ and 256, i.e., the speed increases for even matrix size N and decreases for odd matrix size N. This happens due to the effect of loading the elements of the whole cache line while reading one element of the same cache line.

Probably the most important result is to report the platform impact of reconfigurable cache, i.e. what the user can choose for configuration of the L1 cache to achieve maximum processing speed and avoid associativity problems.

Future work will cover further research on these phenomenons, as well as analysis of correlation of power consumption with the L1/L2 capacity and associativity, since the results show that L1 cache size does not impact the algorithm performance, but different cache associativity configuration due to different cache size configuration can reduce the power consumption.

Also, we plan to measure the number of generated L1 and L2 cache misses to determine the performance drawbacks and performance discrepancies more precisely.

References

[1] J. L. Hennessy and D. A. Patterson, *Computer Architecture, Fifth Edition: A Quantitative Approach*. MA, USA: Elsevier, 2012.

[2] D. Tarjan, J. Meng, and K. Skadron, "Increasing memory miss tolerance for simd cores," in *Proc. of the Conf. on High Performance Computing Networking, Storage and Analysis*, ser. SC '09, 2009, pp. 22:1–22:11.

[3] M. Gusev and S. Ristov, "Performance gains and drawbacks using set associative cache," *Journal of Next Generation Information Technology (JNIT)*, vol. 3, no. 3, pp. 87–98, 31 Aug 2012.

[4] K. Matsumoto, N. Nakasato, and S. Sedukhin, "Implementing a code generator for fast matrix multiplication in opencl on the gpu," in *Embedded Multicore Socs (MCSoC), 2012 IEEE 6th International Symposium on*, sept. 2012, pp. 198–204.

[5] B. Batson and T. N. Vijaykumar, "Reactive-associative caches," in *Proceedings of the 2001 International Conference on Parallel Architectures and Compilation Techniques*, ser. PACT '01, 2001, pp. 49–60.

[6] B. Calder, D. Grunwald, and J. Emer, "Predictive sequential associative cache," in *Proceedings of the 2nd IEEE Symposium on High-Performance Computer Architecture*, ser. HPCA '96, 1996, pp. 244–253.

[7] S. Williams, L. Oliker, R. Vuduc, J. Shalf, K. Yelick, and J. Demmel, "Optimization of sparse matrix-vector multiplication on emerging multicore platforms," *Parallel Comput.*, vol. 35, no. 3, pp. 178–194, 2009.

[8] H. Yoon, T. Zhang, and M. H. Lipasti, "Sip: Speculative insertion policy for high performance caching," Computer Sciences Department University of Wisconsin-Madison, Tech. Rep. 1676, 2010.

[9] X. Ding, K. Wang, and X. Zhang, "Ulcc: a user-level facility for optimizing shared cache performance on multicores," in *Proceedings of the 16th ACM symposium on Principles and practice of parallel programming*, ser. PPoPP '11. ACM, 2011, pp. 103–112.

[10] L. Djinevski, S. Arsenovski, S. Ristov, and M. Gusev, "Performance drawbacks for matrix multiplication using set associative cache in gpu devices," in *MIPRO, 2013 Proceedings of the 36th International Convention, IEEE Conference Publications*, Croatia, 2013, pp. 213–218.

[11] L. Djinevski, S. Ristov, and M. Gusev, "Superlinear speedup for matrix multiplication in gpu devices," in *ICT Innovations 2012*, ser. AISC. Springer Berlin Heidelberg, 2013, vol. 207, pp. 285–294.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 511–514

Analyzing of Some Performance Measures for Parallel Matrix Multiplication

Halil Snopce
South East Europian University,
CST Faculty, Tetovo 1200, R.
Macedonia
Email: h.snopce@seeu.edu.mk

Azir Aliu
South East Europian University,
CST Faculty, Tetovo 1200, R.
Macedonia
Email: azir.aliu@seeu.edu.mk

Abstract—**In order to make a proper selection for the given matrix-matrix multiplication operation and to decide which is the best suitable algorithm that generates a high throughput with a minimum time, a comparison analysis and a performance evaluation for some algorithms is carried out using the identical performance parameters**

Keywords- parallel algorithms for matrix multiplication, systolic array, linear transformation, nonlinear transformation, performance measures, number of processor elements Introduction.

I. INTRODUCTION

Most of the parallel algorithms for matrix multiplication use matrix decomposition that is based on the number of processors available. This includes the systolic algorithm [1], Cannon's algorithm [2], Fox's and Otto's Algorithm [3], PUMMA (Parallel Universal Matrix Multiplication) [4], SUMMA (Scalable Universal Matrix Multiplication) [5] and DIMMA (Distribution Independent Matrix Multiplication) [6]. The standard method for multiplying $n \times n$ matrices requires $O(n^3)$ multiplications. Most existing parallel algorithms are parallelization of the standard method. All implementations of the standard method have a cost, i.e., time-processor product of at least $O(n^3)$. Therefore, it is interesting to develop highly parallel and processor efficient algorithms that have less than $O(n^3)$ cost.

II. WHY SYSTOLIC ARRAY?

The MPI technique needs two kinds of time to complete the multiplication process, t_c and t_f, where t_c represents the time it takes to communicate one data between processors and t_f is the time needed to multiply or add elements of two matrices. It is assumed that matrices are of type $n \times n$. The other assumption is that the number of processors is p. Each processor holds n^2/p elements and it was assumed that n^2/p is set to a new variable m^2.

The number of arithmetic operations units will be denoted by f. The number of communication units will be denoted by c. The quotient $q = f/c$ will represent the average number of flops per communication access. The speedup is $S = q \cdot (t_f / t_c)$. We assume that the number of processors is $p = 4$. The dimension of the matrix is $n = 600$. The last assumption is that $t_f / t_c = 0.1$. Also we need to use the Efficiency formula $E = S/p$. In the table 1 we record the results that we obtain for all algorithms, under the assumptions that we made. [2, 3, 4, 7, 8, 9, 10, 11].

III. DEFINITION OF SOME PERFORMANCE MEASURES FOR SYSTOLIC ARRAYS

Definition 1: The array size (Ω) is the number of PEs in the array.

Definition 2: The computation time (T) is the sum of the time for the date input in the array-T_{in}, the time for the algorithm executing-T_{exe} and the time necessary for dates leaving the array-T_{out}, i.e.

$$T = T_{in} + T_{exe} + T_{out} \tag{1}$$

Definition 3: The execution time, T_{exe}, is defined as:

$$T_{exe} = 1 + \max_{(t,x,y) \in P_{ind}} t - \min_{(t,x,y) \in P_{ind}} t_{out} \tag{2}$$

Theorem 1: [12] The execution time is given by the relation:

$$T_{exe} = 1 + \sum_{j=1}^{3} (N_j - 1) \cdot \left| \min_t t_{ij} \right| \tag{3}$$

Definition 4: The Pipelining period (α)**:** The time interval between two successive computations in a PE. If λ is a scheduling vector and u is a projection direction, then the pipelining period is given by the relation:

$$\alpha = \lambda^T u \tag{4}$$

978-83-60810-53-8/$25.00

Definition 5: The geometric area (g_a) of a two-dimensional systolic array is the area of the smallest convex polygon which bounds the PEs in the (x, y)-plane. The geometric area is given by the formula:

$$g_a = (N_1-1)(N_2-1)|T_{13}| + (N_1-1)(N_3-1)|T_{12}| + (N_2-1)(N_3-1)|T_{11}| \quad (5)$$

Definition 6: The Speedup (S) of a systolic array is the ratio of the processing time in the SA to the processing time in a single processor (T_1), i.e.

$$S = \frac{T_1}{T} \quad (6)$$

Definition 7: The Efficiency (E) is defined as the ratio of the speedup to the number of PEs in the array i.e.

$$E = \frac{S}{\Omega} = \frac{T_1}{T x \Omega} \quad (7)$$

Theorem 2: [12, 13] The number of processors on SHSA array is (the array where we have no using the linear transformation):

$$\Omega = 3N^2 - 3N + 1 \quad (8)$$

Theorem 3: [14] The number of processing elements in 2-dimensional systolic array for the algorithm of matrix-matrix multiplication for which is used the projection direction $u = [1 \ \ 1 \ \ 1]^T$, could be reduced and given with $\Omega = N^2$.

Theorem 4: [15] The number of PEs for the systolic array which is constructed by using the nonlinear transformation the number of PEs is given by the relation:

$$\Omega = n \cdot \left\lfloor \frac{3n-1}{2} \right\rfloor \quad (9)$$

Definition 8: The transformation matrix T maps the index point $(i, j, k) \in P_{ind}$ into the point $(t, x, y) \in T \cdot P_{ind}$, where P_{ind} is the set of index points and

$$Tt = T_1 [i \ \ j \ \ k]^T = i + j + k \quad (10)$$

IV. THE ADVANTAGE OF USING THE LINEAR TRANSFORMATION IN DESIGNING THE SYSTOLIC ARRAY FOR MATRIX MULTIPLICATION

From theorem 2, For N=4 then $\Omega = 37$ (which can be seen from fig.2 too). Because of theorem 3, the number of processors (which can be seen from fig. 1) is $\Omega = 4^2 = 16$. So, we can conclude how the number of processors on the array can be reduced using the linear transformation. For N=4 is 16 vis-à-vis 37 without using the transformation. On the table 3 we give the comparison for number of PE for different values of N. This information is taken from [13].

TABLE 3: COMPARISON FOR NUMBER OF PE

N	Without using L	By using L
5	61	25
10	271	100
50	7351	2500
100	29701	10000

For the pipeline period, in the case of the array in fig. 2, using relation (4) we get

$$\alpha = \lambda^T u = 3 \quad (11)$$

If this formula is used for the systolic array which is constructed by using the linear transformation matrix (the array in fig.1), we get $\alpha = 1$. This means that in the case of fig. 1 the PEs perform in every step.

If k is an index point, then of course, $\max k = (N_1, N_2, N_3)$ and $\min k = (1,1,1)$. Using the relation (10) we have that:

$$\max_{(t,x,y)\in P_{ind}} t \ \ t = N_1 + N_2 + N_3; \ \min_{(t,x,y)\in P_{ind}} t = 1+1+1 = 3 \quad (12)$$

From relations (2) and (12) the execution time may is:

$$T_{exe} = 1 + N_1 + N_2 + N_3 - 3 = N_1 + N_2 + N_3 - 2 \quad (13)$$

If one uses the fact that if $N_1 = N_2 = N_3$, then $T_{exe} = 3N - 2$. On the other hand $T_{in} = T_{out} = N - 1$, so the computation time is:

$$T = 5N - 4 \quad (14)$$

In the case of the array in fig.1 the execution time will be ordered by using the theorem 1:

$$T_{exe} = 1 + (N_1 - 1) \cdot 0 + (N_2 - 1) \cdot 1 + (N_3 - 1) \cdot 1 = N_2 + N_3 - 1 \quad (15)$$

If $N_2 = N_3$ then $T_{exe} = 2N - 1$. In this case $T_{in} = N - 1$ and $T_{out} = 0$, therefore the computational time is

$$T = 3N - 2 \quad (16)$$

In the case of the array which was constructed by using the nonlinear transformation, we have that $T_{exe} = 3N - 1$, $T_{in} = N - 1$ and $T_{out} = 0$. Therefore the computation time is

$$T = 4N - 2 \quad (17)$$

For the geometric area in the case of the array in fig.2, if one takes $N_1 = N_2 = N_3 = N$ then

$$g_a = 3N^2 - 6N + 3 = 3(N-1) \tag{18}$$

For the second case (the array in fig. 1) the geometric area If one takes $N_1 = N_2 = N_3 = N$ will be calculated as

$$g_a = (N-1)^2 \tag{19}$$

In the case of the array with nonlinear transformation, one can calculate the geometric area in a similar way as above

$$g_a = 2(N-1)^2 \tag{20}$$

Since the duration of matrix multiplication on a system with only one processor is $T_1 = N^3$, the speedup and efficiency in the case of the array in fig.2, using relations (6) and (7), will be respectively:

$$S = \frac{N^3}{5N-4} \tag{21}$$

$$E = \frac{N^3}{(5N-4)(3N^2-3N+1)} \left(\lim_{N\to\infty} E = \frac{1}{15} = 6.7\% \right) \tag{22}$$

The same parameters in the case of using linear transformation matrix are:

$$S = \frac{N^3}{3N-2} \tag{23}$$

$$E = \frac{N}{(3N-2)} \left(\lim_{N\to\infty} E = \frac{1}{3} = 33.3\% \right) \tag{24}$$

And finally these parameters for the array where nonlinear transformation has been used are:

$$S = \frac{N^3}{4N-2} \tag{25}$$

$$E = \frac{N^2}{(4N-2)\cdot\left\lfloor \frac{3n-1}{2} \right\rfloor} \left(\lim_{N\to\infty} E = \frac{1}{6} = 16.7\% \right) \tag{26}$$

Using the results obtained by the relations (14-26), as well as theorems 1, 2, and 3, one can construct the corresponding table, where all the results can be compared. In table 2 it is given a comparison of performance characteristics for some values of N.

V. CONCLUSION

In this paper are analyzed some performance measures for parallel matrix multiplication. We emphasized the systolic approach as most efficient. We can conclude that using the identical performance parameters, for each parameter, the array which is constructed using linear transformation matrix has better performances. Especially for the efficiency when N tends to the infinity we have that it is approximately five times better than the array without using the linear transformation. From the table 2 we can deduce the advantage of using the linear transformation.

REFERENCES

[1] Choi, J., J.J. Dongarra and D.W. Walker, Level 3 BLAS for distributed memory concurrent computers, 1992. CNRS-NSF Workshop on Environments and Tools for Parallel Scientific Computing, Saint Hilaire du Touvet, France, Sept. 7-8, Elsevier Sci. Publishers.

[2] Alpatov, P., G. Baker, C. Edwards, J. Gunnels, G. Morrow, J. Overfelt, Robert Van de GEijn and J. Wu, 1997. Plapack: Parallel Linear Algebra Package, Proceedings of the SIAM Parallel Processing Conference.

[3] Agarwal, R.C., S.M. Balle, F.G. Gustavson, M. Joshi and P. Palkar, 1995. A 3-Dimensional Approach to Parallel Matrix Multiplication, IBM J. Res.Develop., Volume 39, Number 5, pp: 1-8, Sept.

[4] Choi, J., J.J. Dongarra and D.W. Walker, 1994. PUMMA: Parallel Universal Matrix Multiplication Algorithms on distributed memory concurrent computers, Concurrency: Practice and Experience, Vol 6(7): 543-570.

[5] Cannon, L.E., 1969. A Cellular Computer to Implement the Kalman Filter Algorithm, Ph.D. Thesis Montana State University.

[6] Chtchelkanova, A., J. Gunnels, G. Morrow, J. Overfelt, R. van de Geijn, 1995. Parallel Implementation of BLAS: Ceneral Techniques for Level 3 BLAS, TR-95-40, Department of Computer Sciences, University of Texas, OCT.

[7] Agarwal, R.C., F.G. Gustavson and M. Zuibar, 1994. A high-performance matrix multiplication algorithm on a distributed memory parallel computer using overlapped communication, IBM J. Res. Develop., Volume 38, Number 6.

[8] Ziad Alqadi and Amjad-Jazzar, 2005. Analysis of program methods used for optimizing matrix multiplication, J.Eng., Vol.15, NO. 1: 73-78.

[9] Anderson, E., Z. Bai, C. Bischof, J.Demmel, J. Dongarra, J. DuCroz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen, 1990 Lapack: A Portable Linear Algebra Library for High Performance Computers, Proceeding of Supercomputing '90, IEEE Press, pp: 1-10.

[10] Edwards, C., P. Geng, A. Patra and R. Vande Geijn, 1995. Parallel matrix distributions: have we been doing it all wrong?, Tech. Report TR-95-40, Dept. of Computer Sciences, The University of Texas at Austin.

[11] Fox, G., S. Otto and A.Hey, 1987. Matrix Algorithms on a Hypercube I: matrix multiplication, Parallel Computing 3, pp:17-31.

[12] C.N. Zhang, J.H. Weston, Y. F. Yan: Determining object functions in systolic array designs. IEEE Trans. VLSI Systems 2, No. 3 (1994), 357-360.

[13] M.P. Bekakos, Highly Parallel Computations-Algorithms and Applications, Democritus University of Thrace, Greece, pp. 139-209, 2001.

[14] Snopce, H., Elmazi, L., Reducing the number of processors elements in systolic arrays for matrix multiplication using linear transformation matrix, Int. J. of Computers, Communications and Control, Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 486-490.

[15] Gusev, M., and Evans, D.J., A new matrix vector Product Systolic Array, Parallel Algorithms and Applications, 22, 346-349, 1994.

Appendix

TABLE 1: THEORETICAL RESULTS

Algorithm	f	c	q	**S**	**E**
Systolic algorithm	55080000	1440000	38.25	3.825	0.956
Cannon's algorithm	271800000	108720000	2.5	0.25	0.0625
Fox's algorithm with square decomposition	162360000	270720000	0.599	0.0599	0.015
Fox's algorithm with scattered decomposition	54360000	109440000	0.497	0.0497	0.0124
PUMMA	54360000	1620000	33.55	3.355	0.839
SUMMA	54360000	1800000	30.2	3.02	0.755
DIMMA	54360000	1800000	30.2	3.02	0.755

TABLE 2: COMPARISON OF PERFORMANCE CHARACTERISTICS

	Without using L			By using L			By nonlinear transf.		
	N=4	N=10	N=100	N=4	N=10	N=100	N=4	N=10	N=100
Ω	37	271	29701	16	100	10000	20	140	14900
T	16	46	496	10	28	298	14	38	398
S	4	21.7	2016	6.4	35.7	3355	4.5	26.3	2512.6
E	10.8%	8%	6.8%	40%	35.7%	33.5%	23%	19%	16.9%
g_a	27	243	29403	9	81	9801	18	162	19602

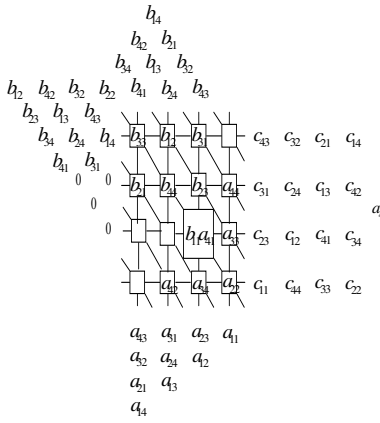


Fig. 1. Systolic array using theorem 3

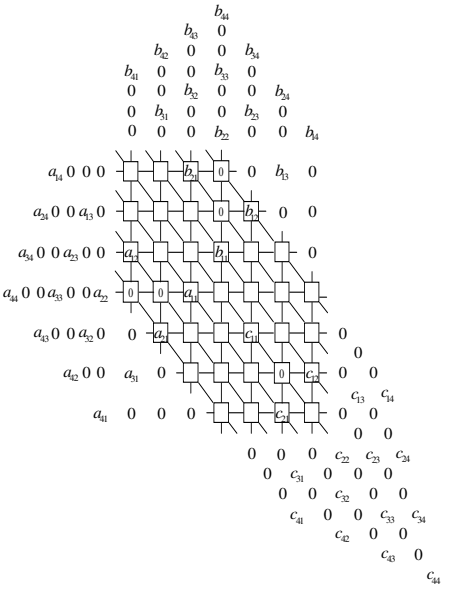


Fig. 2 The SHSA array for N=4

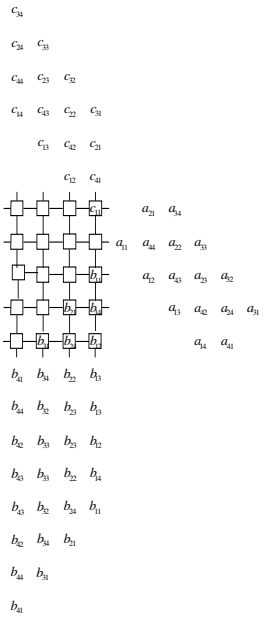


Fig. 3 the systolic with nonlinear mapping

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 515–519

Template Library for Multi-GPU Pseudorandom Number Recursion-based Generators

Dominik Szałkowski
Institute of Mathematics, Maria Curie-Skłodowska University,
Pl. M. Curie-Skłodowskiej 1, Lublin, Poland
Email: dominisz@umcs.lublin.pl

Przemysław Stpiczyński
Maria Curie-Skłodowska University, Lublin, Poland
Institute of Theoretical and Applied Informatics of
the Polish Academy of Sciences, Gliwice, Poland
Email: przem@hektor.umcs.lublin.pl

***Abstract*—The aim of the paper is to show how to design and implement fast parallel algorithms for Linear Congruential, Lagged Fibonacci and Wichmann-Hill pseudorandom number generators. The new algorithms employ the *divide-and-conquer* approach for solving linear recurrence systems. They are implemented on multi GPU-accelerated systems using CUDA. Numerical experiments performed on a computer system with two Fermi GPU cards show that our software achieve good performance in comparison to the widely used NVIDIA CURAND Library.**

I. Introduction

PSEUDORANDOM numbers are very important in practice and pseudorandom number generators are often central parts of scientific applications such as simulations of physical systems. They are used by Monte Carlo methods, especially in case of multidimensional numerical integration [1], [4], [9]. In [8] we showed the general techniques for implementing recursion-based generators of pseudorandom numbers on GPU-accelerated systems which are much more efficient than their sequential counterparts.

NVIDIA CURAND Library [5] provides routines for simple and efficient generation of high-quality random numbers. It comprises two types of generators:

- XORWOW, MRG32K3A and MTGP32 are *pseudorandom number generators* which means that a sequence of random numbers which they produce satisfy most of desired statistical properties of a truly random sequence and they work on 32-bit numbers,
- SOBOL32, SCRAMBLED_SOBOL32, SOBOL64, SCRAMBLED_SOBOL64 are *quasirandom number generators*, n-dimensional points obtained from these fill n-dimensional space evenly, SOBOL32, SCRAMBLED_SOBOL32 use 32-bit arithmetic and SOBOL64, SCRAMBLED_SOBOL64 use 64-bit arithmetic.

Unfortunately, these generators utilize only a single GPU device, thus if we want to perform computations using multiple GPUs, we should apply some parametrization techniques for parallel generation of pseudorandom numbers [3] what can lead to possible unwanted correlations between numbers resulting in their poor statistical properties [6]. It should be noticed that only one generator from CURAND produces fully 64-bit results.

In this paper we show how to design fast parallel algorithms for Linear Congruential, Lagged Fibonacci [3] and Wichmann-Hill [10] pseudorandom number generators which employ the *divide-and-conquer* approach for solving linear recurrence systems [7] and can be easily used in computations on multi-GPU systems. Our generators have exactly the same statistical properties as their sequential counterparts.

Numerical experiments performed on a computer system with two Fermi GPU cards show that they achieve good speedup in comparison to the standard CPU-based sequential algorithms [8] and implementations provided by NVIDIA CURAND Library. Our implementation is freely available as the C++ template library which requires only CUDA Toolkit. It can be downloaded from `http://dominisz.umcs.lublin.pl/gpu-rand`.

II. Parallel Pseudorandom Number Generators

We consider the following three pseudorandom number generators:

1) **Linear Congruential Generator (LCG)**: $x_{i+1} \equiv (ax_i + c)(\mod m)$, where x_i is a sequence of pseudorandom values, $m > 0$ is the *modulus*, a, $0 < a < m$ is the *multiplier*, c, $0 \leq c < m$ is the *increment*, x_0, $0 \leq x_0 < m$ is the *seed* or *start value*,
2) **Lagged Fibonacci Generator (LFG)**: $x_i \equiv (x_{i-p_1} + x_{i-p_2})(\mod m)$, where $0 < p_1 < p_2$,
3) **Wichmann-Hill Generator (WHG, [10])**:

$$\begin{aligned} x_i &\equiv 11600x_{i-1}(\mod 2147483579) \\ y_i &\equiv 47003y_{i-1}(\mod 2147483543) \\ z_i &\equiv 23000z_{i-1}(\mod 2147483423) \\ t_i &\equiv 33000t_{i-1}(\mod 2147483123) \\ W &\equiv x_i/2147483579.0 + y_i/2147483543.0 \\ &\quad + z_i/2147483423.0 + t_i/2147483123.0 \\ W_i &\equiv W - \lfloor W \rfloor. \end{aligned} \tag{1}$$

It should be noted that, in fact, WHG combines four LCG generators, each with the increment 0 (such generator is also called Multiplicative Congruential Generator, MCG). This generator has much better statistical properties than LCG. Its period is about 2^{121}. It passes Big Crush test from TestU01 Library [2].

978-83-60810-53-8/$25.00

In case of LCG and LFG, $m = 2^M$, where $M = 32$ or $M = 64$, thus these generators produce numbers from $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$. It allows the modulus operation to be computed by merely truncating all but the rightmost 32 or 64 bits, respectively. Thus, when we use `unsigned int` or `unsigned long int` data types, we can neglect "$(\mod m)$". In case of WHG, we have moduli given explicitly. Note that the integers x_k are between 0 and $m-1$. They can be converted to real values $r_k \in [0, 1)$ by $r_k = x_k/m$.

It is clear that LCG, LFG, WHG generators can be considered as special cases of linear recurrence systems [7]. Indeed, LCG can be defined as

$$\begin{cases} x_0 = d \\ x_{i+1} = ax_i + c, \quad i = 0, \ldots, n-2, \end{cases} \tag{2}$$

and similarly for LFG we have

$$\begin{cases} x_i = d_i & i = 0, \ldots, p_2 - 1 \\ x_i = x_{i-p_1} + x_{i-p_2}, & i = p_2, \ldots, n-1. \end{cases} \tag{3}$$

The details of our single-GPU implementations of LCG and LFG generators can be found in [8]. Here we only recall the most important formulas. The parallel version LCG can be expressed as follows

$$\begin{cases} \mathbf{x}_0 = A^{-1}\mathbf{f}_0 \\ \mathbf{x}_i = \mathbf{t} + x_{is-1}\mathbf{y}, \quad i = 1, \ldots, r-1, \end{cases} \tag{4}$$

where $\mathbf{x}_i = (x_{is}, \ldots, x_{(i+1)s-1})^T \in \mathbb{Z}_m^s$, $\mathbf{f}_0 = (d, c, \ldots, c)^T \in \mathbb{Z}_m^s$, $\mathbf{f} = (c, \ldots, c)^T \in \mathbb{Z}_m^s$, and

$$A = \begin{bmatrix} 1 & & & \\ -a & 1 & & \\ & \ddots & \ddots & \\ & & -a & 1 \end{bmatrix} \in \mathbb{Z}_m^{s \times s}.$$

Moreover $\mathbf{t} = A^{-1}\mathbf{f}$ and $\mathbf{y} = A^{-1}(a\mathbf{e}_0)$, where $\mathbf{e}_0 = (1, 0, \ldots, 0)^T \in \mathbb{Z}_m^s$.

Similarly, for LFG we have

$$\begin{cases} \mathbf{x}_0 = A_0^{-1}\mathbf{f} \\ \mathbf{x}_i = \sum_{k=0}^{p_2-1} x_{is-p_2+k}\mathbf{y}_k + \sum_{k=0}^{p_1-1} x_{is-p_1+k}\mathbf{y}_k, \\ \qquad\qquad\qquad\qquad i = 1, \ldots, r-1, \end{cases} \tag{5}$$

where matrix A_0 and vectors $\mathbf{f}$, $\mathbf{y}_k$ are defined analogously as for LCG case (see [8] for details). Note that (5) is the generalization of (4).

III. Multi-GPU implementation

To implement the parallel algorithms efficiently on GPU, we will form the following matrix

$$Z = [\mathbf{x}_0, \ldots, \mathbf{x}_{r-1}] \in \mathbb{Z}_m^{s \times r}, \tag{6}$$

where all vectors $\mathbf{x}_i$ are defined by (4) or (5). This allows to use fast coalesced memory access and makes possible the use of shared memory.

The equation (4) has a lot of potential parallelism. The algorithm comprises the following steps. First (Step 1) we have to find $\mathbf{y}$, $\mathbf{t}$. Then (Step 2) we find the last entry of each vector $\mathbf{x}_i$, $i = 1, \ldots, r-1$. Finally (Step 3), we find $s-1$ entries of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_{r-1}$ in parallel. In case of multi-GPU implementation vectors $\mathbf{y}$, $\mathbf{t}$ are computed by CPU and then sent to all GPU devices. The generator seed required to compute Step 2 is received from the previous GPU device and sent to the next one after Step 2 is completed locally (Figure 1). Then all GPUs perform Step 3 independently.

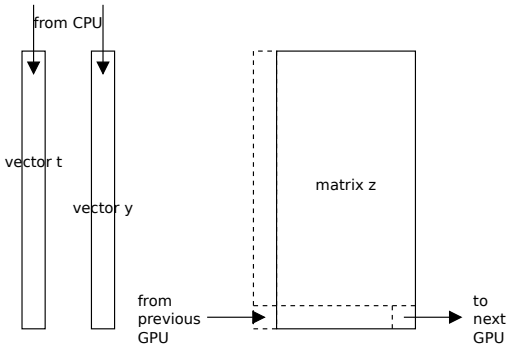


Fig. 1. LCG: data structures on a GPU device and communication scheme

We can develop a similar parallel algorithm for LFG. During the first step we have to find vector $\mathbf{y}_0$. This vector is computed by CPU and sent to all GPUs. It is easy to verify that

$$\mathbf{y}_k = (\underbrace{0, \ldots, 0}_{k}, 1, y_1, \ldots, y_{s-1-k})^T.$$

Then (Step 2) using (5) we find p_2 last entries of $\mathbf{x}_1, \ldots, \mathbf{x}_{r-1}$. Finally (Step 3) we use (5) to find $s - p_2$ first entries of these vectors in parallel. Note that Step 2 requires communication (sending and receiving the seed consisting of p_2 numbers) between GPU devices (Figure 2).

The parallel algorithm for WHG is a simple extensions of the parallel LCG. Instead of vectors $\mathbf{t}$ and $\mathbf{y}$, we have four instances of $\mathbf{y}$, each for one MCG. We also have four separate "last rows" of matrix Z, corresponding to appropriate MCG, which are required during Step 2 (Figure 3).

IV. Results of experiments

The considered algorithms have been tested on a computer with Intel Xeon X5650 (2.67 GHz, 48GB RAM) and NVIDIA Tesla M2050 (448 cores, 3GB GDDR5 RAM with ECC off), running under Linux with `gcc` and NVIDIA `nvcc` compilers and CURAND Library ver. 5.0 provided by the vendor. The results of experiments are presented in Figures 4-6. We can conclude the following:

- Parallel LCG is the fastest among the considered generators. It produces $\approx 32 \cdot 10^6$ `unsigned int` pseudorandom numbers per second, while the fastest CURAND pseudorandom generator achieves the speed of $\approx 12 \cdot 10^6$.

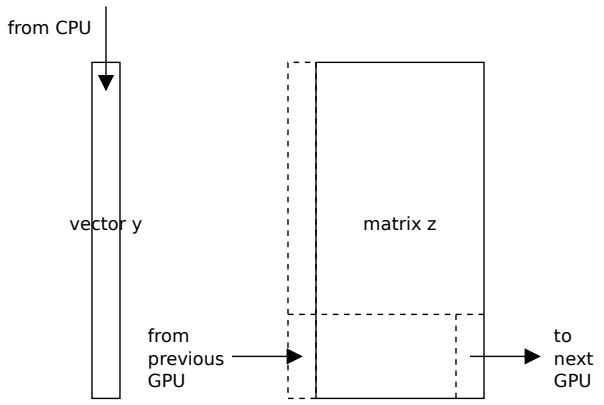


Fig. 2. LFG: data structures on a GPU device and communication scheme

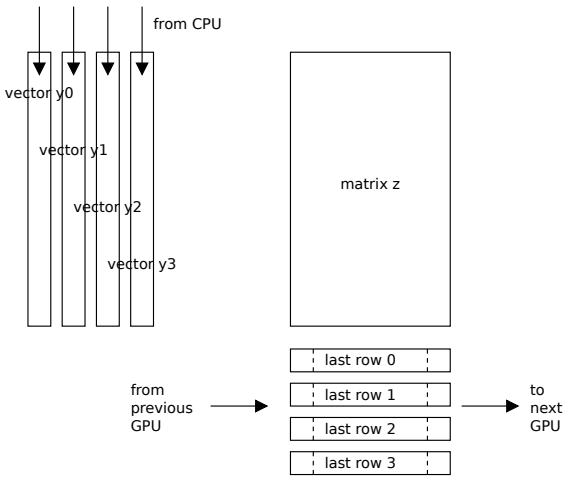


Fig. 3. WHG: data structures on a GPU device and communication scheme

- Parallel WHG is about 6 times slower than LCG. However it uses more computations and communications in comparison to LCG. It also has better statistical properties, so it should be use instead of LCG, when the performance is not so important.
- The performance of parallel LFG depends on the values of p_1 and p_2.
- Our template library provides both 32-bit and 64-bit versions of all generators. CURAND does not support 32-bit or 64-bit arithmetic in all cases (hence missing bars in Figure 4).
- The use of two GPUs accelerates the overall time of computations (Figure 6). In case of LCG we obtain almost linear speedup. The scalability of WHG is worse because of longer lasting Step 2. Unfortunately, the scalability of LFG is poor for large values of p_1, p_2.

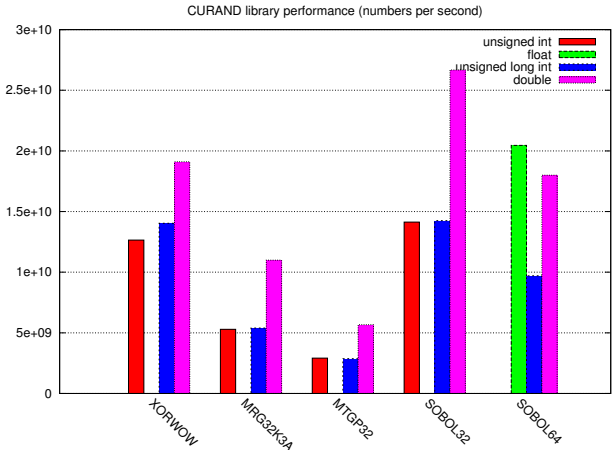


Fig. 4. CURAND Library performance: generation of random number using various generators

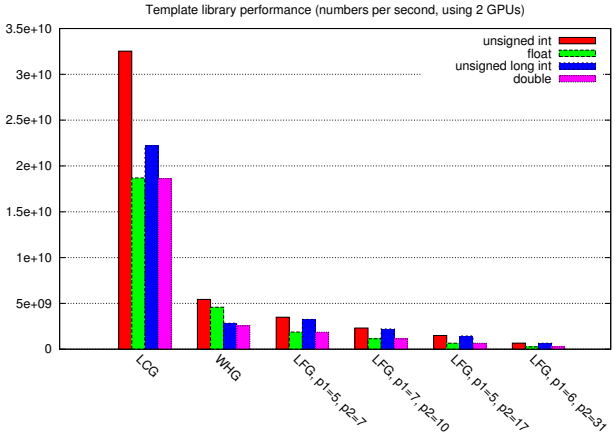


Fig. 5. Our template library performance: generation of random number using various generators

V. Using template library

Let us consider the use our template library in case of LCG. The following class template should be used (we only show `public:` part of it).

```
template <class T>
class LcgGpu {
  public:
    LcgGpu(T multiplier, T increment,
           T seed, size_t count);
    void generate();
    void generateFloat();
    void generateDouble();
    T* getNumbersFromDevice(int device);
    size_t getCountFromDevice(int device);
    int getDeviceCount();
    ...
}//class LcgGpu
```

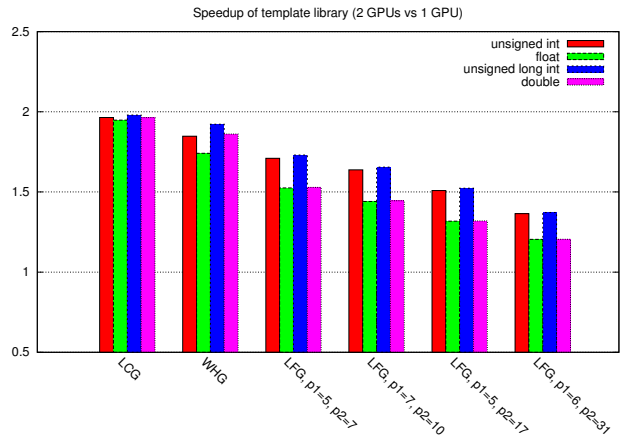


Fig. 6. Our template library: speedup 2 GPUs vs 1 GPU

In order to use the generator we should create an object providing desired parameters (multiplier, increment, seed of the generator and the number of pseudorandom numbers to generate).

```
unsigned int multiplier=1664525;
unsigned int increment=1013904223;
unsigned int seed=31;
size_t count=100000000;

LcgGpu<unsigned int> lcg
  =new LcgGpu<unsigned int>(multiplier,
                            increment,
                            seed, count);
```

Then we generate random numbers (e.g. uniformly distributed real numbers from interval $[0, 1)$).

```
lcg->generateFloat();
```

Generated numbers are stored in global memories of all GPU devices. We use the following routines to obtain the number of GPU devices, which produce numbers, the number of pseudorandom numbers generated by a given device and the address of memory block containing the numbers:

```
int getDeviceCount();
size_t getCountFromDevice(int device);
T* getNumbersFromDevice(int device);
```

Generated numbers can be used directly by each GPU or can be transferred to CPU memory using loop for accessing all devices.

```
for (i=0; i<lcg->getCountFromDevice(); i++) {
        cudaMemcpy(cpuNumbers+offset,
          lcg->getNumbersFromDevice(i),
          lcg->getCountFromDevice(i)
            *sizeof(unsigned int),
          cudaMemcpyDeviceToHost);
    offset=offset
      +lcg->getCountFromDevice(i);
}//for
```

Generation of numbers can be repeated as many times as desired to obtain very long sequence of pseudorandom numbers. Finally we can delete the object.

```
delete lcg;
```

Using LFG generator is quite similar. The only difference is when the object is created. For example, we can use the following code.

```
unsigned int p1=24;
unsigned int p2=55;
unsigned int seed[]={...}; //array of
                           //length p2
size_t count=100000000;

LfgGpu<unsigned int> lfg=
  new LfgGpu<unsigned int>(p1, p2
                           seed, count);
```

Analogously for WHG we use the following.

```
unsigned int seedX=389933028;
unsigned int seedY=148667295;
unsigned int seedZ=146045161;
unsigned int seedT=767880647;

WhgGpu<unsigned int> whg=
  new WhgGpu<unsigned int>(seedX, seedY,
                           seedZ, seedT,
                           count);
```

When we parametrize template with `unsigned int` type then we can use the following routines to generate 32-bit pseudorandom numbers (integer or real numbers).

```
void generate();
void generateFloat();
```

When we need 64-bit precision we use `unsigned long int` type to parametrize template and the following routines.

```
void generate();
void generateDouble();
```

VI. Conclusions

We have showed how to implement fast parallel LCG, LFG and WHG pseudorandom number generators using the *divide-and-conquer* approach on contemporary multi-GPU systems. Numerical experiments performed on a computer system with modern Fermi GPU cards showed that our routines achieve good performance in comparison to the widely used NVIDIA CURAND Library. Our template library is easy to use and it is freely available for the community.

ACKNOWLEDGEMENTS

The work has been prepared using the supercomputer resources provided by the Institute of Mathematics of the Maria Curie-Skłodowska University in Lublin.

REFERENCES

[1] J. M. Bull and T. L. Freeman, "Parallel globally adaptive quadrature on the KSR-1," *Adv. Comput. Math.*, vol. 2, pp. 357–373, 1994. [Online]. Available: http://dx.doi.org/10.1007/BF02521604

[2] P. L'Ecuyer and R. J. Simard, "TestU01: A c library for empirical testing of random number generators," *ACM Trans. Math. Softw.*, vol. 33, no. 4, 2007. [Online]. Available: http://doi.acm.org/10.1145/1268776.1268777

[3] M. Mascagni and A. Srinivasan, "Algorithm 806: SPRNG: a scalable library for pseudorandom number generation," *ACM Trans. Math. Softw.*, vol. 26, no. 3, pp. 436–461, 2000.

[4] H. Niederreiter, "Quasi-Monte Carlo methods and pseudo-random numbers," *Bull. Am. Math. Soc.*, vol. 84, pp. 957–1041, 1978.

[5] NVIDIA, *CUDA Toolkit 5.0. CURAND Guide*. NVIDIA Corporation, 2012.

[6] A. Srinivasan, M. Mascagni, and D. Ceperley, "Testing parallel random number generators," *Parallel Computing*, vol. 29, no. 1, pp. 69–94, 2003.

[7] P. Stpiczyński, "Solving linear recurrence systems on hybrid GPU accelerated manycore systems," in *Proceedings of the Federated Conference on Computer Science and Information Systems, September 18-21, 2011, Szczecin, Poland*. IEEE Computer Society Press, 2011, pp. 465–470. [Online]. Available: http://fedcsis.eucip.pl/proceedings/pliks/148.pdf

[8] P. Stpiczyński, D. Szałkowski, and J. Potiopa, "Parallel GPU-accelerated recursion-based generators of pseudorandom numbers," in *Proceedings of the Federated Conference on Computer Science and Information Systems, September 9-12, 2012, Wroclaw, Poland*. IEEE Computer Society Press, 2012, pp. 571–578. [Online]. Available: http://fedcsis.org/proceedings/fedcsis2012/pliks/380.pdf

[9] D. Szałkowski and P. Stpiczyński, "Multidimensional monte carlo integration on clusters with hybrid gpu-accelerated nodes," 2013, submitted to PPAM2013.

[10] B. A. Wichmann and I. D. Hill, "Generating good pseudo-random numbers," *Comput. Stat. Data Anal.*, vol. 51, no. 3, pp. 1614–1622, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.csda.2006.05.019

International Symposium on Multimedia Applications and Processing

ORGANIZED BY

Software Engineering Department, Faculty of Automation, Computers and Electronics, University of Craiova, Romania "Multimedia Applications Development" Research Centre

Background and Goals

Multimedia information has become ubiquitous on the web, creating new challenges for indexing, access, search and retrieval. Recent advances in pervasive computers, networks, telecommunications, and information technology, along with the proliferation of multimedia mobile devices - such as laptops, iPods, personal digital assistants (PDA), and cellular telephones - have stimulated the development of intelligent pervasive multimedia applications. These key technologies are creating a multimedia revolution that will have significant impact across a wide spectrum of consumer, business, healthcare, educational and governmental domains. Yet many challenges remain, especially when it comes to efficiently indexing, mining, querying, searching, retrieving, displaying and interacting with multimedia data.

The Multimedia - Processing and Applications 2013 (MMAP 2013) Symposium addresses several themes related to theory and practice within multimedia domain. The enormous interest in multimedia from many activity areas (medicine, entertainment, education) led researchers and industry to make a continuous effort to create new, innovative multimedia algorithms and applications.

As a result the conference goal is to bring together researchers, engineers, developers and practitioners in order to communicate their newest and original contributions. The key objective of the MMAP conference is to gather results from academia and industry partners working in all subfields of multimedia: content design, development, authoring and evaluation, systems/tools oriented research and development. We are also interested in looking at service architectures, protocols, and standards for multimedia communications - including middleware - along with the related security issues, such as secure multimedia information sharing. Finally, we encourage submissions describing work on novel applications that exploit the unique set of advantages offered by multimedia computing techniques, including home-networked entertainment and games. However, innovative contributions that don't exactly fit into these areas will also be considered because they might be of benefit to conference attendees.

Call for Papers

MMAP 2013 is a major forum for researchers and practitioners from academia, industry, and government to present, discuss, and exchange ideas that address real-world problems with real-world solutions.

The MMAP 2013 Symposium welcomes submissions of original papers concerning all aspects of multimedia domain ranging from concepts and theoretical developments to advanced technologies and innovative applications. MMAP 2013 invites original previously unpublished contributions that are not submitted concurrently to a journal or another conference.

Papers acceptance and publication will be judged based on their relevance to the symposium theme, clarity of presentation, originality and accuracy of results and proposed solutions.

Topics

Topics of interest are related to Multimedia Processing and Applications including, but are not limited to the following areas:

- Audio, Image and Video Processing
- Animation, Virtual Reality, 3D and Stereo Imaging
- Multimedia File Systems and Databases: Indexing, Recognition and Retrieval
- Machine Learning, Data Mining, Information Retrieval in Multimedia Applications
- Multimedia in Internet and Web Based Systems:
  - E-Learning, E-Commerce and E-Society Applications
- Human Computer Interaction and Interfaces in Multimedia Applications
- Multimedia in Medical Applications
- Entertainment and games
- Security in Multimedia Applications: Authentication and Watermarking
- Distributed Multimedia Systems
- Network and Operating System Support for Multimedia
- Mobile Network Architecture
- Intelligent Multimedia Network Applications

General Chair

Burdescu, Dumitru Dan, University of Craiova, Romania

Steering Committee

Badica, Costin, University of Craiova, Romania
Deserno, Thomas M., Aachen University, Germany
Furht, Borko, Florida Atlantic University, USA
Kosch, Harald, University of Passau, Germany
Obaidat, Mohammad S., Monmouth University, USA
Pitas, Ioannis, University of Thessaloniki, Greece
Uskov, Vladimir, Bradley University, USA

Organizing Committee

Badica, Costin, University of Craiova, Romania
Brezovan, Marius, University of Craiova, Romania

Burdescu, Dumitru Dan, University of Craiova, Romania

Mihaescu, Cristian Marian, University of Craiova, Romania

Stanescu, Liana, University of Craiova, Romania,

Publicity Chair

Badica, Amelia, University of Craiova, Romania

Burlea Schiopoiu, Adriana, University of Craiova, Romania

Program Committee

Badica, Amelia, University of Craiova, Romania

Böszörmenyi, Laszlo, Klagenfurt University, United States

Camacho, David, Universidad Autonoma de Madrid, Spain

Cano, Alberto, University of Cordoba, Spain

Cardoso, Jaime S., Universidade do Porto, Portugal

Cretu, Vladimir, Politehnica University of Timisoara, Romania

Debono, Carl James, University of Malta, Malta

Fomichov, Vladimir, State University - Higher School of Economics, Russia

Giurca, Adrian, Brandenburg University of Technology, Germany

Grosu, Daniel, Wayne State University, United States

Groza, Voicu, University of Ottawa, Canada

Grundspenkis, Janis, Riga Technical University, Latvia

Hendrix, Maurice, Coventry University, United Kingdom

Kannan, Rajkumar, Bishop Heber College Autonomous, India

Korzhik, Valery, State University of Telecommunications, Russia

Kotenko, Igor, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Russia

Kriksciuniene, Dalia, Vilnius University, Lithuania

Lamas, David, Tallin University, Estonia

Lau, Rynson, City University of Hong Kong, Hong Kong S.A.R., China

Lloret, Jaime, Polytechnic University of Valencia, Spain

Logofatu, Bogdan, University of Bucharest, Romania

Luna, Jose, University of Cordoba, Spain

Mangioni, Giuseppe, DIEEI - University of Catania, Italy

Miyata, Hitoshi, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Japan

Mocanu, Mihai, University of Craiova, Romania

Morales-Luna, Guillermo, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico

Ogiela, Marek, AGH University of Science and Technology, Poland

Ohzeki, Kazuo, Shibaura Institute of Technology, Japan

Õunapuu, Enn, Tallinn University of Technology, Estonia

Paltoglou, Georgios, University of Wolverhampton, United Kingdom

Popescu, Dan, CSIRO, Spain

Salem, Abdel-Badeeh M., Ain Shams University, Egypt

Sari, Riri Fitri, University of Indonesia, Indonesia

Smedberg, Asa, Stockholm University, Sweden

Tejera, Mario Hernández, University of Las Palmas de Gran Canaria, Spain

Trausan-Matu, Stefan, Politehnica University of Bucharest, Romania

Trzcielinski, Stefan, Poznan University of Technology, Poland

Tsihrintzis, George, University of Piraeus, Greece

Vega-Rodríguez, Miguel A., University of Extremadura, Spain

Velastin, Sergio, Kingston University, United Kingdom

Wotawa, Franz, Technische Universitaet Graz, Austria

Zurada, Jacek, University of Louisville, United States

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 523–527

Design of Digital Watermarking System Robust to the Number of Removal Attacks

Sergey Anfinogenov
State University of Telecommunications,
St. Petersburg, Russia
Email: serganff@gmail.com

***Abstract*—In this paper it is proved that in fact the zero-bit digital watermarking system based on local maxima embedding in frequency area heuristically proposed recently is resistant to a number of removal attacks. It is shown how the watermark can survive after such conversions as shift cropping rescaling rotation and jpeg transform. The theoretical base of each transformation is given. Also it is shown how the image Fourier amplitude spectrum is affected by the image distortions and how the watermark can overcome those distortions and stay untouched.**

I. Introduction

THE MAIN idea of the watermarking method offered in [1] is an embedding of a zero-bit watermark (identification key) into the positions of maxima of the local areas, which are selected in the amplitude spectrum of the two dimensional discrete Fourier transform (DFT), calculated from the original image.

Now let us remember how this algorithm works step-by-step. First we generate a binary key K, which can be represented as the two dimensional matrix $K(n,m)$ where the number of columns N and rows M is equal to the width and height of the image respectively. Then we calculate the DFT of the image and get the amplitude spectrum. Next we change the amplitude spectrum according to the rule: If $K(n,m)=1$ we build the local area $(n-a..n+a, m-a..m+a)$ with the size (2a+1)x(2a+1) around this point, where a is a constant value which determines local area size. Then maximum of each local area is calculated. This maximum is multiplied by β value ($\beta > 1$) and placed in the point $K(n,m)$. Later we combine this new amplitude spectrum with a phase untouched before and perform the inverse DFT to get the watermarked image.

During the extraction process we calculate the amplitude again and using previously saved key K build the same local areas and verify if the maximum of each area is situated in the point $K(n,m)$. Next we count all positive answers and divide this value by the total number of local areas. Percepts of watermark are recognised if this value exceeds some threshold. If the watermarked image was untouched there would be no errors and all key points would be recognised. If some attack is applied to the watermark image, then some maxima can be lost, but the watermark will still be detectable sometimes.

The current method of zero-bit WM embedding and extraction seems to be robust against such transforms of an image as cyclic shifting, rotation, removal of rows and columns, noise addition, JPEG transform and cropping, but these conclusions have been based on simulation. In the next section we are going to present the proof of this claim based on the properties of DFT.

II. The proof of robustness of the proposed zero-bit WM system to different attacks

Now let us concentrate on the robustness of the algorithm and answer two main questions. After what image distortions a watermark can survive and why? The direct and inverse Fourier transforms for 2D signal $h(n,m)$(Image in our case) with N columns and M rows are as usually given by:

$$F(h) = \widehat{h}(k,l) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} e^{-i(\omega_k n+\omega_l m)} h(n,m) \quad (1)$$

$$h(F) = \frac{1}{NM}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1} e^{-i(\omega_k n+\omega_l m)} \widehat{h}(k,l) \quad (2)$$

Often it is convenient to express frequency in vector notation with $\vec{k} = (k,l)^t$, $\vec{n} = (n,m)^t$, $\vec{\omega}_{kl} = (\omega_k, \omega_l)^t$ and $\vec{\omega}^t\vec{n} = \omega_k n + \omega_l m$. The vector form will help us when we talk about DFT properties. In this section we will show

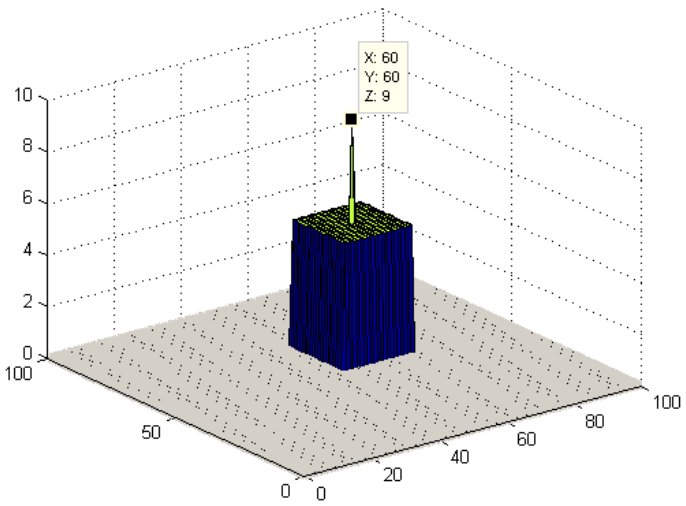


Fig. 1. Test amplitude spectrum

978-83-60810-53-8/$25.00

how the proposed watermarking algorithm can stand against different transformations. The set of transformations that are usually performed to remove a watermark are described in [5] and [4]. Now let us discuss each transformation one by one. To show an effect of each transformation we generate a test amplitude spectrum Fig. 1 where we have only one local area with maximum in the center. Such model differs from the real situation where there are many local maxima, but this simplified model helps us to show how each transformation affects on the behaviour of local maxima positions in each local area.

A. Translation

Using the shift property of the Fourier transform

$$F[f(\vec{x}-\vec{x}_0)] = \exp(-i)\vec{\omega}^t\vec{x}_0)\widehat{f}(\vec{\omega}) \tag{3}$$

it is easy to see that only phase of the DFT is affected by the translation of the image. The amplitude spectrum where the watermark is embedded remains untouched. So that transformation has no impact on a watermark detection.

B. Rotation

According to FFT property a rotation of the image causes the rotation of the FFT amplitude.

$$F(x,y) \rightarrow F(x\cos\theta + y\sin\theta, x\sin\theta + y\cos\theta) \tag{4}$$

To overcome a rotation problem if the watermark is not found initially the detection process is repeated after rotation of the image on a small angle. Another solution can be used with a normalisation algorithm described in [3] where the image is converted to the domain invariant to rotation. In fact, the image rotation on more than 10 degrees can be distinguished from the original. So it is possible to reduce number of calculations and image rotations. The last way to deal with rotation is to extend the size of local areas and detect maxima not in one certain point but in several points around the embedded maxima. That will help especially in case of small rotation angles.

C. Noise Addition

Let us define a set of n points $x_1, x_2, x_3, ...x_n$ with constant amplitude A and a point $x0$ with amplitude $\beta A(\beta > 1)$. At all points we add zero mean i.i.d Gaussian noise. The probability that the maximum stay in the previous position after the addition of noise is the following:

$$P = Pr(\tilde{x_0} \geq x_1, \tilde{x_0} \geq x_2 ... \tilde{x_0} \geq x_n) = ? \tag{5}$$

where

$\tilde{x_0} = \beta A + n_0, \tilde{x_i} = A + n_i$

$i = 1, 2..n..n_i \in i.i.dN(0, \sigma^2)$

It is easy to see that:

$$P = \int_{-\infty}^{+\infty} \omega_0(y) \prod_{i=1}^{n} (P_\tau(x_i) \leq y)dy \tag{6}$$

where

$$Pr\{\tilde{x} \leq y\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} e^{-\frac{t-A^2}{2\sigma^2}} dt, \tag{7}$$

$$\omega_0(y) = \frac{1}{\sqrt{2\pi, \sigma^2}} e^{-\frac{(y-\beta A)^2}{2\sigma^2}} \tag{8}$$

Substituting (7) and (8) in (6) we get:

$$P = \frac{1}{\sqrt{2\pi, \sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(y-\beta y)^2}{2\sigma^2}} \cdot (\int_{-\infty}^{y} e^{-\frac{(t-A)^2}{2\sigma^2}})^n dy, \tag{9}$$

It is easy to find the lower bound of that probability using the equation:

$$P \geq \prod_{i=1}^{n} P_\tau\{\tilde{x_0} \geq \tilde{x_i}\} = (P_\tau\{\tilde{x_0} \geq \tilde{x_i}\})^n \tag{10}$$

where $(P_\tau\{\tilde{x_0} \geq \tilde{x_i}\})^n = (P_\tau\{\tilde{x_0} - \tilde{x_i}\} \geq 0)^n$

$$(P_\tau\{\tilde{x_0} - \tilde{x_i}\} \geq 0)^n = (\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{t-A(\beta-1)}{2\sigma^2}} dt)^n \tag{11}$$

But unfortunately it is the most interesting for us to find the upper bound of that probability, because we want to know when the local maximum changes its position. Taking into account that a calculation by (9) is very tedious procedure, we can try to solve it by simulation. Fig. 2 shows the effect of noise addition and Table I demonstrates the results of correct maxima recognition for $A = 100$, $\beta = 1.5$, $\sigma = 0.097927$.In the similar manner we can calculate the results for other embedding parameters.

We can see from Table I that maxima are recognised whenever signal-to-noise ratio $\frac{\beta^2}{\sigma^2}$ is greater than 0.49808 ($\sigma < 2.1254$).

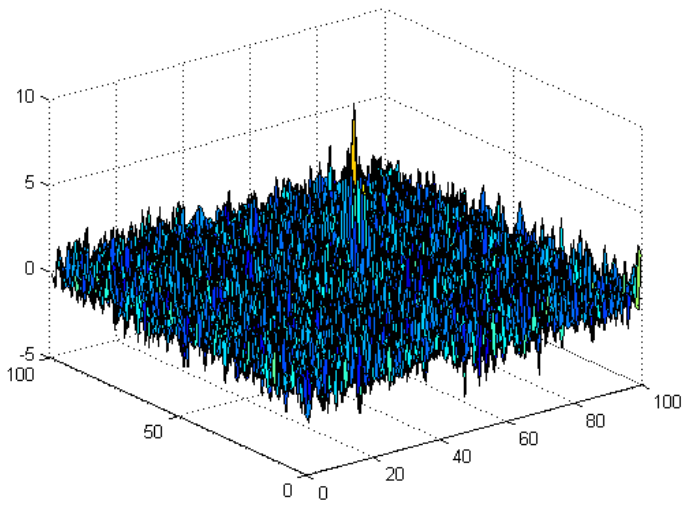


Fig. 2. Amplitude spectrum after noise addition for $A = 100$, $\beta = 1.5$, $\sigma = 0.097927$

TABLE I
THE RESULTS OF CORRECT MAXIMUM RECOGNITION AFTER NOISE ADDITION PERFORMED BY SIMULATION

Variance	Detected
0.097927	Y
0.19737	Y
0.26537	Y
0.38688	Y
0.56174	Y
0.56124	Y
0.72913	Y
0.72439	Y
0.9224	Y
0.98988	Y
1.2446	Y
1.292	Y
1.3197	Y
1.5344	Y
1.6075	Y
1.8164	Y
2.1254	N
1.8201	N
2.1679	N
2.259	Y
2.225	N
2.4485	N
2.5048	N

D. Cropping

During the cropping process some parts of the image are removed, and as the result some frequency components can be changed. Let's analyse this process in more details.

We can present cropping of the image as a multiplication of window by raster image. That is represented in one dimensional form (for the simulation) as one local area of the image amplitude Fig. 4 where cropping is given by the rectangular window function. According to the convolution theorem of the Fourier transform [2] the Fourier transform of the product of the two functions is equal to the convolution of their individual transforms.

So we get:

$$f(n,m)h(n,m) \rightarrow F\left(n',k'\right) * *H\left(n',k'\right) \tag{12}$$

where $f(n,m)$ - raster image,
$h(n,m)$ - the window of cropping. So now we can look on those functions separately.

In the frequency domain window function (in the 1-D case) is defined as follows:

$$h(t) = \frac{\sin\frac{\omega t}{2}}{\frac{\omega t}{2}} \tag{13}$$

The frequency ω is defined by the size of the window. Let's calculate the convolution between $h(t)$ and the test function with one local area (rectangular impulse with the maxima in the center) as follows:

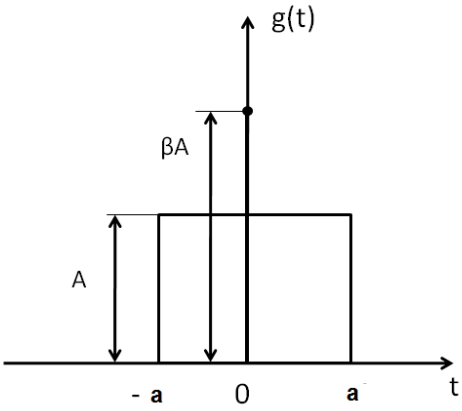


Fig. 3. Local area of the amplitude spectrum

$$y(t) = g(t) * h(t) \tag{14}$$

$$y(t) = FT[I_w(x,y)Rect(c_x(x-x_0), c_y(y-y_0))] \tag{15}$$

where c_x, c_y, x_0, y_0 - cropping parameters.

$$y(t) = \frac{1}{c_x c_y} I'_w(u,v) * e^{-i2\pi(c_x x_0 + c_y y_0)} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} \times sinc\frac{\pi u}{c_x} sinc\frac{\pi v}{c_y} \tag{16}$$

where $sinc(x) = \frac{sin(x)}{x}$ if $x \neq 0$, $sinc(x) = 1$ if $x = 0$.

Now we can represent $I'_w(u,v)$ as the sum of amplitude of the original image and key $K(u,v)$ multiplied by β', where β' is the max value of local area multiplied by a constant β.

$$y(t) = [K(u,v)\beta' + I(u,v)] * * \frac{e^{-i2\pi(c_x x_0 + c_y y_0)}}{c_x c_y} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} sinc\frac{\pi u}{c_x} sinc\frac{\pi v}{c_y} \tag{17}$$

To make the equation more simple we will denote the expression $\frac{e^{-i2\pi(c_x x_0 + c_y y_0)}}{c_x c_y} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} sinc\frac{\pi u}{c_x} sinc\frac{\pi v}{c_y}$ as E and use the distributivity property.

$$y(t) = [K(u,v)\beta'] * E + [I(u,v)] * E \tag{18}$$

The key $K(u,v)$ can have only two values 0 and 1.

If $K(u,v) = 0$ then $y(t) = I(u,v) * E$

else $y(t) = \beta' * E + I(u,v) * E$

If we want the maxima to survive the value of the amplitude in where $K(u,v) = 1$ should be greater than the other points.

$$\beta' * E + [I(u,v)] * E > [I(u,v)] * E \tag{19}$$

$$\beta' * \frac{e^{-i2\pi(c_x x_0 + c_y y_0)}}{c_x c_y} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} sinc\frac{\pi u}{c_x} sinc\frac{\pi v}{c_y} > 0 \tag{20}$$

Let us substitute the cropping parameters and see when the maximum would be recognised. Table II shows the results of calculation for the different size of the window function. We

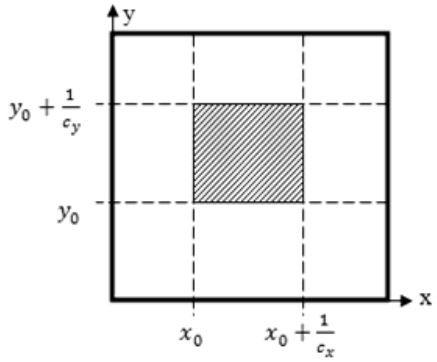


Fig. 4. Cropped area of an image

TABLE II
RESULTS OF MAXIMA RECOGNITION AFTER CROPPING BY WINDOW WITH COORDINATES x_0, $x_0 + \frac{1}{c_x}$, y_0, $y_0 + \frac{1}{c_y}$ AND TOTAL IMAGE SIZE 100X100

x_0	$x_0 + \frac{1}{c_x}$	y_0	$y_0 + \frac{1}{c_y}$	Detected
1	99	1	99	Y
2	98	2	98	Y
3	97	3	97	Y
4	96	4	96	Y
5	95	5	95	Y
6	94	6	94	Y
7	93	7	93	Y
8	92	8	92	Y
9	91	9	91	Y
10	90	10	90	Y
11	89	11	89	Y
12	88	12	88	Y
13	87	13	87	Y
14	86	14	86	Y
15	85	15	85	Y
16	84	16	84	Y
17	83	17	83	Y
18	82	18	82	Y
19	81	19	81	Y
20	80	20	80	N
21	79	21	79	N
22	78	22	78	N
23	77	23	77	N
24	76	24	76	N
25	75	25	75	N
26	74	26	74	N
27	73	27	73	N
28	72	28	72	N
29	71	29	71	N
30	70	30	70	N

gradually reduce the size of the window Fig. 4 and check how the detection process is performed. We can see that the maxima can be still detected after removing a half of an image.

E. Resize

Resizing the image results in inverse resizing of an amplitude spectrum. Resize can be represented as a multiplication of the coordinates on a corresponding constant value. If we look on the similarity theorem:

$$f(an, bm) \rightarrow \frac{1}{|ab|} F\left(\frac{n'}{a}, \frac{m'}{b}\right) \tag{21}$$

we can see that the resize in spatial domain causes frequency shift in the spectra. In combination with the resize maxima remains on the same distance from the center. So the maxima in the amplitude spectra will not change their positions.

F. Jpeg transform

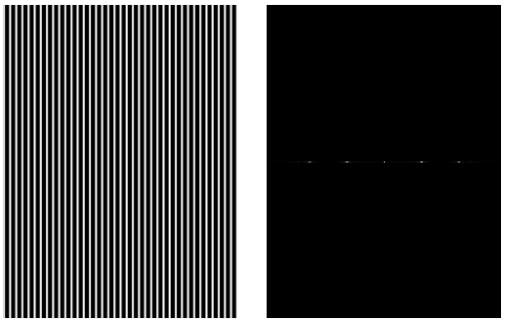


Fig. 5. Image after JPEG transform and the amplitude spectrum

Converting an image using JPEG algorithm produce specific kind of distortions. Many watermarking algorithms can not stand against such transform. The proposed algorithm can survive after JPEG transform performed with up to 30 present quality factor. This value may vary from image to image depending on image type and especially image size. Looking on the amplitude of the image after such transform we can see that some extra maxima appeared Fig. 5 right image. But all the values of those additional maxims are as the result much smaller than original ones. As long as we are searching for the max values those additional maxima give small effect on the extraction. We can see it only when additional maxima appear in the neighbour local areas with the smaller main maxima. Another effect of the JPEG transform is a removal of the high frequencies. After such transform most of the high frequencies are erased including the embedded maxima. The total number of maxima in the real system is about 350 for 100x100 pixel image. But the number of survived maxima is enough to detect the watermark.

The results of the experiments presented in Table III show that the probability of false detection appears equal to 0. The probability of successful detection of a WM is equal or close to 1 also after the cyclic shift on 50% on a vertical and a horizontal, and removal of 10% of the rows and columns.

In the Table III the recognised maxima number ratio to their total number of embedded maxima are presented. Total number of the embedded and extracted maxima is a mean value of the number of maxima, calculated as a result of 100

TABLE III
EXPERIMENTAL RESULTS

Characteristic	(1)	With embedding of a WM				
		(2)	(3)	(4)	(5)	(6)
Detected maxima number	25	295	209	252	240	231
Detected maxima %	8	100	72	85	81	78
Probability of successful WM extraction	0	1	1	0.92	0.93	0.97

(1): No embedding
(2): Without distortions
(3): Cyclic shift of 50% on a vertical and a horizontal axis
(4): Noise adding 5%
(5): Removal of 10% of rows and columns
(6): Cropping 20% of the image

images testing. For all experiments the parameters $a = 2$, $\beta = 1.5$ have been selected.

The probability of successful data extraction is sometimes less than 1, but it remains still acceptable, for the thing after adding a noise (5% of the image brightness range). However, the commercial value of the images after such strong conversions is low, and it is very unlikely to be applied to the images by pirates.

III. CONCLUSION

So in this paper we tried to explain why the watermarking system can survive after image distortions. We showed that in spite of the fact, that image distortions affect on the amplitude spectra the most part of local maxima survives and therefore zero-bit watermark can be recognised with great probability.

ACKNOWLEDGMENT

The author would like to thank Dr. Valery Korzhik for help and support.

REFERENCES

[1] S. Anfinogenov, V. Korzhik, and G. Morales-Luna. Robust digital watermarking system for still images. In *FedCSIS*, pages 685–689, 2011.
[2] C. Solomon. *Fundamentals of Digital Image Processing.* Wiley-Blackwell, 2011.
[3] Dong, P., Brankov, J., Galatsanos, N., Yang, Y., Davoine, F. *Digital watermarking robust to geometric distortions*, IEEE Transactions on Image Processing, vol. 14, no. 12, pp. 2140–2150, 2005.
[4] Liu, K. J. R., Trappe, W., Wang, Z. J. *Multimedia fingerprinting forensics for traitor tracing.* Hindawi, 2011.
[5] Ingemar J. Cox, Miller M.L, Bloom J.A, Fridrich J, Kalker T. *Digital Watermarking and Steganography.* Morgan Kaufmann Publishers, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 529–534

A Robust Cattle Identification Scheme Using Muzzle Print Images

Ali Ismail Awad[1,*], Hossam M. Zawbaa[2,*], Hamdi A. Mahmoud[2], Eman Hany Hassan Abdel Nabi[3,*]
Rabie Hassan Fayed[3,*], Aboul Ella Hassanien[4,*]
[1]*Faculty of Engineering, Al Azhar University, Qena, Egypt*
Email: aawad@ieee.org
[2]*Faculty of Computers and Information, BeniSuef University, BeniSuef, Egypt*
Email: hossam.zawbaa@gmail.com
[3]*Faculty of Veterinary Medicine, Cairo University, Cairo, Egypt*
Email: dr.emy2010@gmail.com, Email: rhfayed@hotmail.com
[4]*Faculty of Computers and Information, Cairo University, Cairo, Egypt*
Email: aboitegypt@gmail.com
[*]*Scientific Research Group in Egypt, (SRGE), http://www.egyptscience.net*

***Abstract*—Cattle identification receives a great research attention as an important way to maintain the livestock. The identification accuracy and the processing time are two key challenges of any cattle identification methodology. This paper presents a robust and fast cattle identification scheme from muzzle print images using local invariant features. The presented scheme compensates some weakness of ear tag and electrical-based traditional identification techniques in terms of accuracy and processing time. The proposed scheme uses Scale Invariant Feature Transform (SIFT) for detecting the interesting points for image matching. For a robust identification scheme, a Random Sample Consensus (RANSAC) algorithm has been coupled with the SIFT output to remove the outlier points and achieve more robustness. The experimental evaluations prove the superiority of the presented scheme as it achieves 93.3% identification accuracy in reasonable processing time compared to 90% identification accuracy achieved by some traditional identification approaches.**

I. Introduction

RECENTLY, governments pay a great attention to the livestock by providing vaccination against the most of diseases. They seek to overcome some food problems and keep the livestock as huge as possible. Cattle identification plays an important role in controlling the disease outbreak, vaccination management, production management, cattle traceability, and cattle ownership assignment [1]. Traditional cattle identification methods such as ear notching, tattooing, branding, or even some electrical identification methods such as Radio Frequency Identification (RFID) [2] are not able to provide enough reliability to the cattle identification due to theft, fraudulent, and duplication. Therefore, the need to a robust cattle identification scheme is a vital requirement.

Human biometrics is a key fundamental security mechanism that assigns unique identity to an individual according to some physiological or behavioral features [3], [4]. These features are sometimes called as biometrics modalities, identifiers, traits, or characteristics. Human biometrics identifiers must fulfill some operational and behavioral characteristics such as uniqueness, universality, acceptability, circumvention, and accuracy [5].

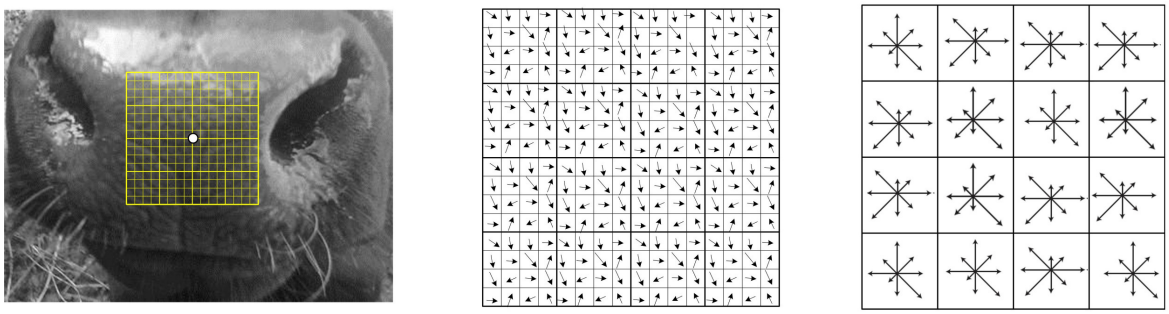
Adopting human biometric traits into animals is a promising technology for cattle identification domain. It has many applications such as cattle classification, cattle tracking from birth to the end of food chain, and understanding animal diseases trajectory and population. On the other side, using animal biometrics in computerized systems faces great challenges with respect to accuracy and robustness as the animal movement can not be easily controlled. Driven from this perspective, adopting human biometrics to cattle identification can overcome plenty of the current cattle identification weaknesses.

Muzzle print, or nose print, was investigated as distinguished pattern for animals since 1921 [6]. It is considered as a unique animal identifer that is similar to human fingerprints. Paper-based or inked muzzle print collection is inconvenient and time inefficient process. It needs special skill to control the animal and get the pattern on a paper. Furthermore, the inked muzzle print images do not have sufficient quality, and hence, it is difficult be used in a computerized manner [7]. Therefore, there is a lack of a standard muzzle print benchmark. Driven from this need, the first contribution of this research is to collect a database of live captured muzzle print images that works as a benchmark for evaluating the proposed cattle identification scheme.

A local feature of an image is usually related to a change of an image property such as texture, color, and pixel intensity [8]. The advantage of local features is that they are computed at multiple points in the image, and hence, they are invariant to image scale and rotation. In addition, they do not need further image pre-processing or segmentation [9]. Scale Invariant Feature Transform (SIFT) [10] is one of the popular methods for image matching and object recognition. SIFT features have been used by some researchers in human biometrics with applications on fingerprints [11], [12] and palmprints [13]. SIFT efficiently extracts robust and unique features, therefore it has been used to overcome different image degradation factors such as noise, partiality, scale, and rotation.

The identification accuracy is the foremost important fac-

978-83-60810-53-8/$25.00



(a) (b) (c)

Fig. 1. The process of building a single SIFT keypoint descriptor: (a) A Single SIFT keypoint extracted from muzzle print image, (b) A 16 × 16 pixel orientations, (c) A single 4 × 4 cells descriptor with 8 pixel orientations. The default length of a single SIFT keypoint descriptor is 4 × 4 × 8 = 128 element.

tor for measuring the performance of any automatic cattle identification approach. This paper presents a robust cattle identification scheme that uses SIFT features for calculating the similarity score between the input muzzle print image and the template one. The superiority of the proposed scheme is the assured cattle identification robustness provided by combining the robust SIFT features with a RANdom SAmple Consensus (RANSAC) algorithm for robust SIFT features matching [14].

The reminder part of this paper is organized as follows: Section II covers some preliminaries topics. Section III explains the design and the implementation of the proposed scheme. Section IV shows the evaluation phase of the proposed scheme. Conclusions and future work are written in Section V.

II. Preliminaries

A. SIFT Features

The SIFT feature extraction works through sequential steps of operations. These steps can be summarized as scale space extrema detection, keypoints localization, keypoint orientation assignment, and building the keypoints descriptor [15]. The Difference-of-Gaussian (DOG) is used to detect the keypoints as the local extrema of DOG function. The pixel is compared against 26 neighboring pixels (8 in the same scale, 9 in the above scale, and 9 in the below scale) to detect the local pixel extrema and minima. Following on, the detected keypoint is localized by determining its neighborhoods, and examine them for contrast and edge parameters. The keypoints with low contrast and weak edge responses are rejected. The keypoint neighborhoods region is used to build the histogram of the local gradient directions, and the keypoint orientation is calculated as the peak of the gradient histogram [10], [15], [16]. The default SIFT feature extraction produces keypoint associated with a descriptor of 128 element length. The descriptor is constructed from (4 × 4 cells) × 8 orientations [17]. The cascaded operations of building a single SIFT keypoint descriptor from muzzle print image are shown in Fig. 1.

Applying SIFT feature extraction translates the muzzle print image into a set of keypoints according to the local maxima. The extracted keypoint is associated with a descriptor related to the orientations of the surrounded pixels. In this paper, a standard SIFT extraction has been used for keypoint detection and building the associated descriptor. SIFT features have been extracted and matched using the VLFeat library [18]. The output of matching process is a similarity score between the input image and the template that is enrolled in the database.

B. Identification Accuracy

In order to measure the accuracy of the presented cattle identification scheme, The identification *Error Rate* (*ER*) is considered. The ER is defined as "the rate that the identified animal, the animal who corresponds to the template image, is different from the animal of the input image". We also consider the standard verification error rates as FAR, FRR, and ERR [19]. The *False Acceptance Rate* (*FAR*) is the rate that the similarity between the images of different animals is greater than a threshold. Whereas the *False Rejection Rate* (*FRR*) is calculated as the rate that the similarity between two images of the same animal is less than a threshold. Thus, the FAR and FRR depend on the similarity threshold. The *Equal Error Rate* (*EER*) is the value of FRR and FAR at the point of the threshold where the two error rates are identical [20].

III. Proposed Identification Scheme

Analogy to the human fingerprints, cattle muzzle prints have some discriminative features according to the grooves, or valleys, and beads structures. These uneven features are distributed over the skin surface in the cattle nose area. These features are defined by the white skin grooves, or by the black convexes surrounded by the grooves [7], see Fig. 3 for consulting the convexes and the grooves in muzzle print images taken from two different animals.

Minagawa et al. [7] used the joint pixels on the skin grooves as a key feature for muzzle print matching. Some long preprocessing steps were conducted to extract the joint pixels. This approach achieved maximum and minimum matching scores as 60% and 12%, respectively. It achieved unsatisfactory identification performance (accuracy) that was around 30% measured over a database of 43 animals.

Noviyanto and Arymurthy [21] applied Speeded-Up Robust Features (SURF) on muzzle print images for enhancing the identification accuracy. A U-SURF method was applied on 8

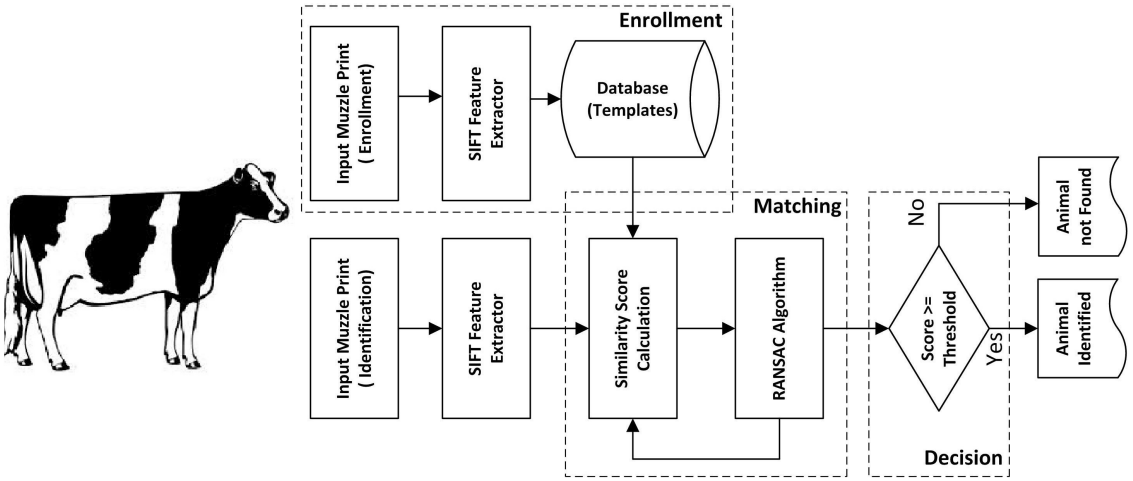


Fig. 2. A block diagram of a complete cattle identification system using muzzle print images. The components of the enrollment phase and the identification phase are emphasized in the block diagram. The proposed identification scheme is represented as a combination between SIFT features and RANSAC algorithm.

animals with 15 images each. The experimental scenario used 10 muzzle print images in the training phase, and the other 5 images were used as input samples. The maximum achieved identification accuracy under rotation condition was 90%.

The presented scheme in this research is robust from two perspectives. First, it invests the robustness of the SIFT features to image scale, shift, and rotation. Second, it uses the RANSAC algorithm as a robust inliers estimator for enhancing the matching results of SIFT features, and ensure the robustness of the matching process. The proposed scheme includes SIFT feature extraction, SIFT feature matching, and RANSAC algorithm. Fig. 2 shows a generic and complete muzzle print based cattle identification system, and highlights the cascaded components of the presented scheme.

RANSAC algorithm has been developed by Fischler and Bolles [14] especially for computer vision, and it works as robust estimator for features matching. In many images matching cases, RANSAC is an effective robust estimator, which can handle around 50% mismatch contamination levels of the input samples. The integration of the extracted local invariant features and RANSAC is valuable for optimizing the images similarity score measurement using SIFT features [22].

Admittedly, the generic animal identification system, shown in Fig. 2, works the same way of the human identification one. It has two phases; enrollment phase and identification phase. In the enrollment phase, a muzzle print image is presented and the SIFT feature vector is constructed. Then, the extracted feature vector is stored as a template in the database. The identification phase includes the same enrollment procedure plus matching and decision sub-phases. For calculating the similarity score, the SIFT features of the input image is matched against the templates stored in the database as (1:N) matching procedure. The muzzle print image corresponding to the feature vector that has a shortest distance to the input feature vector is considered as the most similar one, and it is given the highest similarity score. RANSAC homography algorithm comes at the end of the matching process to remove the matching outliers, mismatched SIFT keypoints, data and ensure the robustness of the similarity score. The animal identity is then assigned according to the highest estimated similarity score between the input image and the template one.

IV. Performance Evaluation

The experiments in this paper have been conducted using a PC with Intel® Core™ i3-2120 running at 3.30 GHz, and 8 GB of RAM. The PC is empowered by Matlab® and Windows® 64-bit. The VLFeat library [18] has been used for extracting, processing, and matching the SIFT features. VLFeat has been installed and optimized for the mentioned experimental environment.

A. Database

The lack of a standard muzzle print images database was a challenge for conducting this research. Therefore, collecting a muzzle print images database was a crucial decision. The database has been collected from 15 cattle animals with 7 live captured muzzle print images each. A sample of muzzle print images captured from two individual animals is shown in Fig. 3. A special care has been given to the quality of the collected images. The collected images cover different quality levels and degradation factors such as image rotation and image partiality for simulating some real time identification conditions.

The identification scenario works as follows: 7 images of each animal have been swaped between the enrollment phase an identification phase, and the similarity scores between all of them are calculated. Therefore, similarity score matrix with dimension of 105 $\times$ 105 have been created. The animal is correctly identified if the similarity score between the input sample and the template samples is greater than or equal a specific threshold. The template of a single animal has been constructed from 6 images which were marked as $T_1, T_2, T_3, ..., T_6$. The remaining 1 image has been used as

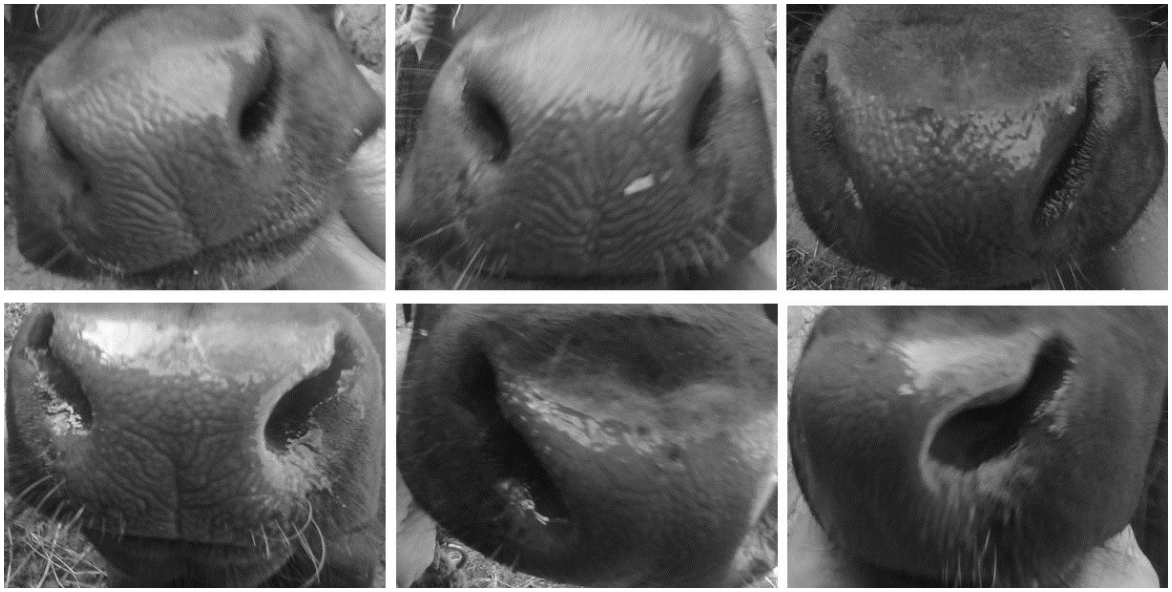


Fig. 3. A sample of the collected muzzle print images database from live animals. The represented muzzle print images have been taken from two different animals. The muzzle print images show different deteriorating factors include orientated images, blurred images, low resolution images, and partial images.

input, and was marked as I_1, S was a similarity function, and H was a similarity score. A correctly identified animal should strictly following the next equation as:

$$S(I_1, T_1) \parallel S(I_1, T_2), ..., \parallel S(I_1, T_6) \geq H \qquad (1)$$

The FAR, FRR, and ERR have been calculated according to the criteria mentioned in Section II-B.

B. Evaluation Results

Preceding to any experimental work, the database images have been processed in terms of image enhancement, image segmentation, and image normalization. The first experimental scenario is directed toward setting the best SIFT parameters that compromise the number of extracted features (keypoints) with the consumed processing time. The preparatory experiments showed that the most effective parameter is the peak threshold (`PeakThresh`) [15], [18], thus the objective of this scenario is to optimize the peak threshold. The results of the conducted experiments is shown in Fig. 4. The reported results are the average of value of 105 feature extraction processes and 5565 matching operations. The maximum number of features is achieved with (`PeakThresh = 0.0`), however with (`PeakThresh = 0.001`), the extracted features are reduced by 30, and the extraction time is reduced by 5 ms. The other `PeakThresh` values achieve unacceptable number of features regardless of the time factor. The optimum `PeakThresh` value is selected as 0.0 seeking for more SIFT features, and hence, more robustness in feature matching. Following on, the SIFT peak threshold is set to that optimum value, whereas the other parameters were kept as defaults.

In real time identification, 6 images of each individual animal have been processed and enrolled in the database, the

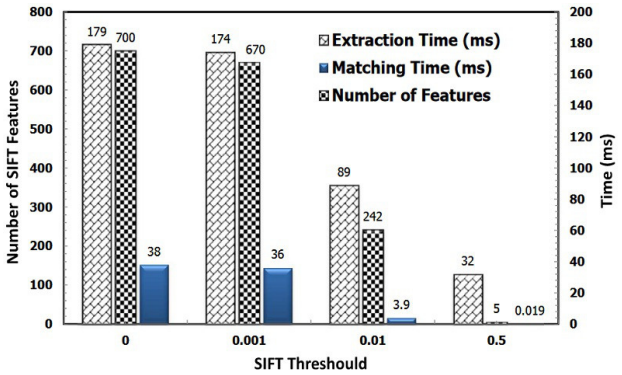


Fig. 4. The behavior of SIFT feature extraction with different peak threshold (`PeakThresh`) values with respect to the number of features, the extraction time, and the matching time. The optimum value is `PeakThresh = 0.0`.

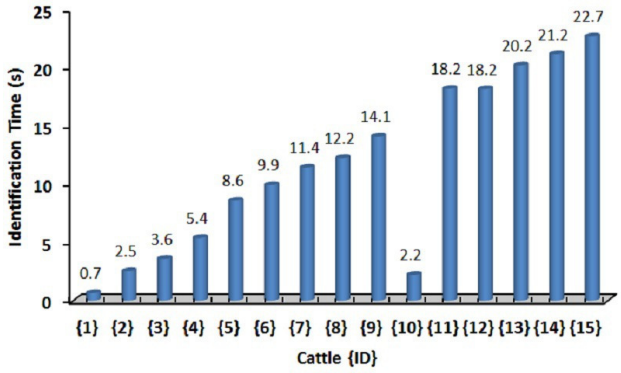


Fig. 5. The identification time for each input animal. Due to the linear search approach, the identification time linearly increases as the order of the template inside the database increases. The cattle with ID $\{10\}$ is wrongly identified.

TABLE I
THE INPUT IMAGES, THE MATCHED IMAGES, THE MATCHING SCORES, AND THE IDENTIFICATION STATUS FOR THE IDENTIFICATION SCENARIO.

Input Image	Matched Image	Matching Score	Identification Status
101_5	101_3	71.56	Correct
102_5	102_4	45.45	∼
103_5	103_1	73.33	∼
104_5	104_1	42.00	∼
105_5	105_7	45.00	∼
106_5	106_3	51.85	∼
107_5	107_3	87.77	∼
108_5	108_1	48.00	∼
109_5	109_3	95.37	∼
110_5	**102_3**	**45.76**	**False**
111_5	111_7	39.00	Correct
112_5	112_1	70.37	∼
113_5	113_3	89.59	∼
114_5	114_1	57.14	∼
115_5	115_1	50.00	∼

total images in the database were ($6 \times 15 = 90$), and 1 image has been used as input to simulate the identification operation. According to equation 1, 14 animals out of 15 have been correctly identified which achieves equivalent identification accuracy value as 93.3%. It is worth notice that the average consumed feature extraction time is 179 ms and the average individual matching time is 38 ms including RANSAC optimization, which are consistent with Fig. 4. However, both times are considered very short for single feature extraction and matching operation, the total identification time still long, around ≈ 23 s at maximum, because a linear database research method has been used, and the identification time is based on the location of the template inside the database. The identification time of each input animal is shown in Fig. 5.

Table I summarises outcomes of the conducted real time identification phase in terms of the input image, the matched template image, the matching score, and the status (correctness) of the identification operation. The image naming scheme works as $1XX\_Y$, whereas XX is the cattle ID (1 to 15), and Y is the image order (1 to 7). The Table shows that the cattle with ID $\{10\}$ is wrongly identified because the similarity score with a template image from cattle ID $\{2\}$ is greater than the defined threshold. The reported results in the Table are consistent with Fig. 5 as the wrongly identified animal consumes very short identification time, and violates the linearity of the incremental identification time with the increased cattle ID.

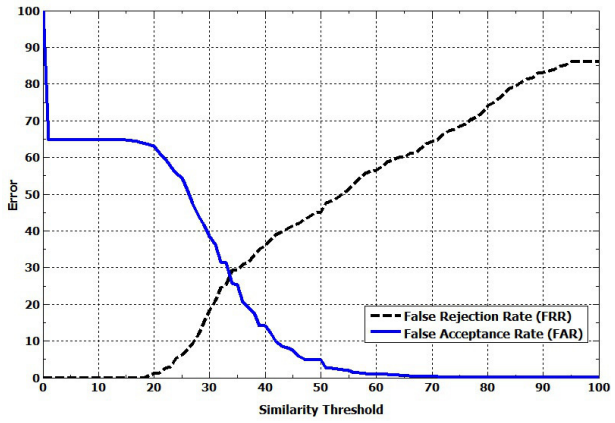


Fig. 6. False Acceptance Rate (FAR) and False Rejection Rate (FRR) plotted versus the similarity threshold. The Equal Error Rate (ERR) is shown as the cross point between FAR and FRR. ERR is ≈ 27.4 with threshold is ≈ 34.0.

The wrongly identified animal is considered as false matched or false accepted input because the match occurred with a template that does not correspond to the input sample. The FAR in this case is 6.67%, and it is equal to the identification ER. The relations between FAR, FRR, and ERR are determined according to the similarity threshold. Fig. 6 shows FAR versus FRR related to the similarity threshold. In order to achieve FAR equal to 6.67%, the similarity threshold should be selected around 45. However, the conducted experiments showed that FAR equals to 6.67% has been achieved with a threshold equals to 39, and with FRR equals to 0. We do believe that this is because of combining multiple images from the same animal in one database template.

V. CONCLUSIONS AND FUTURE WORK

This paper has presented a robust cattle identification scheme that uses muzzle print images as input to SIFT feature extraction and matching. Due to the lack of standard muzzle print database, we have collected 105 images from 15 animals to work as a benchmark for the presented scheme. In order to evaluate the robustness of the scheme, the collected images cover different deteriorating factors such as rotated images, blurred images, partial images, and low resolution images. The achieved identification accuracy is 93.3% compared to 90% reported in the literature. The superiority of the presented scheme comes from the coupling of local invariant features with RANSAC homography as a robust outliers removal algorithm. Muzzle print images database extension and standardization for international benchmark of muzzle print related algorithms is one of the future work. Additionally, the reduction of the identification time in a large database is an interesting challenge that will also be tackled in the future.

REFERENCES

[1] M. Vlad, R. A. Parvulet, and M. S. Vlad, "A survey of livestock identification systems," in *Proceedings of the* 13^{th} *WSEAS International Conference on Automation and Information, (ICAI12)*. Iasi, Romania: WSEAS Press, June 2012, pp. 165–170.

[2] C. Roberts, "Radio frequency identification (RFID)," *Computers & Security*, vol. 25, no. 1, pp. 18–26, 2006.

[3] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Springer, 2011.

[4] R. Giot, M. El-Abed, and C. Rosenberger, "Fast computation of the performance evaluation of biometric systems: Application to multibiometrics," *Future Generation Computer Systems*, vol. 29, no. 3, pp. 788–799, 2013, Special Section: Recent Developments in High Performance Computing and Security.

[5] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, January 2004.

[6] W. Petersen, "The identification of the bovine by means of nose-prints," *Journal of Dairy Science*, vol. 5, no. 3, pp. 249–258, 1922.

[7] H. Minagawa, T. Fujimura, M. Ichiyanagi, and K. Tanaka, "Identification of beef cattle by analyzing images of their muzzle patterns lifted on paper," in *Proceedings of the Third Asian Conference for Information Technology in Agriculture, (AFITA 2002): Asian Agricultural Information Technology & Management*, Beijing, China, October 2002, pp. 596–600.

[8] K. Mikolajczyk and T. Tuytelaars, "Local image features," in *Encyclopedia of Biometrics*, S. Li and A. Jain, Eds. Springer US, 2009, pp. 939–943.

[9] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, Jul. 2008.

[10] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of* 7^{th} *IEEE International Conference on Computer Vision (ICCV'99)*, Kerkyra, Corfu, Greece, September 1999, pp. 1150–1157.

[11] G. Iannizzotto and F. L. Rosa, "A SIFT-based fingerprint verification system using cellular neural networks," in *Pattern Recognition Techniques, Technology and Applications*. InTech, 2008, pp. 523–536.

[12] U. Park, S. Pankanti, and A. K. Jain, "Fingerprint verification using SIFT features," in *Proceedings of SPIE Defense and Security Symposium*, 2008.

[13] J. Chen and Y.-S. Moon, "Using SIFT features in palmprint authentication," in *Procedings of* 19^{th} *International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[16] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[17] S. Saleem, A. Bais, and R. Sablatnig, "A performance evaluation of SIFT and SURF for multispectral image matching," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, A. Campilho and M. Kamel, Eds. Springer Berlin / Heidelberg, 2012, vol. 7324, pp. 166–173.

[18] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008. [Online]. Available: http://www.vlfeat.org/

[19] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd ed. Springer-Verlag, 2009.

[20] A. I. Awad and K. Baba, "Evaluation of a fingerprint identification algorithm with SIFT features," in *Proceedings of the* 3^{rd} *2012 IIAI International Conference on Advanced Applied Informatics*. Fukuoka, Japan: IEEE, September 2012, pp. 129–132.

[21] A. Noviyanto and A. M. Arymurthy, "Automatic cattle identification based on muzzle photo using speed-up robust features approach," in *Proceedings of the* 3^{rd} *European Conference of Computer Science, ECCS12*. Paris, France: WSEAS Press, December 2012, pp. 110–114.

[22] L. Cheng, M. Li, Y. Liu, W. Cai, Y. Chen, and K. Yang, "Remote sensing image matching by integrating affine invariant feature extraction and RANSAC," *Computers & Electrical Engineering*, vol. 38, no. 4, pp. 1023–1032, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 535–540

Logo identification algorithm for TV Internet

Marta Chodyka
Institute of Computer Science, Pope John Paul II
University in Biala Podlaska, Poland
Email: m.chodyka@dydaktyka.pswbp.pl

Volodymyr Mosorov
Institute of Applied Computer Science, Lodz
University of Technology, Poland
Email:
volodymyr..mosorov@.p.lodz.pl

***Abstract*—Content inappropriate for children on Internet television is a serious problem in today's multimedia world. There are numerous methods which are used to control the content of the transmitted television programmes. However, these well-known methods do not solve the above mentioned problem completely. The paper presents a more effective method for automatic identification of the provider's logo based on an original image sequence analysis. The automatic identification of the provider's logo can be used to block access to video programmes of the selected providers. The method has been tested on some chosen video transmissions on-line, achieving over 98% of correct identification.**

I. Introduction

THE problem of underage persons' easy access to the multimedia video with inappropriate content and its consequences is well known [3,14]. One of the sources enabling the access to such video programmes is the widely available Internet TV. There are numerous methods which are used to control the content of the television programmes transmitted via the Internet. These include, among others, blocking video materials at certain hours [19,20] or filtering chosen IP addresses and keywords on web pages [18]. There are also parental control modules, which can be embedded in the anti-virus software, web browsers and operation systems. All these well-known methods do not, however, solve the above mentioned problem completely. Thus, in the case of a temporary access block on Internet TV, parents must be involved in the process of programme assessment and selection. With regard to IP address filtering, the problem concerns a rapidly growing number of keywords, which the filter should block, as well as easily made changes of the IP addresses by Internet providers.

Another solution is to do an analysis of the provider's logo transmitted together with the video stream. In a video production, logos are used to convey information about the provider's programme content, which can be used in the selection of age-appropriate programmes while broadcasting video. There are related applications which try to identify brand logotypes in video data [5, 6, 11, 15] by using the static character of the logo. In order to identify the logo, some logo detection algorithms use neural network and image analysis procedures [1,7,8,10,16,17]. However, the selection of an adequate neural network's models, their over-fitting capacity and the high computational cost of the methods limit their applications in practice.

The logo identification in the programme categorisation is presented by Cozar et al. 2007 [4]. This method performs a temporal and spatial segmentation by calculating the minimal luminance variance region of the set of frames and the non-linear diffusion filtering. However, 95% of correct identification has been achieved only when the analysis is conducted on-line. A different solution is presented by Ozay, Sankur [13]. This time, an algorithm performs a detection of the logo by morphological operations. Nevertheless, online tests for detection and recognition on running videos have achieved lower than 96% average accuracy. In [2] logo detection techniques have been used to differentiate advertisements from TV programmes. This approach assumes that a logo exists if a region with stable contours can be found in the image. No temporal information is used and the method has not been tested on video material in a real time transmission, which has resulted in many false detection cases.

Contrasting the aforementioned methods, the paper presents a more effective method for automatic identification of the provider's logo based on an original image of sequence analysis. The automatic identification of the provider's logo allows to block access to video programmes of the selected providers. It takes place regardless of the transmission time, IP address or the keywords used to find a required website.

The method has been tested on some transmitted video, achieving over 98,7% of correct identification.

The article consists of several parts. Section 2 contains a description of the logo detection algorithms based on spatial segmentation. It additionally presents the logo identification and its comparison with logo patterns. Section 3 concentrates on testing the presented method on chosen video streams and illustrating the results of its application. The article ends with Section 4, which includes main conclusions and presents plans for the further development of the above method.

II. Algorithm Description

The video streaming Internet TV programme is a set of ordered frames through time. These frames can include one or several superimposed logos. Usually, a logo is defined as

978-83-60810-53-8/$25.00

a small graphic or picture that appears behind the anchor person on the screen. Logo image areas show luminance variance values in narrower interval than other image areas, depending on the logo transparency. An important feature of a logo image is that the logo contours are stable, while the background varies during video broadcasting. Besides, during video broadcasting a logo can be present or absent, for instance during an interruption of the programme transmission. Logotypes are usually placed at any of the four corners of a frame. Therefore, four image corners should be considered as the regions of interest (ROIs). Moreover, their size is limited, since logos should not perturb video viewing (see Fig. 1). Furthermore, logo areas do not significantly change from frame to frame.

Fig. 1 Example of the frames from broadcasting TV

A logo is a characteristic feature of any programme provider as well as its contents. Logo identification enables verification of various programmes providers, which makes it a tool of parental control, enabling blocking unsuitable programmes for underage viewers. A child's parent or guardian chooses logo patterns from a providers' base which are regarded as inappropriate for children. When a programme transmission takes place, its logo is identified and compared with the ones selected as unwelcome by the parent or guardian. Depending on the received information, the video signal is either blocked or allowed to flow.

When the logo of the transmitted on-line programme is not included in the data base, the system can add this new candidate logo to the logo patterns base. The new candidate undergoes a process of segmentation, yet it is not included in the currently transmitted logo identification process. Automatic logo adding to the logo data base can take place after it has been projected and recognised several times.

Let a mathematical model of a logo image be a matrix, $I = I(i, j), i = 1..m, j = 1..n$, where m and n define the size of the logo image.

Initially, the digital image I of the analyzed logo region is converted to the monochrome image $\mathbf{I}'$. This operation includes the calculation of the brightness $I'(i, j)$, $0 \le I'(i, j) \le 255$, for each pixel of the RGB colour components $I'(i, j)$.

To extract contours of the logo regions of the monochromatic image $\mathbf{I}'$, the Sobel operator [9] is applied. Due to this operation an image of the logo contours is created $\mathbf{I}^*$. However, the extracted contours of the logo regions are often not salient because the result of the extraction depends considerably on the time variable background where logos appear. In order to achieve better quality of the contours, the adopted method averages the sequence of the logo contours $\mathbf{I}^*$:

$$\bar{I} = \frac{1}{K} \sum_{k=1}^{K} I_k^*, \qquad (1)$$

where I_k^* is the logo contours image and K – is the number of frames, $\bar{\mathbf{I}}$ -the average image of the logo's contours.

It seems clear that in the sequence (see eq. 1) the number of the processed frames K depends mostly on the characteristics of the video stream.

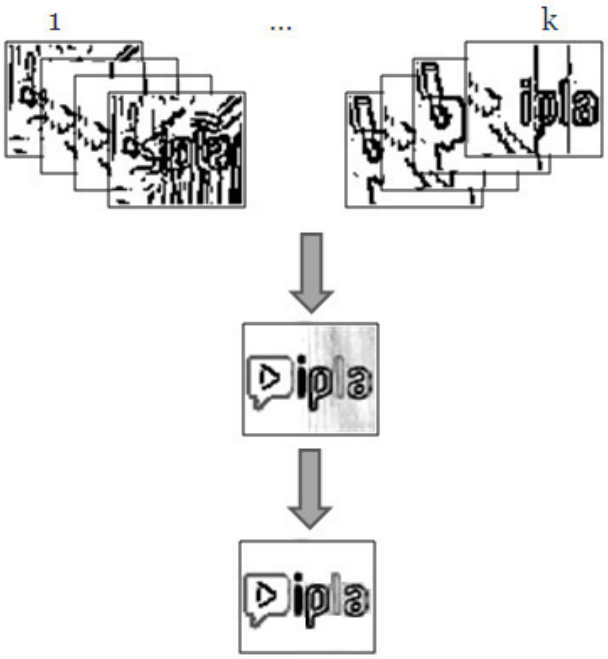


Fig. 2 Images obtained in each stage of the algorithm for a real sequence through time: logo contour image , average logo contour image and binary logo image

Thus, a video with a dynamic sequence of frames needs fewer frames to generate stable logo contours than the one with a static sequence. Therefore, K value should be chosen experimentally. It seems plausible that a large value K can guarantee better detection for logos which are static for a long period of time. However, a wrong logo contour image is obtained if logo changes occur within the K frames. In this case, the number of frames K used for the logo extraction must be decreased.

In the next stage, a spatial segmentation of logo contours is conducted binarizing of $\mathbf{I}^*$:

$$\mathbf{B} = B(i,j) = \begin{cases} 0, for & \bar{I}(i,j) \geq p_1 \\ 1, for & \bar{I}(i,j) < p_1 \end{cases}, \; i = 1..m, \; j = 1..n \tag{2}$$

where **B** is the binary image of the logo contours and the threshold level p_1 which are arbitrarily determined from a histogram. An appropriate choice of the threshold level p_1 is the basis of a proper process of identifying the logo contours from the image. In order to calculate the required level, average histograms of the logo contours are determined $\bar{\mathbf{I}}$, which, due to different backgrounds, vary considerably (see fig. 3).

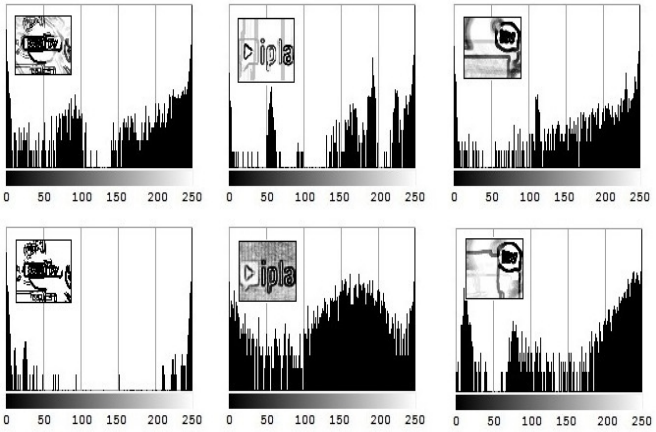


Fig. 3 Images of averaged logo contours $\bar{\mathbf{I}}$ and their respective histograms

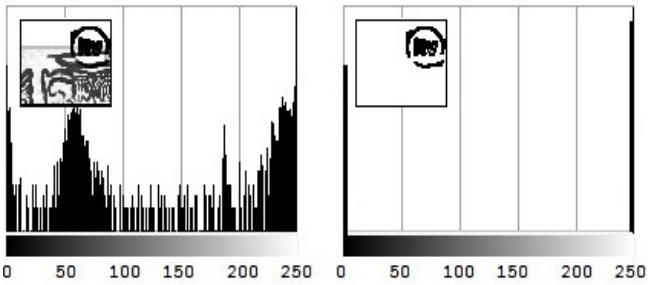
The optimum level p_1 is calculated by means of the Otsu [13] method according to formula 3.

$$p_1 = \arg\max_{p} \left(\omega_0 \omega_1 (\mu_1 - \mu_0)^2 \right) \tag{3}$$

where ω_0 - constitutes a standardised quantity of the logo contours (a quotient of the number of points belonging to the contours and the number of the image points), ω_1 is the standardised number of the background quality, μ_0 and μ_1 are the averaged qualities of the points brightness for the contours and background respectively, $0 \leq p \leq 255$.

Fig.4 presents example of an image and histogram before and after the application of the Otsu method.

Generally, there are cases when the analysed video stream does not comprise any logo, for instance, during commercial breaks. To recognise such a case the following procedure of logo histogram analysis is proposed. Examples of images without logo $\bar{I}$ and their respective histograms are presented in Fig. 5.



a) b)

Fig. 4 An image and its histogram before (a) and after the application of the Otsu binarisation method (b).

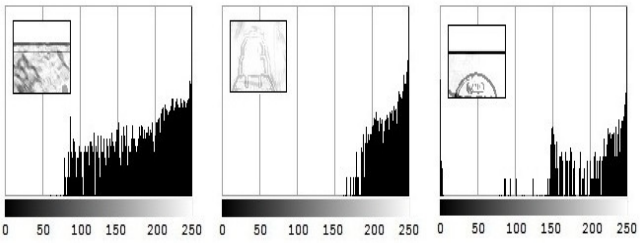
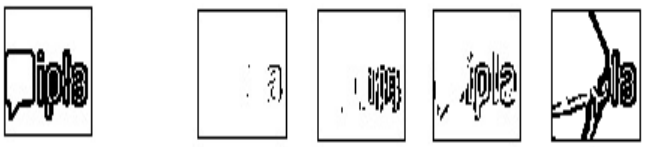


Fig. 5 Some images without logo $\bar{I}$ and their respective histograms.

The next step includes calculating sums S_1 and S_2 how often grey scale values $h(p)$ larger than $0.5h_{max}$ appear into the two ranges <0... p_1 > and (p_1 ...255> respectively, where h_{max} indicates the maximum of a histogram. If $S_1 \leq S_2$, it may be inferred that the logo is not included in the image.

When images undergo the analysis process, two kinds of errors may take place. The first one concerns a situation when the logo is present but has not been identified by our algorithm. This happens when algorithm reads incomplete logo contours, i.e. when it identifies light contours in a light background. The case is illustrated by figure 6.



a) b)

Fig. 6. Examples of binary image contours of the logo B presenting the logo of IPLA provider in real time sequences: full logo contours (a), and incomplete logo contours (b)

The other error connected with the logo identification may take place when the logo is not present and algorithm identifies static contours of an object as the logo, and subsequently adds the identified contours to the data base as a new pattern. Some examples concerning such situations are presented in figure 7.

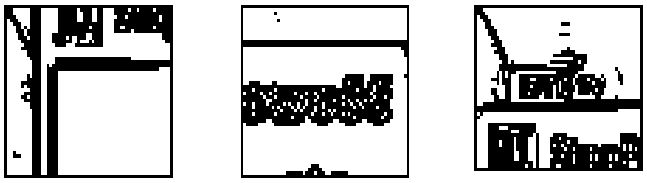


Fig. 7. Examples of binary logo B contours identified inappropriately as potential candidates for new logos.

The above situations may take place due to the nature of the discussed problem. Proper recognition of such cases by our algorithm is, however, difficult.

To identify the logo, it is first of all necessary to define the logotype database as a set of the logo patterns representing different broadcast providers of the Internet TV programmes. Let $\left\{B_r^z\right\}$, $r=1..R$, be the reference set of the R logo patterns. Each pattern B_r^z is obtained by the same procedure as the one described above, when the background is stable.

A good descriptor of the binary image **B** of the logo contours is the shape itself, but a long feature vector would be created. An important reduction of the feature vector size, without a great loss of accuracy, can be achieved if the x-axis and y-axis shape projections are used.

Let $w_i=\sum_{j=1}^{n} B(i,j)$, $i=1..m$, and $k_j=\sum_{i=1}^{m} B(i,j)$, $j=1..n$ mean x-axis and y-axis shape projections of a binary image **B** of the logo contours.

Then, a good metric to compare the feature vectors [**w**, **k**] and [**w**^{z}, **k**^{z}] of **B** and B_r^z respectively is the distance given by the following expression:

$$\min_r \left(\sum_{i=1}^{n} |w_i^I - w_{r,i}^z| + \sum_{j=1}^{m} |k_i^I - k_{r,i}^z| \right), \quad r=1..R \qquad (6)$$

Algorithm enables an automatic supplementation of the pattern data base. A candidate analysis of a new pattern is conduced according to of the rank of correlative factors τ Kendala [10] between the analysed image and patterns. The method enables qualifying if the logo included in the transmitted programme exists in the data base or whether it should be added as a potential candidate.

III. Method Verification

„*StopPlay*", a novel application shown in Figure 3, has been written in the C# language. The our application analyses a video stream of the selected Internet television programmes in on-line regime. In order to verify the correctness of the algorithm in the process of the logo recognition, a set of six patterns of the logo $\left\{B_r^z\right\}$, $r=1..6$ (see Fig. 8) of popular Internet televisions was defined. The Internet addresses of Internet television programmes used in the tests include Inter Alia: http://www.itv.net.pl, http://www.ipla.pl.

The logo images of dimensions 60x50 pixels are automatically extracted from each frame in the video stream during the transmission of Internet television programme. The number of binary images needed to create an average contour image was set at K=40.

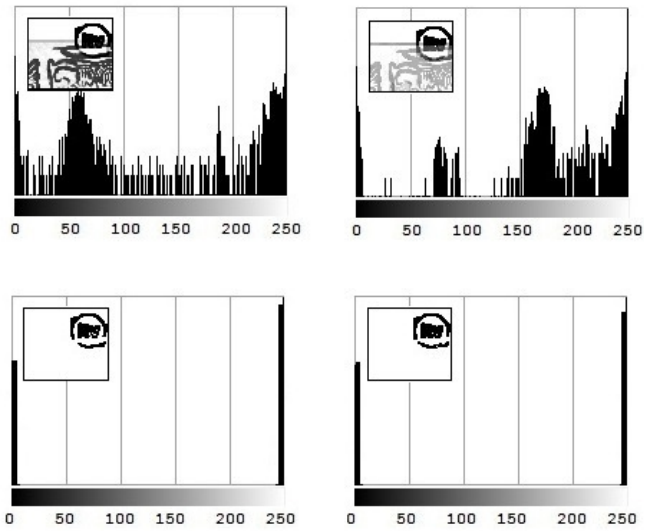
Fig. 8 An exemplary set of chosen logotypes and main application window

As it is argued in Section 2, in order to rid the programme of disturbances and get clear contours of the logo in its background, qualities p_1 cannot be taken arbitrarily. Fig. 9 presents examples of averaged contours of the logo and their respective binary images and histograms. The threshold levels p_1 are chosen according to the Otsu method and depend on the levels of grey shades in the image.

The use of such values allows to achieve approximately 99% of correct identification in the logo detection procedure.

The only activity left for the user is to choose the names of the provider (providers), whose logo should be recognised from a particular set of programmes. Fig. 10 presents an analysis of the tested logos of the television programmes. The tests were conducted on an average of 20 000 video frames during approximately three hours' time on the three available TV sites: ITV, EZO, IPLA. The algorithm was tested during the TV programme transmission as well as during commercial breaks. The obtained results show that the presented algorithm detects the logo with an accuracy of over 99%.

The lack of proper recognition of the logo is due to cases when the logo and the background are in the same colours, i.e. without visible logo's contours, as well as cases when some permanent objects are present in the logo. When recording consecutive frames of video sequences these additional objects become regions identified in the algorithm as a logo.



a) b)

Fig. 9. Averaged images of the logo contours and their respective binary images with appropriate quality levels $p_1 = 35$ (a), $p_2 = 100$ (b) presented according to the Otsu method and their histograms.

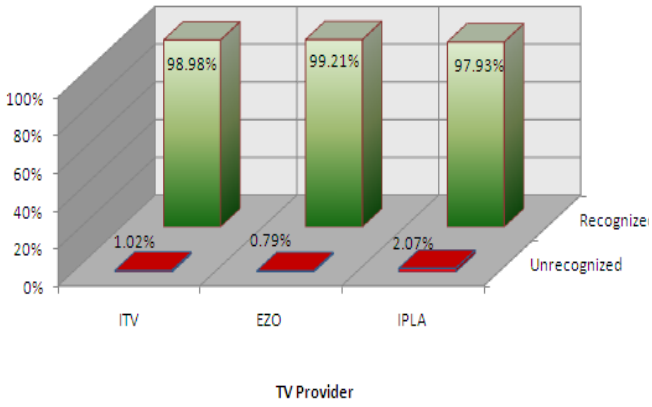


Fig. 10 Percentage chart for positive recognised logos in the selected broadcasting Internet video

IV. Conclusions

This article presents our logotype recognition algorithm and its application in the television programme providers in the on-line regime.

The suggested method takes advantage of a multi-step segmentation of temporal and spacious logo, which enables detecting the image contours and eliminating the background objects from the on-line video images. A comparison of the achieved images of the logo with the patterns allows an automatic identification of the transmitted programme. The identification process takes into account situations when the logo is not present due to, for instance, an interruption of the transmission process.

It has been proved that the implemented algorithm is capable of detecting images of the logo with an accuracy of over 98,7%. The cases which are problematic are due to situations when the logo and background images are in the same colours and when permanent objects appear in the logo region. Under such circumstances the algorithm identifies the entire regions as logos. However, in contrast to many other object recognition algorithms, the proposed algorithm does not require preparation of any learning set or application of any advanced methods for image processing. This allows for its practical and easy use in the application of automatic identification of television programmes and minimises the potential negative effects of Internet television on children.

The future works will be focused on development of the proposed algorithm for cases when known logo is slightly modified by TV providers in mourning days.

References

[1] T. Acharya, A. K. Ray "*Image Processing: Principles and Applications"*, John Wiley, pp. 428, 2005

[2] A. Albiol, M. J. Fulla, L. Torres "*Detection of TV commercials"*, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. 3, pp. 541–544, 2004

[3] A. Bryant "*The Children's Television Community",* Lawrence Erlbaum Associates, Mahwah, New Jersey, 2007

[4] J. R. Cozar, N. Guil, J.M. Gonzalez-Linares, E.L.Zapata, E. Izquierdo "*Logotype detection to support semantic-based video annotation"*, Signal Processing: Image Communication 22, Elsevier B.V, pp. 669–679, 2007

[5] S. Duffner, C. Garcia "*A neural scheme for robust detection of transparent logos in TV programs"*, Lecture Notes in Computer Science - II, vol. 4132, Springer, Berlin, pp.14–23, 2006

[6] A. Ekin, R. Braspenning "*Spatial detection of TV channel logos as outliners from the content"*, Proceedings of SPIE - The International Society for Optical Engineering, San Jose, USA vol. 6077, January 18, 2006.

[7] R. C. Gonzalez, R.E. Woods "*Digital Image Processing",* Prentice Hall, Upper Saddle River, New Jersey, 2001

[8] "*Handbook of Pattern Recognition and Computer Vision"*, 4th Edition edited by C H Chen (University of Massachusetts Dartmouth, USA), pp. 796. 2009

[9] B. Jähne "*Digital Image Processing",* Springer, Berlin, Heidelberg, New York, 2002

[10] M. G. Kendall "*Rank Correlation Methods*", Edition 1. London: Charles Griffin, 1948

[11] H. Kim, W.-Y. Loh "*Classification Trees With Unbiased Multiway Splits"*, Journal of the American Statistical Association, pp. 598-604, 2001

[12] G. Diamantopoulos; S. Salehi "*Learning VirtualDub*", Packt Publishing, 212 p. 2005

[13] N. Otsu "*A threshold selection method from grey-level histograms"*, IEEE Trans. System Man Cybernet, 9(1), pp. 62-69, 1979

[14] S Livingstone *"Children and the Internet",* Edition 1, Wiley, 272 pages

[15] S. J. Kirsh "*Children, adolescents, and media violence: a critical look at the research*", Sage, California, 2006

[16] W. Q. Yan, J. Wang, M.S. Kankanhalli M "*Automatic video logo detection and removal"*, Multimedia Systems 10(5), pp. 379-391, 2005

[17] I. T. Young, J. J. Gerbrands., L. J. van Vliet "F*undamentals of Image Processing"*, The Netherlands at the Delft University of Technology, 1998

[18] Patent "*System and Method for Subscriber Controlled Signal Blocking"* no CA2266982, 1999

[19] Patent "*Method for child lock in Internet Television"* no KR20010037415, 2001

[20] Patent "*Method and apparatus for permitting a potential viewer to view a desired program"* noUS2004015985, 2004

[21] Patent "*Content control system"* no US2005028191, 2005

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 541–548

Semantic Multi-layered Design of Interactive 3D Presentations

Jakub Flotyński, Krzysztof Walczak
Poznań University of Economics, Poland
email: {flotynski, walczak}@kti.ue.poznan.pl

***Abstract*—Dependencies between interactive 3D content elements are typically more complex than dependencies between standard web pages as they may relate to different aspects of the content—spatial, temporal, structural, logical and behavioural. The Semantic Web approach helps in making data understandable and processable for both humans and computers by providing common concepts for describing web resources. However, semantic concepts may be also used to improve the process of designing content. In this paper, a new approach to semantic multi-layered design of interactive 3D content is presented. The proposed solution provides a semantic representation of 3D content in multiple layers reflecting diverse aspects of 3D content. The presented solution conforms to well-established 3D content and semantic description standards and—therefore—may facilitate creation, dissemination and reuse of 3D content in a variety of application domains on the web.**

***Index Terms*—3D web, 3D content, semantic modelling, ontology, RDF, OWL, RDFS**

I. Introduction

WIDESPREAD use of interactive 3D technologies has been recently enabled by increasing hardware performance, rapid growth in the network bandwidth as well as availability of cheap 3D accelerators and input-output devices. However, the potential of 3D technologies in everyday use may be fully exploited only if accompanied by easy-to-use methods of creating, searching and combining distributed interactive three-dimensional content.

Creating, searching and combining distributed interactive 3D content are much more complex and challenging tasks than in the case of typical web pages. Relationships between components of an interactive three-dimensional virtual scene may include, in addition to its basic meaning and presentation form, also spatial, temporal, structural, logical, and behavioural aspects. The Semantic Web approach makes the described data understandable for both humans and computers achieving a new quality in building web applications that can ”understand” the meaning of particular components of content and services as well as their relationships.

Semantic Web standards may be applied to 3D content, leading to much better methods of creating, searching, reasoning, combining and presenting content. 3D content models based on commonly used semantic concepts may be independent of particular description methods and languages. The use of common concepts facilitates dissemination and reuse of individual components of the content that may be semantically assembled to create particular VR/AR presentations depending on the user location, preferences, device used, etc. Moreover, extending geometrical, structural, spatial, logical, and behavioural components with their semantic descriptions permits reasoning on complex semantically described 3D scenes that include these components.

Although a number of methods and languages for programming 3D presentations, and several solutions for creating semantic descriptions of 3D content have been proposed, they do not support layered design of interactive 3D presentations and do not enable semantic representation of components of 3D objects and scenes.

Layered design of 3D presentations provides a separation of concerns between particular semantic layers corresponding to different aspects of the presentation. It reduces the complexity and the number of connections between particular components of the content, facilitating their implementation and exchange, and simplifying the creation of the desirable final presentation.

The main contribution of this paper is an approach to semantic multi-layered representation of interactive 3D presentations. The proposed solution provides a complex representation of 3D content in multiple layers reflecting various aspects of the content—geometry, structure, appearance, scene, logic, and behaviour. The solution conforms to well-established 3D content and semantic description standards.

The remainder of this paper is structured as follows. Sections II and III provide an overview of the current state of the art in the domains of 3D content presentation and semantic description of web resources. Section IV introduces a novel semantic multi-layered model of 3D content. Section V discusses an implementation of the proposed approach. An example of the semantic design of 3D content is explained in Section VI. Finally, Section VII concludes the paper and indicates the possible directions of future research.

II. Interactive 3D Presentations

A number of technologies (including languages, libraries, frameworks, and game engines) have been devised for creating interactive 3D content presentations, in particular built into web applications.

The Virtual Reality Modeling Language (VRML) [1] is an open, textual language devised by the Web3D Consortium for describing static and animated 3D content in a declarative way. A VRML scene is represented as a graph with nodes reflecting different aspects of the described 3D content—geometry, structure, appearance, space, logic, and behaviour.

978-83-60810-53-8/$25.00

VRML also supports linking external multimedia resources—images, audio and video. In addition to the use of specific behavioural VRML nodes, the logic and behaviour of the presented 3D objects may be described by embedded imperative ECMAScript code. Several implementations of VRML browsers are available, e.g., ParallelGraphics Cortona3D [2], Bitmanagament BS Contact [3], FreeWRL [4], and InstantReality [5].

The Extensible 3D (X3D) [6] is a successor to VRML, also designed by the Web3D Consortium. X3D introduces several functional extensions to VRML, such as Humanoid Animation, NURBS and CAD geometry. Furthermore, it supports additional XML-based and binary encoding formats as well as basic means for metadata description. Depending on the set of implemented features, different X3D profiles may be selected for the presentation of particular 3D content. Currently, X3D is implemented by a few browsers, e.g., BS Contact, FreeWRL and InstantReality. VRML and X3D enable standardized presentation of 3D content on the web, accessible with additional browser plug-ins. To enable seamless integration of X3D content with web pages, X3DOM [7] has been designed. It is an open source framework intended as a potential extension to HTML5. The content encoded with X3DOM can be presented without additional plug-ins by the majority of modern web browsers.

PDF3D [8] is another approach to 3D content presentation. It utilizes the U3D [9] file format for model representation and a proprietary JavaScript API for programming its behaviour. A PDF document with 3D content may be directly embedded in a web page, and presented with the Abobe Reader plug-in.

A number of libraries have been developed for creating 3D content presentations. Such libraries usually permit programming of the logic and behaviour of the content, while 3D objects in the scene are represented by external resources, which are encoded in, e.g., JSON, COLLADA, AWD, Wavefront OBJ or 3DS. Several libraries have been implemented on the basis of JavaScript and OpenGL to enable 3D presentations built into web pages—WebGL [10], GLGE [11], JebGL [12], Oak3D, [13], and O3D [14]. Other libraries (Papervision3D [15], Alternativa3D [16], Away3D [17] or Sandy 3D [18]) have been developed for the ActionScript—an object-oriented dialect of ECMAScript that is used for web applications compatible with the Adobe Flash Player. Web presentations of 3D content can also be built using Java applets implemented with JOGL [19] or Java3D [20] libraries.

Another group of solutions incorporates game engines, which allow for the development of complex 3D web applications enriched with additional aspects, such as physics, collision detection, artificial intelligence and networking. For instance, Unity [21] and Unreal [22] permit 3D presentations accessible with the Unity Web Player and the Adobe Flash Player.

III. Semantic Descriptions of 3D Content

In this section, the state of the art in the area of semantic description of web content is presented. In particular, basic techniques for describing the semantics of web resources, metadata and ontologies for 3D multimedia content as well as methods of semantic creation of 3D content are considered.

A. Foundations for the Semantic Web

The primary technique for describing data semantics on the web is the Resource Description Framework (RDF) [23]—a standard devised by the W3C. RDF introduces general rules for making statements about resources. Each statement is comprised of three elements: *a subject* (a resource described by the statement), *a predicate* (a property of the subject) and *an object* (the value of the property).

RDF introduces classes (as types of resources), containers and lists to provide basic concepts for semantic descriptions. However, these notions are often insufficient for describing the semantics of complex resources. The RDF Schema (RDFS) [24] and the Web Ontology Language (OWL) [25] are W3C standards based on RDF providing higher expressiveness for semantic descriptions of web resources, e.g., hierarchies of classes and properties, constraints, property restrictions as well as operations on sets. OWL defines a set of profiles, which differ in complexity and decidability. Semantic Web Rule Language (SWRL) [26] is an extension to OWL devised for describing semantic Horn-like rules. While RDF and RDF-based techniques permit the creation of ontologies and knowledge bases, SPARQL [27] is a language for querying RDF data sources.

RDF and RDF-based technologies have been intended as the basis of the Semantic Web. Hence, they are applicable to any type of web resources, but they do not address specific aspects of particular content types (especially 3D). That is why application-specific ontologies are required to describe content of various types on the web.

B. Metadata and Ontologies for 3D Content

To provide a common space for the classes and properties of resources on the web, several vocabularies, metadata schemas and ontologies have been proposed for various application domains and types of resources, in particular for 3D multimedia content. The Multimedia Content Description Interface (MPEG-7) [28] is an extensive standard that defines a set of tools for creating metadata—Descriptors, Description Schemes, the Description Definition Language and Coding Schemes. There is a wide range of target media types that may be described with MPEG-7—images, audio, video and 3D objects, including multimedia content in VR applications [29].

A few ontologies have been proposed for multimedia content. The Ontology for Media Resources [30] has been devised by the W3C on the basis of RDF, RDFS and OWL, as a common solution for describing multimedia published on the web. It provides an interoperable core vocabulary that is mapped to a set of metadata formats for media content (e.g., MPEG-7). The Core Ontology for Multimedia (COMM) [31], [32] is another solution designed for describing media content such as images, audio, video and 3D objects. COMM is based on MPEG-7, but it represents knowledge with open Semantic Web

solutions avoiding some interoperability problems that occur in MPEG-7, e.g., with semantically equivalent descriptors that are processed in different manners [32].

C. Semantic Creation of 3D Scenes

Several works have been devoted to the semantic creation of 3D content. In [33], an approach to creating interoperable RDF-based Semantic Virtual Environments, with system-independent and machine-readable abstract descriptions has been presented. In [34][35][36], a rule-based framework based on MPEG-7 has been proposed for the adaptation of 3D content, e.g., geometry and texture degradation or filtering of objects. The content can be described with different encoding formats (in particular X3D), and it is annotated with an indexing model. In [37], an integration of X3D and OWL using scene-independent ontologies and the concept of semantic zones are proposed to enable querying 3D scenes at different levels of semantic detail and they have been used to implement a tour through the Venetian Palace.

In [38], a method of structured design of VR content has been proposed. In [39][40][41], an approach to generating virtual words upon mappings of domain ontologies to particular 3D representation languages (e.g., X3D) has been considered. The following three content generation stages are distinguished: specification of a domain ontology, mapping of the domain ontology to a 3D description language, and generation of the final presentation. The solution stresses spatial relations (position and orientation) between objects in the scene.

Several works have been conducted on the modelling of behaviour of VR objects. In [42], the Beh-VR approach and the VR-BML language have been proposed for the dynamic creation of behaviour-rich interactive 3D content. The proposed solution aims at simplification of behaviour programming for non-professionals. Another method facilitating the modelling of content behaviour [43][44][45] provides a means for expressing primitive and complex behaviours as well as a set of temporal operators. Finally, a rule-based ontology framework for feature modelling, consistency check and feature modelling, has been explained in [46].

IV. Semantic Model of Interactive 3D Content

Although several approaches have been devised for semantic modelling of 3D content, they lack solutions for semantic representation of 3D content. Layered design of 3D presentations provides a separation of concerns between particular layers corresponding to different aspects of the designed presentation, which are described in individual, specific manners. It reduces the complexity and the number of connections between components, which are incorporated in different layers, facilitating their implementation and exchange, and simplifying the creation of the desirable content. In addition, possible implementation profiles of a structured solution may cover only layers reflecting the required aspects of the context—geometry, structure, appearance, space, logic, or behaviour.

In this section, a novel approach to the semantic design of interactive 3D presentations is proposed. The presented solution is based on a multi-layered semantic representation of 3D content. The model complies with the Semantic Web approach, and it has several important advantages in comparison to the available solutions for modelling of 3D content. First, the components of semantically described content may be searched, explored and reused by applying well-established Semantic Web standards. Second, it allows for reasoning on the content, which further enables discovering knowledge that has not been explicitly encoded. Third, 3D content described by commonly used concepts is platform- and standard-independent, and it may be transformed to final presentations encoded in different languages, depending on particular requirements, e.g., the context of interaction, client device used, user preferences.

The proposed semantic model of 3D presentations is depicted in Fig. 1. It includes six layers corresponding to distinct aspects of 3D content and different stages of the development of 3D presentations—*Geometry Layer*, *Structure Layer*, *Appearance Layer*, *Scene Layer*, *Logic Layer*, and *Behaviour Layer*. The subsequent layers are partly dependent—every layer uses only its own concepts and the concepts specified in the lower layers (gray arrows), i.e. a 3D presentation may fully utilize the components of a particular layer without referring to its higher layers. Like in OWL, the concepts defined are classes as well as data properties and object properties describing respectively the attributes and relations between class instances. The specification of the relations between class instances in the class definitions indicates optional and obligatory dependencies between components, which are specified during the modelling process while creating instances of these classes.

In the proposed approach, a 3D presentation may be created at an arbitrary layer and the development process includes the creation of components which are defined in the selected layer and its lower layers. For instance, design of a complex 3D scene with behaviour covers all of the layers of the presented model. However, presentations that consist only of lower layers are also possible, e.g., reusable structural 3D objects without appearance that are to be injected into different complex presentations can be created at layer 2.

Two types of classes are distinguished—*abstract* and *concrete*. With the presented solution, a developer designs a presentation by creating instances of the selected concrete classes. The properties in the presented model are specified as optional or obligatory with a given cardinality. A component in the resulting scene is assigned the properties of its class and the superclasses. Created objects may be described with desirable data properties and linked one to another with object properties.

The proposed semantic model has been designed with regards to concepts commonly used in well-established 3D content representation languages and libraries, such as X3D, Unity, etc. In the diagram presented in Fig. 1, several data properties which are typical for different 3D content representation standards as well as exact data types and ranges of

properties have been omitted as they are not crucial for the proposed idea. The presented model contains key concepts used in designing 3D presentations, but it may be extended with new classes and properties depending on the particular use.

The proposed model is semantically complete—no classes or properties are created or removed during the content presentation, only the values of properties may change. For instance, a moving object may be stopped by turning off its motion animation, but not by removing it.

The following sections describe the semantic design of 3D presentations with regards to the particular layers of the proposed model. The design starts with the description of basic shapes included in the created presentation (the *Geometry Layer*). Second, the basic objects are assigned spatial properties to be combined into arbitrary complex structural objects—complex shapes with spatial dependencies (the *Structure Layer*). Third, appearance properties are added to the complex structural objects to create visual components, which may be illuminated and enriched with environmental effects (the *Appearance Layer*). Next, the components with appearance are included in a scene with a viewpoint and navigation modes (the *Scene Layer*). Finally, logic and behaviour may be added to the scene and all its components (the *Logic* and *Behaviour* layers).

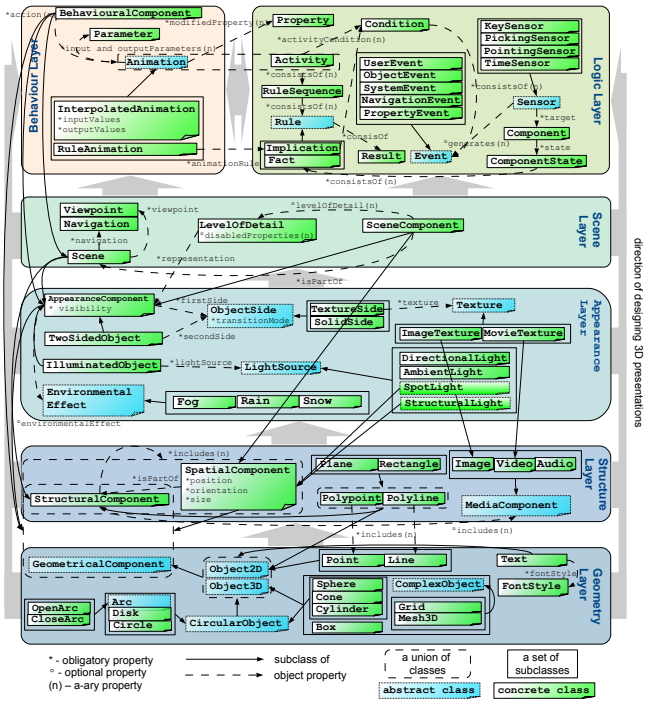


Fig. 1. The semantic model of interactive 3D content

A. Geometry Layer

The *Geometry Layer* is the base layer of the proposed semantic model and it includes concepts of basic 2D and 3D geometry shapes, which physically form the presentation. In the proposed model, this layer is aimed at the low-level manipulation of the geometry of scene components. The primary `GeometricalComponent` class is abstract, thus all shapes created are instances of its descendants. The classes of the *Geometry Layer* are relatively simple, in comparison to the classes defined in the other layers, as they are mainly determined by using data properties specific to 2D and 3D content. Since the components of this layer have no common point of spatial reference, their spatial properties (exact size, position and orientation) cannot be given. As this layer does not specify other aspects of 3D content (space, appearance, logic and behaviour), it allows only for modelling simple separated objects that do not combine into complex models and scenes. Hence, this is not sufficient for building practical 3D presentations. `GeometricalComponents` have limited expressiveness in 3D presentations, as they can only represent isolated integral objects, e.g., 3D models of sculptures or shapes of buildings.

B. Structure Layer

The *Structure Layer* is the second layer of the model and it depends only on the *Geometry Layer*. While the `GeometricalComponents` describe basic shapes included in a scene, `StructuralComponents` enable creating logical and spatial combinations of them into complex objects. `StructuralComponents` may recursively include other `StructuralComponents` as well as `Media` (`Images`, `Audio` and `Video`) and `Spatial` components. `SpatialComponents` are `Geometrical` and `Structural` components with spatial properties (`position`, `orientation` and `size`) set relatively to the parent `StructuralComponent`.

Thanks to the specific meaning of the `include` inverse functional property, `StructuralComponents` may be considered as a whole while assigning some properties (sub-properties of the `structurallyTransitive` property) in higher layers, e.g., appearance or spatial properties are automatically set for all the subcomponents of a complex `StructuralComponent` when the property is set to it.

Although the components of this layer have no appearance and thereby they are not sufficient for practical 3D presentations, they describe its structure and are managed from within higher layers, e.g., when determining logic and behaviour.

C. Appearance Layer

The *Appearance Layer* is aimed at adding appearance to `Geometrical` and `Structural` components that are defined in the previous layers. The primary `AppearanceComponent` may be either a single- or a two-sided object. Each side can be covered with textures (images or movies), or described by typical appearance properties (`colour`, `transparency`, etc.). The same appearance properties may be set for a whole `StructuralComponent` with all its subcomponents by specifying `transitionMode`—to ignore or respect the individual settings of subcomponents. In addition, `AppearanceComponents` may be illuminated

by LightSources of several types and enriched with EnvironmentalEffects. At this layer, a 3D presentation consists of a set of logically structured components with appearance that have no logic and behaviour, e.g., static museum artefacts, furniture, buildings, etc.

D. Scene Layer

The primary class of the *Scene Layer* is the Scene that is an AppearanceComponent or a StructuralComponent. It has assigned a list of Viewpoints and a list of Navigation modes. A background and EnvironmentalEffects of the Scene may be specified at the lower *Appearance Layer*. The subcomponents of Scene are SceneComponents inheriting from either Spatial or Appearance components. Complex SpatialComponents may be presented with different LevelsOfDetail depending on the current distance between the object and the observer. Each LevelOfDetail indicates a set of AppearanceComponents that are the ingredients of a particular target object, to be visible, and a set of appearance properties to be disabled. At this layer, designing complex navigable 3D presentations (without logic and behaviour) is feasible, e.g., static virtual museum exhibitions, models of cities, etc.

E. Logic Layer

The *Logic Layer* is intended as a framework providing concepts for describing the logic of components that are defined in other layers. This layer does not introduce apparent effects to 3D presentations, as opposed to the previous layers. As the logic may be related to different aspects of the created 3D content, the *Logic Layer* links all of the previous layers. This layer has been designed according to the rule-based approach that enables both complex declarative descriptions and reasoning on the created 3D presentations. The primary entity of logic description is a Rule that may be either a Fact or an Implication (if a Condition—a conjunction—is satisfied then a Result is also satisfied). While Facts describe the ComponentStates of any object defined in any layer of the presented semantic model, Implications are mainly used for describing complex RuleSequences. In a RuleSequence, the result of an Implication is the Condition of its following Implication. Such a chain permits an ordered performance of consecutive steps, like in typical imperative programming. In turn, several independent sequences may create an Activity that is initiated when a common required alternative of Conditions is satisfied. A Condition may refer to Events (generated either by a user interaction, an object, the navigation, the system, or a change of a property value). Events are generated by Sensors (e.g., KeySensor or PointingSensor for user interactions, PickingSensor for object interactions, TimeSensor for system interactions). Events and Sensors, as well as the relations between them defined in the model are similar to the corresponding concepts widely-used in other technologies for describing 3D, thus they are not described in detail in Fig. 1.

F. Behaviour Layer

The *Behaviour Layer* provides concepts that introduce apparent behavioural effects to 3D presentations built upon the previous layers. Like the *Logic Layer*, the *Behaviour Layer* leverages all the lower layers, as the behaviour of a component may concern different aspects of 3D content. In particular, selected classes of the *Logic Layer* are especially important for defining behaviour. The primary BehaviouralComponent class extends the Geometrical, Structural, Appearance components, or Scenes with an arbitrary number of actions—Activities or Animations. Activities, which are conditionally dependent on Events, enable programming interactions. While Activities are entirely defined in the *Logic Layer*, Animations are behavioural objects with sets of input and output Parameters used to control modifiedProperties. Two types of Animations may be distinguished. RuleAnimations utilize Implications to bind input and output Parameters in a functional manner. InterpolatedAnimations specify sets of values of input and output Parameters. The changes of classification attributes are gradual. For numeric attributes with continuous domains, intermediate input and output values that are not specified explicitly, are calculated from their neighbouring values. At this layer, the created presentations may be dynamic 3D scenes with components including all the aspects of 3D content. The presentation may change in time due to interactions between objects in the scene, user interactions, system interactions, etc.

V. Implementation of the Proposed Semantic Model

The proposed semantic model of interactive 3D content has been implemented as an ontology with well-established Semantic Web standards (RDF, RDFS, OWL, SWRL and SPARQL), using the Protege editor [47]. The implementation rules of the model are explained below.

A. Class Definitions

The classes of the model are described by the owl:Class type. Inheritance between classes is described by the rdfs:subClassOf property. In some cases, the subclasses of a class union have been defined, e.g., BehaviouralComponents may inherit either from Geometrical, Appearance, StructuralComponents, or Scenes; behavioural Actions may be either Activities or Animations.

B. Disjointness of Classes

In the presented model, the subclasses of a common superclass (surrounded by solid rectangles in the diagram) are mutually disjoint classes described with the owl:disjointWith property, e.g., 2D and 3D objects, media components, events.

C. Property Definitions

Data and object properties are reflected by the owl:DatatypeProperty and the owl:ObjectProperty, respectively. In some cases,

properties defined for a superclass need to be implemented individually for different subclasses, e.g., the `size` specifies two values (two dimensions) for 2D components, and three values (three dimensions) for 3D objects. Such dependencies have been encoded as hierarchies of properties using the `rdfs:subPropertyOf`.

D. Optionality and Cardinality of Properties

Some classes of the model may allow or require their instances to have selected properties specified with a particular cardinality, e.g., a `SpatialComponent` included in a `StructuralComponent` must be assigned a single `position`, a single `orientation` and a single `size`. Classes with such requirements have been indicated as subclasses or equivalent classes (`owl:equivalentClass`) of restricted superclasses (the `owl:Restriction` property). Optional unary, optional n-ary, obligatory unary, and obligatory n-ary properties of components are described, respectively, by the `owl:maxQualifiedCardinality`, `owl:someValuesFrom`, `owl:qualifiedCardinality`, and `owl:minQualifiedCardinality` properties set to 1. However, to check if the exact and maximal cardinality are satisfied, the closed world assumption should be made. This requirement can be met for 3D presentations designed strictly for the particular use, whose components come from limited or well-known sources.

E. Domains and Ranges of Properties

To unambiguously connect properties to the corresponding classes, domains and ranges have been determined for properties. In the vast majority of cases, domains and ranges enclose single classes, e.g., the sides of an `AppearanceObject`, the `Viewpoint` of a `Scene`. However, a few domains and ranges are unions of classes, e.g., a `consistOf` property may be indicated for an `Activity`, a `RuleSequence` or a `Rule`; an `includes` property can indicate a `Geometrical` or a `StructuralComponent`.

F. Structurally Transitive Properties

Structurally transitive properties influence not only the component they are specified for, but also the subcomponents it `includes`. Structurally transitive properties have been defined with SPARQL rules (Listing 1)—if a component has a property (line 2) that is a structurally transitive property (3), and the component has a subcomponent (4), then this subcomponent is also assigned the property with the same value (1).

In the presented semantic model, structurally transitive properties are related to appearance and spatial aspects of the content, e.g., colour, transparency, light sources (the parts of an illuminated object are also illuminated).

Listing 1. A SPARQL rule spreading structurally transitive properties

```
1 construct {?subcomponent prop ?value.}
2 where { ?component ?prop ?value.
3         ?prop rdfs:subPropertyOf STproperty.
4         ?component includes ?subcomponent. }
```

G. Logic Definitions

To enable complex descriptions of logic and reasoning on the content, components of the *Logic Layer* are implemented in SWRL. An `Implication` with a `Condition` and a `Result` is encoded as a `ruleml:imp` element containing a body (`ruleml:_body`) and a head (`ruleml:_head`). A `RuleSequence` is mapped to a list of `ruleml:imp` elements, in which the head of an implication is the body of the next one. Several independent sequences which are attainable by a common `Condition` (their initial rules have the same body) form an `Activity`.

VI. Example Layered Design of a 3D Artefact Model

In this section, an example of the use of the implemented semantic model is presented in the context of designing a 3D presentation of a virtual museum artefact, e.g., in development of educational games, creating commercial presentations, etc. The artefact is a complex static reusable 3D component that represents a bronze statue with a hat. The example focuses on modelling activities performed by a developer in the consecutive layers of the proposed model.

The design process is illustrated and based on an instance of the model (a knowledge base indicated by the `museum` prefix, Listing 2), which is compliant with the implemented model ontology (the `sm` prefix) and includes descriptions of the created objects, their properties and logic rules. The knowledge base is manually transformed to a concrete 3D scene. However, a tool for automatic transformation could be developed. In the presented example, the resulting 3D content description is encoded in X3D (Listing 3), however, other formats could be used as well.

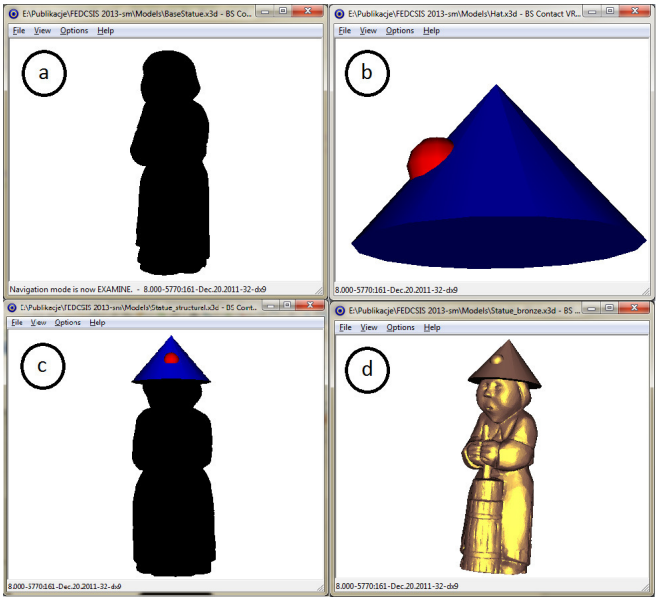


Fig. 2. The results of the consecutive stages of the example layered design of the 3D artefact: the geometry of the artefact (a), the hat element (b), the artefact as a structural component (c), the artefact as a component with appearance (d)

Fig. 2 depicts particular steps of the design process. During the first stage of the process, a `Mesh3D` geometrical object (Listing 2, line 17) is created by a scanner or a modelling tool (e.g., Blender or 3ds Max) to describe the shape of the body of the statue—its coordinates and coordinate indexes (2/20-21). The generated X3D code (3/18-19) leads to the view depicted in 2a. This is not a full 3D presentation due to the lack of properties describing the appearance—the created geometrical object is theoretically invisible.

In the second stage, a desirable `Hat` is found in another existing knowledge base and retrieved—e.g., by a SPARQL query with given conditions (Fig. 2b). In contrast to the body of the statue, the `Hat` is not limited only to geometry, but it has also structure, appearance, and spatial properties applied to its subcomponents (a brim and a decoration, Listing 2, lines 4,5). In the resulting X3D document the `Hat` is represented by a `Transformation` element with nested `Shape` elements (3/7-16).

Listing 2. An example of a semantically designed virtual 3D artefact

```
1  museum:Hat rdf:type owl:NamedIndividual ;
2    rdf:type sm:SpatialComponent ,
3    sm:SceneComponent ;
4    sm:includes museum:HatBrim ;
5    sm:includes museum:HatDecoration ;
6    sm:size "50 50 50" ;
7    sm:position "−5 100 −205" ;
8    sm:orientation "0 0 0" .
9  museum:BronzeStatue rdf:type owl:NamedIndividual ,
10   sm:StructuralComponent , sm:AppearanceComponent ,
11   sm:Scene .
12   sm:includes museum:BronzeStatueBody ;
13   sm:includes museum:Hat ;
14   sm:firstSide museum:BronzeStatueSide ;
15   sm:viewpoint museum:Viewpoint;
16   sm:navigation museum:Navigation .
17 museum:BronzeStatueBody rdf:type sm:Mesh3D ,
18   owl:NamedIndividual , sm:SpatialComponent ,
19   sm:SceneComponent ;
20   sm:coordinateIndex "..." ;
21   sm:coordinates "..." ;
22   sm:size "35 30 150" ;
23   sm:orientation "0 0 0" ;
24   sm:position "−5 100 0" .
25 museum:BronzeStatueSide rdf:type sm:SolidSide ,
26   owl:NamedIndividual ;
27   sm:color "0.65 0.45 0.4" ;
28   sm:transitionMode "override" .
29 museum:Viewpoint rdf:type sm:Viewpoint ,
30   owl:NamedIndividual ;
31   sm:position "−11.5858 −6.00235 1038.03" ;
32   sm:orientation "1.0 0.0 0.0 0.0" .
33 museum:Navigation rdf:type sm:Viewpoint ,
34   owl:NamedIndividual ;
35   sm:mode "examine" ;
36   sm:transitionType "linear" .
```

At the third stage, a structure is added to the designed model to represent its complexity. A new `StructuralComponent BronzeStatue` is created, and it includes both the `Hat` and the `BronzeStatueBody` (2/9-13). Since the `Hat` and the `Body` are parts of the `Statue` their `positions`, `orientations` and `sizes` need to be specified relatively to the parent `BronzeStatue` component (2/6-8,22-24). Then the subcomponents are inferred to be `SpatialComponents` (2/2,18), as they are defined as parts of a `StructuralComponent`. This structuralization leads to the X3D code (3/7-20), excluding an instruction describing the appearance of the statue body (3/17). At this stage of the design, the complex `BronzeStatue` may be illustrated as a complex component with spatial relations between its subcomponents (Fig. 2c).

At the next stage, appearance is specified for the whole `BronzeStatue` (2/14). The `BronzeStatueSide` (2/25-28) is a `SolidSide` component that defines a colour (2/27) used for the entire complex `BronzeStatue` overriding colours set for its subcomponents (2/28). The corresponding X3D `diffuseColor` attribute is set appropriately for both the `StatueBody` and the parts of the `Hat` (3/10,14,17). Since the appearance is specified for the whole statue (and not only for its `Hat`), it is inferred to be an `AppearanceComponent` (2/10), as it is a `StructuralComponent` with an `ObjectSide` assigned. The resulting X3D scene is depicted in Fig. 2d and Listing 3, lines 7-20—it is a reusable complex component with the appearance and the spatial relations between its subcomponents.

The last stage encloses activities performed in the *Scene Layer* and it extends the previous `AppearanceComponent` statue with a `Viewpoint` (2/29-32) and a descriptor of `Navigation` (2/33-36). With a `Viewpoint` and a `Navigation` mode specified, the `BronzeStatue` starts to be classified as a `Scene` (2/11), and its subcomponents—the `Hat` and the `Body`—as `SceneComponents` (2/3,19). The corresponding X3D description contains all the instructions in Listing 3, including lines 5,6.

Listing 3. The final representation of the virtual 3D artefact in X3D

```
<?xml version="1.0" encoding="UTF−8" standalone="no"?>                          1
<!DOCTYPE X3D SYSTEM "x3d−3.0.dtd">                                             2
<X3D profile="Immersive" version="3.0">                                         3
 <Scene>                                                                        4
  <Viewpoint position="−11.5858 −6.00235 1038.03" orientation="1 0 0 0" />      5
  <NavigationInfo type="EXAMINE" transitionType="LINEAR" />                     6
  <Transform><Transform>                                                        7
   <Transform rotation="−1.0 0.0 0.0 1.5708" translation="−5 100 −205 " scale="  8
         50 50 50">
    <Shape><Cone height='1.2' />                                                9
     <Appearance><Material diffuseColor="0.65 0.45 0.4" /></Appearance>          10
   </Shape></Transform>                                                         11
   <Transform rotation="−1.0 0.0 0.0 1.5708" translation="−5 125 −200 " scale="  12
         50 50 50">
    <Shape><Sphere radius='0.2' />                                              13
     <Appearance><Material diffuseColor="0.65 0.45 0.4"/></Appearance>           14
   </Shape></Transform>                                                         15
   </Transform>                                                                 16
   <Shape><Appearance><Material diffuseColor="0.65 0.45 0.4" />                  17
    </Appearance><IndexedFaceSet coordIndex="...">                              18
    <Coordinate point="..." /></IndexedFaceSet>                                 19
</Shape></Transform></Scene></X3D>                                              20
```

VII. Conclusions and Future Works

In this paper, a novel approach to the semantic multi-layered design of interactive 3D presentations has been proposed. The presented division of the structure of 3D content into several distinct semantic layers facilitates the content creation process at the level of 3D model representation. However, the considered model does not address 3D content creation at an arbitrary high level of semantic abstraction, in particular by the use of domain concepts and ontologies. In addition, the presentations created with the model require explicit specification of all the components and relationships between them, which need to be presented in the resulting scene. Methods of semantic modelling and composition of 3D content at an arbitrarily

high level of abstraction may be proposed to permit implicit conditional query-based assembly of complex 3D scenes.

Other possible directions of future research incorporate several facets. First, although the presented approach is independent of any modelling tools, the use of semantic editors (e.g., Protege [47]) is highly recommended as it significantly facilitates working with the utilized Semantic Web standards. A specific development environment may be devised to support designing 3D presentations with regard to the consecutive layers of the presented model. Second, the implementation of the model should be evaluated and compared to other platforms in terms of the simplicity of 3D content creation. Furthermore, translators for selected target languages and technologies might be implemented, e.g., Java3D or Unity. To permit semantic exploration of 3D content in real-time, a persistent mapping between the primary semantic representations and the generated final scenes encoded in particular 3D content representation languages should be elaborated. Finally, the context of user–system interaction (e.g., user location, preferences, client device, etc.) can be semantically modelled to enable multi-platform 3D content presentations.

VIII. Acknowledgements

This research work has been partially funded by the Polish National Science Centre grant No. DEC-2012/07/B/ST6/01523.

References

[1] ISO/IEC 14772-1:1997. The Virtual Reality Modeling Language *http://www.web3d.org/x3d/specifications/*. Retrieved May 17, 2013.
[2] Cortona3D, *http://www.cortona3d.com/*. Retr. May 17, 2013.
[3] Bitmanagement, BS Contact, *http://www.bitmanagement.com/products/interactive-3d-clients/bs-contact*. Retrieved May 17, 2013.
[4] FreeWRL, *http://freewrl.sourceforge.net/*. Retrieved May 17, 2013.
[5] Instantreality Framework 2.0. Fraunhofer IGD (2011) *http://instantreality.de/home/*. Retrieved May 17, 2013.
[6] ISO/IEC 19775-1:2008. Extensible 3D (X3D) (2008) *http://web3d.org/x3d/specifications/*. Retrieved May 17, 2013.
[7] X3DOM, *http://www.x3dom.org/*. Retrieved May 17, 2013.
[8] Visual Technology Services, PDF3D—3D Visualization and Technical Publishing Technology, *http://www.pdf3d.com/*. Retrieved May 17, 2013.
[9] U3D File Format, *http://ecma-international.org/*. Retrieved May 17, 2013.
[10] WebGL Specification, *https://www.khronos.org/registry/webgl/specs/1.0/*. Retrieved May 17, 2013.
[11] GLGE, *http://www.glge.org/*. Retrieved May 17, 2013.
[12] JebGL, *http://http://jebgl.com/*. Retrieved May 17, 2013.
[13] Oak3D, *http://www.oak3d.com/*. Retrieved May 17, 2013.
[14] O3D, *http://code.google.com/intl/pl/apis/*. Retrieved May 17, 2013.
[15] Papervision3D, *http://papervision3d.org/*. Retrieved May 17, 2013.
[16] Alternativa, *http://alternativaplatform.com/en/*. Retr. May 17, 2013.
[17] Away3D, *http://away3d.com/*. Retrieved May 17, 2013.
[18] Sandy 3D, *http://www.flashsandy.org/*. Retrieved May 17, 2013.
[19] JOGL, *https://jogamp.org/jogl/www/*. Retrieved May 13, 2013.
[20] Java3D, *http://www.oracle.com/technetwork/java/javase/tech/index-jsp-138252.html*. Retrieved May 13, 2013.
[21] Unity. *http://unity3d.com/*. Retrieved May 13, 2013.
[22] Unreal GE. *http://www.unrealengine.com/*. Retr. May 13, 2013.
[23] Resource Description Framework (RDF). *http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/*. Retrieved May 17, 2013.
[24] Resource Description Framework Schema. *http://www.w3.org/TR/2000/CR-rdf-schema-20000327/*. Retrieved May 17, 2013.
[25] OWL Web Ontology Language Reference. *http://www.w3.org/TR/owl-ref/*. Retrieved May 17, 2013.
[26] SWRL: A Semantic Web Rule Language Combining OWL and RuleML. *http://www.w3.org/Submission/SWRL/*. Retrieved May 17, 2013.
[27] SPARQL Query Language for RDF. *http://www.w3.org/TR/rdf-sparql-query/*. Retrieved May 17, 2013.
[28] Inf. tech. Mult. content desc. interface Part 10: Schema def. *http://www.chiariglione.org/ mpeg/working_documents/mpeg-07/schema_def/cd.zip*. Retrieved May 17, 2013.
[29] Walczak K., Chmielewski J., Stawniak M., Strykowski S., Extensible Metadata Framework for Describing Virtual Reality and Multimedia Contents. In: *Proc. of the 7th IASTED Int. Conference on Databases and Applications*, Innsbruck, Austria, February 14-16, 2006, pp. 168-175.
[30] Ont. for Media Resources 1.0. *http://www.w3.org/TR/mediaont-10/*. Retrieved May 17, 2013.
[31] Core Ont. for Mult.. *http://comm.semanticweb.org/*. Retrieved May 17, 2013.
[32] Arndt R., Troncy R., Staab S., Hardman L., Vacura M., COMM: designing a well-founded multimedia ontology for the web. In: *Proc. of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference*, Busan, Korea, November 11-15, 2007, pp. 30-43.
[33] Otto K. A., Semantic Virtual Environments. In: *Special interest tracks and posters of the 14th international conference on World Wide Web*, Japan, May 10-14, 2005, pp. 1036-1037.
[34] Bilasco I. M., Villanova-Oliver M., Gensel J., Martin H., Semantic-based Rules for 3D Scene Adaptation. In: *Proc. of the twelfth int. conf. on 3D web technology*, Umbria, Italy, April 15-18, 2007, pp. 97-100.
[35] Bilasco I. M., Gensel J., Villanova-Oliver M., Martin H., On Indexing of 3D Scenes Using MPEG-7, In: *Proc. of the 13th annual ACM int. conf. on Multimedia*, Singapore, November 06-12, 2005, pp. 471-474.
[36] Bilasco I. M., Gensel J., Villanova-Oliver M., Martin H., 3DSEAM: a model for annotating 3D scenes using MPEG-7, In: *Proceedings of the eleventh international conference on 3D web technology*, Columbia, MD, USA, April 18-21, 2006, pp. 65-74.
[37] Pittarello F., Faveri A., Semantic description of 3D environments: a proposal based on web standards, In: *Proc. of the 11th Int. Conf. on 3D web Techn.*, Columbia, USA, April 18-21, 2006, pp. 85-95.
[38] Walczak K., Flex-VR: Configurable 3D Web Applications. In: *Proceedings of the International Conference on Human System Interaction HSI 2008*, Krakw, May 25-27, 2008, pp. 135-140, ISBN: 1-4244-1543-8.
[39] De Troyer O., Bille W., Romero R., Stuer P., On Generating Virtual Worlds from Domain Ontologies, In: *Proc. of the 9th Int. Conf. on Multi-Media Modeling*,Taipei,Taiwan, 2003, pp. 279-294, ISBN 957-9078-57-2.
[40] Bille W., Pellens B., Kleinermann F., De Troyer O., Intelligent Modelling of Virtual Worlds Using Domain Ontologies, In: *Proc. of the Workshop of Intelligent Computing (WIC), held in conjunction with the MICAI 2004 conference*,Mexico City,Mexico,2004, pp. 272-279, ISBN 968-489-024-9.
[41] De Troyer O., Kleinermann F., Pellens B., Bille W., Conceptual modeling for virtual reality, In: *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling - Vol. 83*, Australian Computer Society, Inc. Darlinghurst, Australia, 2007, pp. 3-18, ISBN: 978-1-920682-64-4.
[42] Walczak K., Beh-VR: Modeling Behavior of Dynamic Virtual Reality Contents. In: *Proc. of the 12th Int. Conf. on Virtual Systems and Multimedia VSMM 2006*, in: *H. Zha et al. (Eds.): VSMM 2006, LNCS 4270*, Springer Verlag Heidelberg 2006, pp. 40-51.
[43] Pellens B., De Troyer O., Bille W., Kleinermann F., Romero R., An Ontology-Driven Approach for Modeling Behavior in Virtual Environments, In: *Proceedings of Ontology Mining and Engineering and its Use for Virtual Reality (WOMEUVR 2005)*, Springer-Verlag, Agia Napa, Cyprus, 2005, pp. 1215-1224, ISBN 3-540-29739-1.
[44] Pellens B., De Troyer O., Bille W., Kleinermann F., Conceptual Modeling of Object Behavior in a Virtual Environment, In: *Proc. of Virtual Concept 2005*, Springer-Verlag, Biarritz, France, 2005, pp. 93-94.
[45] Pellens B., Kleinermann F., De Troyer O., A Development Environment using Behavior Patterns to Facilitate Building 3D/VR Applications, In: *In Proceedings of the 6th Australasian Conference on Interactive Entertainment*, ACM, Sydney,Australia 2009, ISBN 978-1-4503-0010-0.
[46] Zaid L. A., Kleinermann F., De Troyer O., Applying semantic web technology to feature modeling, In: *Proceedings of the 2009 ACM symposium on Applied Computing*, Honolulu, Hawaii, USA, 2009, pp. 1252-1256, ISBN: 978-1-60558-166-8.
[47] Protege, *http://protege.stanford.edu/*. Retrieved May 18, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 549–556

Microformat and Microdata Schemas for Interactive 3D Web Content

Jakub Flotyński, Krzysztof Walczak
Poznań University of Economics, Poland
email: {flotynski, walczak}@kti.ue.poznan.pl

***Abstract*—The paper presents new Microformat and Microdata schemas for creating descriptions of interactive 3D web content. Microformats and Microdata are increasingly popular solutions for creating lightweight attribute-based built-in semantic metadata of web content. However, although Microformats and Microdata enable basic description of media objects, they have not been intended for 3D content. Describing 3D components is more complex than describing standard web pages as the descriptions may relate to different aspects of the 3D content—spatial, temporal, structural, logical and behavioural. The main contribution of this paper are new Microformat and Microdata schemas for describing 3D web components and 3D scenes with metadata and semantic properties. The proposed schemas may be combined with X3D, a well-established 3D content description standard. Thanks to the use of the standardized solutions, the presented approach facilitates widespread dissemination of 3D content for use in a variety of multimedia applications on the web.**

***Index Terms*—3D content, semantic metadata, Microformats, Microdata, X3D**

I. Introduction

INTERACTIVE 3D technologies have enabled significant progress in the quality and functionality of human-computer interfaces. Widespread use of interactive 3D technologies, including virtual reality (VR) and augmented reality (AR), has been recently enabled by increasing hardware performance, availability of versatile input-output devices, as well as rapid growth in the available network bandwidth. However, the potential of 3D/VR/AR technologies in everyday applications can be fully exploited only if accompanied by the development of efficient and easy-to-use methods of creation, publication and sharing of interactive 3D multimedia content.

Building, searching and combining distributed three-dimensional interactive content are much more complex and challenging tasks than in the case of typical web pages. The relationships between components of an interactive three-dimensional virtual scene may include, in addition to its basic meaning and presentation form, also spatial, temporal, structural, logical, and behavioural aspects.

The aforementioned problems may be alleviated by describing 3D content with appropriate metadata and semantic properties. Research on the Semantic Web was initiated by T. Berners-Lee and the W3C (World-Wide Web Consortium) in 2001. This research aims at evolutionary development of the current web towards a distributed semantic database linking structured content and documents. Semantic description of web content makes it understandable for both humans and computers achieving a new quality in building web applications that can ”understand” the meaning of particular components of content and services as well as their relationships, leading to much better methods of searching, reasoning, combining and presenting web content.

On the basis of Semantic Web recommendations such as the Resource Description Framework (RDF) [1], the RDF Schema [2] and the Web Ontology Language (OWL) [3], a number of vocabularies, schemas and ontologies have been devised for a variety of application domains, in particular for multimedia systems, e.g., the Multimedia Content Description Interface [4], the Ontology for Media Resources [5] and the Core Ontology for Multimedia [6]. Available approaches to creating semantic descriptions of media content introduce a number of common attributes convenient for the general type of web resources, e.g., identifier, title, description, contributor, etc. In addition, they provide specific classes and properties intended for images, audio and video, but not for complex 3D web components, which may be described by multiple specific properties such as interactivity, animations, illumination, levels of detail, etc. Such metadata properties may be useful for exploration and analysis of 3D content, in particular for multimedia retrieval and optimization of queries for 3D content by providing values of attributes that are relatively constant and whose calculation is time-consuming.

Microformats [7] and Microdata [8] are increasingly used approaches to creating built-in semantic descriptions of web content with schemas defined in common repositories on the web. Embedding metadata directly in web content has a few important advantages in comparison to approaches that decouple resources from their descriptions. First, with embedded metadata, resources are unambiguously and inextricably linked with their descriptions. Second, it enables more concise descriptions and faster and less complicated authoring and analysis of semantically described content. Furthermore, it facilitates combining the semantic descriptions of resources with descriptions of web pages that embed the resources. Finally, it permits storage of content in structurally simpler databases. However, although existing Microformats and Microdata enable basic semantic descriptions of several types of multimedia objects, such as images, audio and video, they do not provide support for describing 3D content.

The main contribution of this paper are new Microformat and Microdata schemas for creating semantic metadata of interactive 3D web components and 3D scenes. The proposed

978-83-60810-53-8/$25.00

schemas facilitate indexing and retrieval of 3D content that meets specific criteria. The schemas include a number of specific properties that may be useful for contextual 3D content presentation dependent on, e.g., hardware/software platform, user–system interaction paradigms, user preferences, etc. The schemas may be combined with Extensible 3D (X3D) [9]—a well-established 3D content description standard, but they are not limited to this language. Thanks to the use of standardized solutions, the proposed approach enables flexible description and widespread dissemination of 3D content for use in a variety of multimedia web systems, e.g., in cultural heritage, education, simulations, geospatial visualisations, etc.

The remainder of this paper is structured as follows. Section II provides an overview of the state of the art in the domain of semantic and metadata descriptions of web resources, in particular 3D web content. In Section III, new Microformat and Microdata schemas are proposed for creating built-in semantic metadata of interactive 3D web content. Section IV describes a possible application of the proposed schemas in a system for searching 3D models by their metadata properties. Finally, Section V concludes the paper and indicates the possible directions of future research.

II. Semantic Descriptions of Multimedia Web Content

In this section, the state of the art in the field of semantic and metadata descriptions of multimedia web content is presented. In particular, metadata and ontologies as well as Microformats and Microdata for describing multimedia content are discussed. Next, methods of creating attribute-based embedded semantic and metadata descriptions of interactive 3D web content are considered.

A. Metadata and Ontologies for Multimedia Content

Several vocabularies, metadata schemas and ontologies have been proposed for describing multimedia content. DIG35 [10] defines metadata schemas for digital images. CableLabs [11] introduces vocabularies for both images and videos. The QuickTime File Format Specification [12] provides a schema for describing movie files. The Multimedia Content Description Interface (MPEG-7) [4] is a standard that defines a set of sophisticated tools for creating metadata—Descriptors, Description Schemes, the Description Definition Language and Coding Schemes. There is a wide range of target multimedia content that may be described with MPEG-7, including images, audio, video and 3D objects [13][14], however the standard is strongly focused on audio-visual data. The standards mentioned above typically include a number of generic properties (e.g., resource identifier, title, description, contributor, etc.), and a number of specific properties for describing images, audio and video, but not for complex interactive 3D web content.

A few ontologies have been proposed for multimedia content. The Ontology for Media Resources [5] has been devised by the W3C on the basis of the Resource Description Framework (RDF) [1], the RDF Schema [2] and the Web Ontology Language (OWL) [3] as a common solution for describing multimedia published on the web. It provides an interoperable core vocabulary that is mapped to a set of metadata formats for media content (e.g., DIG35, CableLabs and MPEG-7). A number of concepts defined in this ontology are common for web content of different types. There is a limited set of attributes typical for multimedia content, e.g., frameSize, compression, duration and samplingRate. This ontology lacks classes and properties typical for interactive 3D content, such as illumination, animations, navigation, levels of detail, etc.

The Core Ontology for Multimedia (COMM) [6][15] is another solution designed for describing media content. COMM is based on MPEG-7, but it represents knowledge with open Semantic Web solutions avoiding some interoperability problems that occur in MPEG-7, e.g., with semantically equivalent descriptors that are processed in different manners [15]. This ontology is convenient for describing images, audio and video, but it contains only a limited set of concepts suitable for interactive 3D web content.

Some other works are devoted to metadata for describing interactivity of 3D objects [16] and their interfaces [17]. Such descriptions may be used for finding 3D components by their properties [18]. In [19], metadata schemas for media objects have been proposed in the context of teaching architecture. The Metadata 3D Initiative [20] is a project in which a number of companies and research centres collaborate on the standardization of schemas for 3D (stereoscopic) content to make interoperable lenses, cameras, rigs, stereoscopic image processors, etc. In [21], the Multimedia Web Ontology Language, an extension to OWL, designed for creating ontologies and models for probabilistic reasoning in multimedia processing is presented.

B. Microformat and Microdata Schemas for Multimedia Content

Microformats [7] and Microdata [8] are solutions that are currently increasingly used for encoding semantic metadata of web content. In contrast to RDF-based approaches, which enable semantic descriptions with ontologies distributed across the web, Microformats and Microdata permit rapid creation of lightweight built-in semantic descriptions of content with schemas defined in common repositories on the web [7][22]. Such descriptions may be understandable for widely-used web search engines, such as Google, Yahoo and Bing. Currently, both Microformats and Microdata provide a variety of schemas for describing different types of web resources, in particular for multimedia content including images, audio and video.

The `hMedia` Microformat [23] is used for describing images, audio and video with a common set of properties. This schema is convenient for creation of general semantic descriptions that do not go into specific details with regard to individual media types. The `hAudio` [24] has been designed for describing audio content. There are no specific Microformats for describing images, video and 3D content.

In comparison to Microformats, Microdata provides a few more compound schemas, which form a hierarchical structure

with inheritance of properties. The root of this hierarchy is the `MediaObject` [25] schema, which defines a set of properties common for different types of media objects (like the `hMedia` Microformat)—images, audio and video. In addition, more extensive semantic and metadata descriptions of the particular media types may be created with descendant schemas of the `MediaObject`—`ImageObject` [26], `AudioObject` [27] and `VideoObject` [28]. Like the approaches mentioned above, Microdata does not provide schemas and properties sufficient for creating metadata and semantic descriptions of interactive 3D web content.

C. Attribute-based Embedded Metadata for 3D Web Content

In [29][30], a method of creating lightweight attribute-based semantic descriptions built into interactive 3D web content has been presented. The method enables metadata and semantic descriptions of both real objects and their virtual 3D counterparts by putting metadata into individual X3D metadata nodes. The resulting metadata and semantic descriptions of 3D web content are equivalent to Microformat and Microdata descriptions of typical web pages in terms of expressiveness. Moreover, such descriptions may use the same schemas, thus the method permits bidirectional transformation between descriptions that are built into web pages and descriptions embedded in 3D web content. Due to the use of the standard syntax and structure of X3D documents, the compatibility of the proposed approach with available X3D browsers is preserved.

The proposed approach is depicted in Fig. 1. The primary entity of the semantic description of 3D content is an X3D `MetadataSet` node. Since the method enables semantic descriptions of both real objects and their virtual 3D counterparts, both types of resources are referenced in the same manner—by their URIs. If a particular 3D component is to be described, it has to be assigned a URI in the X3D `DEF` attribute. The `name` attribute of the `MetadataSet` indicates a list of types of the described item, each of which may determine a set of semantic metadata properties. New item properties may be added to a `MetadataSet` independently of the schemas used. The optional `value` attribute specifies the URI (navigable or non-navigable) of the described object. The `reference` attribute of the `MetadataSet` contains a list of references to attributes that have been specified in other semantic descriptions and need to be shared with the primary one.

In addition to specifying the type and the URI of an item, the `MetadataSet` serves also as a container for item properties and relationships with other resources, which are reflected by nested typed metadata nodes (integer, float, double, string). The name and value of a property/relationship are given by the `name` and `value` attributes, respectively. The `reference` attribute is used to distinguish a property from a relationship. Metadata describing a property (e.g., data type) or a relationship may be contained in an additional metadata node of a desirable type, which is nested in the property/relationship typed element.

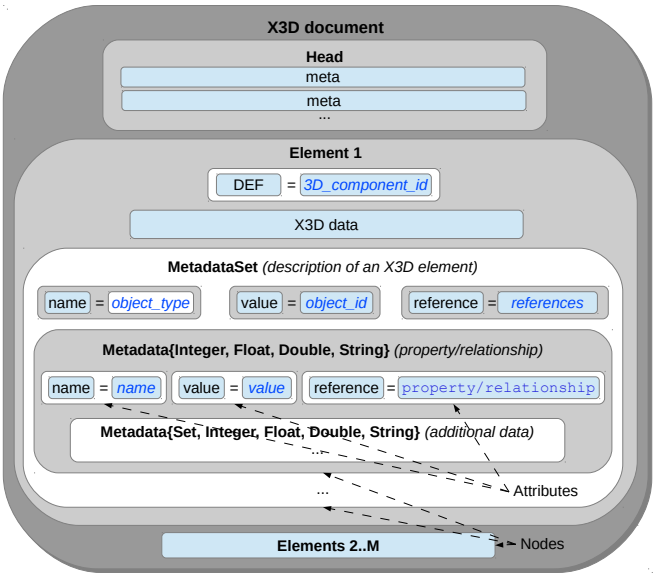


Fig. 1. Attribute-based embedded semantic descriptions of 3D content

In [31], a novel method of harvesting embedded attribute-based semantic metadata descriptions from distributed X3D web content has been proposed. The presented solution is an XSLT-based equivalent to the GRDDL [32] approach (originally intended for typical web pages), which has been designed for selection and processing of semantic descriptions of complex 3D/VR/AR scenes and components distributed across the web. The selection is based on the actual format, type and structure of the components, which are determined by their syntax. Filtering of the metadata to be extracted is necessary for reducing large semantic descriptions of complex 3D/VR/AR content to excerpts relevant to a particular application. The harvesting of metadata is a preliminary stage of the semantic analysis of the content, and it may precede the following activities, such as loading the generated semantic descriptions into a database and querying the system for semantically described 3D components.

III. Metadata Schemas for 3D Content

The aforementioned approaches address different aspects of creating metadata and semantic descriptions of web content, but they do not provide metadata schemas convenient for describing interactive 3D web content. To enable semantic descriptions of interactive 3D web components and complex 3D scenes, new metadata schemas are proposed in this section. They are intended to facilitate indexing, exploration and analysis of 3D content, and searching for 3D components and scenes described with embedded metadata attributes. Furthermore, the schemas include a number of specific properties that may be useful for contextual 3D content presentation dependent on, e.g., hardware/software client platform, user–system interaction paradigms, user preferences.

First, a classification of semantic metadata properties of interactive 3D web content is introduced. Then, new Micro-

format and Microdata schemas are proposed for describing 3D content with metadata and semantics.

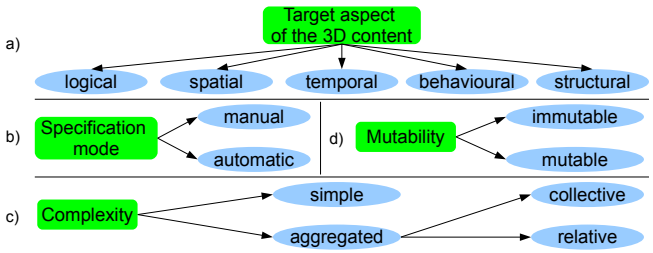


Fig. 2. The classification of metadata properties of interactive 3D web content in terms of the target aspect of the described 3D content (a), the specification mode of the property (b), the complexity of the property (c), and the mutability of the property (d)

A. Classification of Metadata Properties of Interactive 3D Web Content

The classification of metadata properties of interactive 3D web content is depicted in Fig. 2. The proposed classification is discussed in the four following aspects.

1) Target aspect of the 3D content: The metadata properties may be divided into different groups in terms of the described target aspect of the 3D content—*logical* (e.g., content description, presented object), *spatial* (e.g., dimensions, levels of detail), *temporal* (e.g., duration of an animation), *behavioural* (e.g., interactivity, animations) or *structural* (e.g., 3D sub-components).

2) Specification mode: The metadata properties may be distinguished in terms of the mode of specification, which may be *manual* or *automatic*. The first group incorporates properties that cannot be derived from the described 3D content and must be specified manually, in particular, by the author of the 3D content, e.g., the URI of the prototype of the described 3D component, the description of the component, and the semantic roles of its sub-components. The second group includes properties that can be automatically derived from the described 3D content, e.g., dimensions, animations, interactions.

3) Complexity: *Simple* and *aggregated* properties may be distinguished. *Simple* properties may be automatically retrieved from the described 3D component without appealing to its sub-components, e.g., the background of the scene or navigation modes. The main advantage of extracting *simple* properties is a possibility to reference them in queries that are built with a query language.

In contrast to *simple* properties, *aggregated* metadata properties are set with regard to the sub-components of the described 3D content, as opposed to other media types that do not incorporate a complex hierarchy of nested objects. *Aggregated* properties are determined by recursive processing and analysis of the content, which may be difficult and time-consuming, e.g., determining the number of levels of detail, animation types, light sources for the described content and all its sub-components. The main advantage of using *aggregated* metadata properties is that the calculation of them is performed once, before the described 3D content is loaded into a system. The presented approach may be preferred for 3D content that is rarely modified, for which the calculated properties are valid for a relatively long time. In such cases, the presented approach accelerates queries sent to the system, which do not need to initiate time-consuming processing of the content.

Two types of *aggregated* automatically determined metadata properties may be distinguished—*collective* and *relative* properties. *Collective* properties are specified by the analysis of sub-components of the described 3D component and determining a single aggregated value or a list of aggregated values. The results are related to the described 3D content as a whole, e.g., the dimensions, mass or volume of the described component are calculated as the sums of the dimensions, masses and volumes of its particular sub-components.

Like *collective* properties, *relative* properties are determined with regard to the sub-components of the described component, but the value of a *relative* property always describes either a relationship between the described component and its particular sub-components, or a relationship between particular sub-components of the described component. Such properties are convenient for expressing relative physical quantities, e.g., collision detection, velocity, acceleration, angular velocity, angular acceleration that are determined only by relationships between particular objects. In contrast to *collective* properties, *relative* properties are not aggregated.

4) Mutability: Metadata properties may be categorized in terms of the mutability into *immutable* and *mutable* properties. The first group comprises properties that do not change during the content presentation, e.g., the URI of the object presented by the described component or the names of additional packages required for correct content presentation. The second group contains properties that potentially can (but do not have to) change, e.g., dimensions, mass, illumination may change because of disappearing of some sub-components, switching off some light sources, etc. In the presented approach, it is not assumed that *mutable* properties are up to date for the whole period of the 3D content presentation. Instead, an author of the described 3D content may arbitrarily select a condition or a point in time for which the property is specified. It is recommended to select a specific condition/point in time, e.g., the start or the end of an animation that modifies the described property, and to indicate the selected point by the metadata of this property.

B. Metadata Properties of the Proposed Schemas

The list of metadata properties of the proposed `Microformat3D` and `Microdata3D` schemas are presented in Table 1. In the table, the aspect of 3D content (logical, spatial, temporal, behavioural and structural), the name and data types are provided together with the complexity, mutability and description for each property. In the *Type* column, the first data type is specified according to the proposed Microformat, the second one—to the proposed Microdata schema.

Target aspect of the 3D content	Property	Type	Complexity	Mutability	Description
logical	presentedObject	string	simple	immutable	A uniform identifier of the object (prototype) that is presented by the described 3D content. The presented object may be either real or virtual with a URI specified, e.g., as an HTTP address.
	packages	string[]	collective	immutable	The list of the names or URIs of additional packages necessary for the presentation of the whole described 3D content including all its sub-components. It is specific to the particular 3D content description standard used. For instance, X3D content may require such additional packages as Geospatial, NURBS, Human Animation, Distributed Interactive Simulation or CAD.
spatial	dimensions	double[] / float[]	collective	mutable	The width, height and length of the described 3D component including all its sub-components.
	mass	double / float	collective	mutable	The mass of the described 3D component including all its sub-components.
	volume	double / float	collective	mutable	The volume of the described 3D component including all its sub-components.
	fog	double / float	collective	mutable	The minimal visibility range for the described 3D scene and all its sub-components.
	background	string	simple	mutable	The name/value of the color or a URI of the image that is used as the background of the described 3D scene.
	illumination	(string, long)[]	collective	mutable	The list of types and numbers of light sources of the particular type, which are used in the scene and its sub-components. The light types may be, e.g., `point`, `spot`, `directional`, `area`, `model`, `ambient`.
	levelsOfDetail	(long, double/float)[]	collective	mutable	The list of the levels of detail used for the described 3D scene. Each level is specified by the number of polygons and a range. If multiple sub-components are combined within the described complex 3D scene, this parameter should be a combination of the levels of detail with regard to these sub-components.
	collisions	string[][]	relative	mutable	The list of the sets of URIs of 3D sub-components of the described 3D scene, for which collision detection is enabled.
temporal and behavioural	interactivity	string[]	collective	mutable	The list of user interactions allowed for the described 3D component and its sub-components, e.g., `selection`, `manipulation`, `navigation`, `system control`, `symbolic input`, etc.
	navigation	string[]	simple	mutable	The list of navigation modes allowed for the described 3D scene, e.g., `any`, `fly`, `walk`, `examine`, `lookat`, `slide`, `rotate`, `pan`, `game`, `jump`, `none`.
	animations	string[]	collective	mutable	The list of animation types used in the described 3D scene and its sub-components, e.g., `position`, `orientation`, `scale`, `structure`, `shape`, `appearance`, etc. For complex 3D scenes, the list should include animations in relationships between particular sub-components of the scene as well as between the scene and its sub-components, e.g., `position` animation.
structural	imageComponents	(string, string)[]	collective	immutable	The list of image components (textures) that are linked to the described component and its sub-components, with their URIs and semantic roles in the described 3D content, e.g., a texture of a dish, a sculpture, an exhibit, etc.
	audioComponents	(string, string)[]	collective	immutable	The list of audio components that are linked to the described component and its sub-components, with their URIs and semantic roles in the described 3D content, e.g., a background sound, a piano sound, etc.
	videoComponents	(string, string)[]	collective	immutable	The list of video components that are linked to the described component and its sub-components, with their URIs and semantic roles in the described 3D content, e.g., a projection, a movie, etc.
	3DComponents	(string, string)[]	collective with relative properties	immutable	The list of 3D sub-components with their URIs and semantic roles in the described 3D component, e.g., artefact, exhibition stand, wall, floor, furniture, etc.

Table 1. Semantic metadata properties of the proposed Microformat3D and Microdata3D schemas

The proposed Microformat and Microdata schemas are partially based on metadata properties devised in previous research works, e.g., the ARCO (Augmented Representation of Cultural Objects) 3D virtual museum system [33][34]. The presented list focuses only on properties specific for interactive 3D web content, and it is common for the new proposed Microformat and Microdata schemas. The schemas make use of the aforementioned Microformats and Microdata for media resources (`hMedia`, `hAudio`, `MediaObject`, `ImageObject`, `AudioObject`, and `VideoObject`) extending them with new metadata properties for 3D content. Attributes common for different media types have been omitted in the list, as they are inherited from the parent schemas. The inherited properties (not listed in Table 1) are mainly immutable and manually specified (e.g., `title`, `contributor`, `description`) or they are simple attributes, determined automatically without processing of their sub-components (e.g., `encodingFormat`, `uploadDate`). In Table 1, only the `presentedObject` must be specified manually, e.g., by a content creator. Other properties may be automatically determined—usually as aggregations of values from sub-components.

Microformats and Microdata introduce several equivalent metadata schemas, e.g., the `hAudio` and `hMedia` Microformats have as counterparts the `AudioObject` and the `MediaObject` in Microdata. Although these schemas usually contain common sets of properties, their numbers and types of attributes are frequently different, e.g., the `contributor` attribute indicates a `hcard` in Microformats and a `Person` or an `Organization` in Microdata. Hence, although the new proposed `Microformat3D` and `Microdata3D` schemas have the same properties specific for describing 3D content, they differ in the numbers and

types of properties that are inherited from their parent standard-specific schemas.

The proposed schemas do not impose the use of any particular units for individual metadata properties, but it is recommended to conform to the quantities and units specified in the described 3D content, e.g., metres and kilograms may be used for specifying the `dimensions` and the `mass` of the described object. Like any metadata describing a particular semantic property, unit descriptions may be nested into the descriptions of the relevant properties.

Although, overall, it is recommended to embed descriptions directly in the described 3D content, this is not required. The proposed `Microformat3D` and `Microdata3D` schemas are intended to be used directly within 3D content descriptions (in particular in X3D documents), but they may be also used in other types of documents, e.g., web pages embedding the 3D content. Hence, the name of the new Microformat does not start with a letter indicating a particular standard of the parent document—as opposed to the `hMedia` and `hAudio` Microformats that have been originally designed for HTML web pages.

IV. Searching 3D Content by Semantic Metadata Properties

This section explains the architecture of a planned system enabling 3D content retrieval using attribute-based built-in descriptions. The system leverages the proposed `Microformat3D` and `Microdata3D` schemas. First, the client–system interaction is discussed, second, an illustrative example is considered.

A. Client–system Interaction

The interaction between a client (a web browser) and the system is discussed in the two following aspects—loading 3D content into the system and querying the system for 3D content.

1) Loading 3D content into the system: Loading of 3D content into the system (Fig. 3a) is performed by the *Web Client*. Any browser can be used as the client, not imposing any specific and difficult to meet software requirements for the client side. A user sends 3D content with a built-in semantic description to the *3D Loader* web service via a web page with a file input component. The *3D Loader* sends the 3D content with its embedded metadata to the *GRDDL Agent* that extracts the metadata and creates a separate RDF document according to the rules presented in [31]. The document contains a semantic description equivalent to the built-in description. No changes are introduced to the primary X3D document. Finally, both the X3D document with built-in metadata and the generated RDF document are stored in the *Database*.

2) Querying the system for 3D content: Fig. 3b presents the consecutive steps performed every time 3D content is requested the system. First, a user utilizes a *Web Client* to build a query that specifies desirable metadata properties of the content. The query is embedded in an HTTP request and sent to the *Query Handler* web service. The *Proxy* is a Java application that mediates in the communication and extends the request with a context description of the client–system interaction that may specify, e.g., the client device and the software platform, user preferences and location, interaction paradigm, etc. Next, the *Query Handler* web service translates the extended query into a SPARQL [35] statement which is delivered to a *SPARQL Query Engine*. The engine retrieves desirable 3D components from the *Database*. The components are sent back to the *Query Handler* that invokes a *Web Page Builder* to create a representative web page. Finally, the web page is delivered to the *Web Client*.

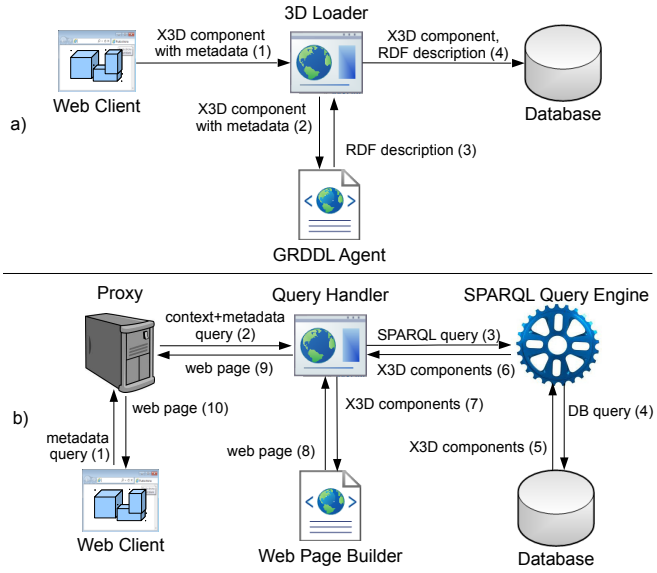


Fig. 3. Loading 3D content into the system (a) and querying the system for 3D content (b)

Listing 1. X3D content described with the proposed `Microformat3D`

```
<Shape DEF='Sculpture'>                                                          1
 <Appearance>...</Appearance><IndexedFaceSet>...</IndexedFaceSet>                2
 <MetadataSet name='http://www.kti.ue.poznan.pl/3DContent' value='Sculpture'>    3
  <MetadataString name='fn' value='Wooden sculpture' reference='property' />     4
  <MetadataString name='enclosure' value='.../sculpt.x3d' reference='relationship' />  5
  <MetadataString name='description' value='An example virtual sculpture' reference  6
      ='property' />
  <MetadataString name='presentedObject' value='http://.../museum/sculpture'     7
      reference='relationship' />
  <MetadataString name='dimensions' value='0.5 0.3 1' reference='property'>      8
   <MetadataString name='unit' value='meter' reference='property' />             9
  </MetadataString>                                                              10
  <MetadataString name='collisions' value='http://.../dish.x3d http://.../handle.x3d'  11
      reference='relationship' />
  <MetadataString name='illumination' value='point' reference='property' />      12
  <MetadataString name='interactivity' value='selection manipulation navigation'  13
      reference='property' />
  <MetadataString name='navigation' value='fly walk' reference='property' />     14
  <MetadataSet name='3DComponents' reference='property'>                         15
   <MetadataSet name='http://www.kti.ue.poznan.pl/StringTuple'>                  16
    <MetadataString name='URI' value='http://.../dish.x3d' reference='property' />  17
    <MetadataString name='role' value='dish' reference='property' />             18
  </MetadataSet></MetadataSet>                                                   19
 </MetadataSet>                                                                  20
</Shape>                                                                         21
```

B. Illustrative Example

Below, two examples of the client–system interaction are presented with regard to the steps described in the previous subsection. In the examples below, 3D content is described

with the proposed Microformat, but an equivalent description could be created as well using the Microdata schema.

1) Loading 3D content into the system: Example X3D content that presents a 3D model of a sculpture is presented in Listing 1. The X3D `head` and several other elements reflecting the geometry and appearance have been omitted as they are not crucial for this example. The method of embedding semantic descriptions into 3D content has been explained in detail in [29][30]. In the presented example, the metadata properties have been specified manually while creating 3D components. However, it is desirable to develop an additional tool automatically calculating and embedding the properties into 3D content.

The 3D content is described with the `Microformat3D` (line 3) with multiple properties inherited from the `hMedia` (`fn`-4, `enclosure`-5, `description`-6), logical (`presentedObject`-7), spatial (`dimensions`-8, `collisions`-11, `illumination`-12), temporal and behavioural (`interactivity`-13, `navigation`-14) as well as structural (`3DComponents` 15-19). Next, the 3D content is sent to the *3D Loader* web service and stored in the *Database*, which is implemented using Oracle XML DB [36].

Listing 2. Example conditions sent to the system by a user (a), and the SPARQL query including contextual requirements inserted by the *Proxy* (b)

```
a) ?component description ?description
     FILTER regex(?description, "artefact").
   ?component material ?material
     FILTER regex(?material, "wood").
   ?component illumination "point".
   ?component animations "position".

b) select ?URI where {
                ?component enclosure ?URI.
                ?component interaction "manipulation".
                ?component navigation "walk".
                ?component levelsOfDetail ?lod.
                ?lod numOfPolygons ?polygons.
                FILTER (?polygons >= 100 000).
                { select count(?lod) as ?n where
                            { ?component levelsOfDetail ?lod. }}.
                FILTER (?n >= 3). }
```

2) Querying the system for 3D content: Querying the system for 3D content starts with specifying conditions using the *Web Client* (e.g., a web page). The *Web Client* builds a SPARQL query, which is encoded with the SPARQL Protocol for RDF [37], built into HTTP address and sent to the *Query Handler* web service. In the presented example, a user specifies semantic properties of the desirable objects by requiring 3D models of artefacts made of wood. In addition, the following metadata properties of the 3D objects are specified—the artefacts should be illuminated by `point` light sources and their `positions` should be animated (Listing 2a). The *Proxy* inserts additional contextual requirements into the request that specify 3D components suitable for desktop devices equipped with a keyboard and a high-resolution screen—with the `manipulation` interaction, the `walk` navigation mode enabled, having at least 100k `polygons` and at least 3 `levelsOfDetail` (Listing 2b). The *Query Handler* conveys the extended query to the *SPARQL Query Engine* (implemented with Apache Jena [38]). Next, 3D components that satisfy the given conditions are retrieved from the *Database* and provided to the *Query Handler*. Finally, the *Web Page Builder* creates a web page (Fig. 4), inserting the found 3D components into a web page template. The resulting document is delivered to the *Web Client*.

Fig. 4. An example web page with the requested 3D components

V. Conclusions and Future Works

In this paper, Microformat3D and Microdata3D schemas have been proposed for describing interactive 3D web content. Although a number of schemas and ontologies have been designed for describing the metadata and semantics of multimedia content on the web, they have been intended mostly for images, audio and video and not for interactive 3D content. The lack of commonly accepted schemas for describing the metadata and semantics of 3D resources is one of the important obstacles to widespread creation, dissemination and reuse of interactive 3D web content.

The proposed Microformat3D and Microdata3D schemas allow for metadata and semantic descriptions of interactive 3D web components and 3D scenes. The proposed schemas make use of the available Microformats and Microdata schemas for flexible semantic descriptions, and they may be combined with X3D, which is the leading standard for describing 3D content on the web. The compatibility with well-established web approaches enables the use of the presented schemas for describing, retrieving and exploring (finding, classifying, clustering, cataloguing, etc.) interactive 3D content in a variety of multimedia web systems, using a number of available tools (editors, validators, parsers, transformers, etc.), with minimal additional effort. The presented approach may be used for query optimization in applications with context-based user–system interaction.

The proposed approach stresses the compatibility of the created descriptions with the current syntax and structure of X3D documents and available X3D browsers. To provide the conformance of the presented solution to popular web search engines, the proposed schemas need to be encoded using the original Microformat and Microdata syntax, which has been intended for web pages and which is not compatible with 3D content standards. Therefore, the metadata should be inserted into web pages embedding the described 3D content.

We plan implementation of the proposed approach as an extension to the ARCO virtual museum system. Such extension

will allow for creating built-in metadata and semantic descriptions of virtual museum exhibitions, and will enable evaluation of the schemas in terms of the achieved optimization of queries to digital repositories of 3D exhibits.

Possible directions of future research incorporate several facets. First, the paper describes only an initial set of properties that may be used for describing 3D content. Based on practical experiences from implementation of systems using this kind of descriptions, the presented schemas may be further extended to include more specific properties describing 3D content. Also, possible values of the properties together with the preferred syntax should be specified. Second, the presented metadata model may be implemented using powerful RDF and RDF-based technologies such as OWL and RDFS. This will permit sophisticated exploration of semantically described content, including querying data sources and reasoning. Third, a tool for automatic computation of the proposed metadata properties should be developed and used for new 3D components loaded into the system. Next, the contextual rules managed by the *Proxy* might be described with Semantic Web standards to be accessible and processable with widely-used semantic tools. Finally, an evaluation of the system should be performed to assess the benefits from the query optimization provided by the proposed metadata schemas.

VI. Acknowledgements

This research work has been partially funded by the Polish National Science Centre grant No. DEC-2012/07/B/ST6/01523.

References

[1] W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax. *http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/*. Retrieved April 25, 2013.

[2] W3C. RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004. *http://www.w3.org/TR/rdf-schema/*. Retrieved April 25, 2013.

[3] OWL. OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. *http://www.w3.org/TR/owl-ref/*. Retrieved April 25, 2013.

[4] ISO/IEC 15938-10:2005. Information technology - Multimedia content description interface – Part 10: Schema definition. *http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm ?csnumber=40527*. Retrieved April 25, 2013.

[5] W3C. Ontology for Media Resources 1.0. W3C Recommendation 09 February 2012. *http://www.w3.org/TR/mediaont-10/*. Retrieved April 25, 2013.

[6] A Core Ontology for Multimedia. *http://comm.semanticweb.org/*. Retrieved April 25, 2013.

[7] Microformats. *http://microformats.org/*. Retrieved April 25, 2013.

[8] W3C. HTML Microdata. Editor's Draft 26 April 2013. *http://www.w3.org/html/wg/drafts/microdata/master/*. Retrieved April 26, 2013.

[9] Web3D. Extensible 3D (X3D) Part 1: Architecture and base components ISO/IEC 19775-1:2008. *http://www.web3d.org/files/specifications/19775-1/V3.2/Part01/Architecture.html*. Retrieved April 25, 2013.

[10] The DIG35 Phase 2 Initiative Group. DIG35 Specification—Metadata for Digital Image. *http://www.bgbm.org/tdwg/acc/Documents/DIG35-v1.1WD-010416.pdf*. Retrieved April 25, 2013.

[11] CableLabs. Video-On-Demand Content Specification Version 2.0. *http://www.cablelabs.com/specifications/MD-SP-VOD-CONTENT2.0-I02-070105.pdf*. Retrieved April 25, 2013.

[12] Apple. QuickTime File Format Specification. *https://developer .apple.com/standards/qtff-2001.pdf*. Retrieved April 25, 2013.

[13] Bilasco I. M., Gensel J., Villanova-Oliver M., Martin H., On Indexing of 3D Scenes Using MPEG-7, In: *Proceedings of the 13th annual ACM international conference on Multimedia*, Singapore, November 06-12, 2005, pp. 471-474.

[14] Bilasco I. M., Gensel J., Villanova-Oliver M., Martin H., An MPEG-7 framework enhancing the reuse of 3D models , In: *Proceedings of the Web3D Symposium 2006*, Columbia, Maryland, USA, 2006, pp. 65-74.

[15] Arndt R., Troncy R., Staab S., Hardman L., Vacura M., COMM: designing a well-founded multimedia ontology for the web. In: *Proc. of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference*, Busan, Korea, November 11-15, 2007, pp. 30-43.

[16] Chmielewski J., Describing Interactivity of 3D Content. In: *Interactive 3D Multimedia Content—Models for Creation, Management, Search and Presentation*, ed. Cellary, W., Walczak K., Springer, London, Dordrecht, Heidelberg, New York, 2012, pp. 195-221.

[17] Chmielewski J., Metadata Model for Interaction of 3D Object, In: *The 1st International IEEE Conference on Information Technology*, ed. Stepnowski, A., M. Moszyński, T. Kochański, J. Dabrowski, Gdańsk, May 18, 2008, Gdańsk University of Technology, 2008, pp. 313-316.

[18] Chmielewski J., Finding interactive 3D objects by their interaction properties, In: *Multimedia Tools and Applications*, Springer, Netherlands, 2012, ISSN: 1380-7501.

[19] Boeykens S., Bogani E., Metadata for 3D Models. How to search in 3D Model repositories? In: *Proceedings of the International Conference of Education, Research and Innovation*, Madrid, Spain, November 17-19, 2008.

[20] Metadata 3D Initiative. *http://www.m3di.org/*. Retrieved April 25, 2013.

[21] Ghosh H., Chaudhury S., Ontology for semantic multimedia web. *http://www.w3cindia.in/conf-site/Hiranmay%20Ghosh%20–%20mowl%20%28ontology %20for%20semantic%20multimedia%20web.pdf*. Retrieved April 25, 2013.

[22] Schema.org. *http://schema.org/*. Retrieved April 25, 2013.

[23] hMedia. *http://microformats.org/wiki/hmedia*. Retrieved May 20, 2013.

[24] hAudio. *http://microformats.org/wiki/hAudio*. Retrieved May 20, 2013.

[25] MediaObject. *http://schema.org/MediaObject*. Retrieved May 20, 2013.

[26] ImageObject. *http://schema.org/ImageObject*. Retrieved May 20, 2013.

[27] AudioObject . *http://schema.org/AudioObject*. Retrieved May 20, 2013.

[28] VideoObject. *http://schema.org/VideoObject*. Retrieved May 20, 2013.

[29] Flotyński J., Walczak K., Describing Semantics of 3D Web Content with RDFa. In: *Proceedings of the First International Conference on Building and Exploring Web Based Environments*, Seville, Spain, January 27-February 1, 2013, pp. 63-68, ISBN 978-1-61208-248-6.

[30] Flotyński J., Walczak K., Attribute-based semantic descriptions of interactive 3D web content. In: *Information Technologies in Organizations : Management and Applications of Multimedia - Toruń*, Wydawnictwa Towarzystwa Naukowego Organizacji i Kierownictwa—Dom Organizatora, Czestochowa, Poland, 2013, pp. 111-138, ISBN 978-83-7285-691-3.

[31] Flotyński J., Harvesting of Semantic Metadata from Distributed 3D Web Content. In: *Proc. of the 6th Int. Conf. on Human System Interaction*, Sopot, Poland, June 6-8, 2013, ISBN 978-1-4673-5636-7.

[32] W3C. GRDDL Primer. W3C Working Group Note 28 June 2007. *http://www.w3.org/TR/grddl-primer/*. Retrieved April 25, 2013.

[33] Mourkoussis N., White M., Patel M., Chmielewski J., Walczak K., AMS Metadata for Cultural Exhibitions using Virtual Reality, In: *Dublin Core International Conference DC 2003*, Seattle, Washington, USA, September 28 - October 2, 2003, pp. 193-201.

[34] Patel M., White M., Mourkoussis N., Walczak K., Chmielewski J., Wojciechowski R., Metadata Requirements for Digital Museum Environments, In: *Int. Journal on Digital Libraries, Special Issue on Digital Museum*, vol 5, no 3, May, 2005, Springer Verlag, pp.179-192.

[35] W3C. SPARQL Query Language for RDF. W3C Recommendation 15 January 2008. *http://www.w3.org/TR/rdf-sparql-query/*. Retrieved April 25, 2013.

[36] Oracle XML DB. *http://www.oracle.com/technetwork/database-features/xmldb/overview/index.html*. Retrieved April 25, 2013.

[37] W3C. SPARQL 1.1 Protocol. W3C Recommendation 21 March 2013. *http://www.w3.org/TR/sparql11-protocol/*. Retrieved April 25, 2013.

[38] Apache Software Foundation. Apache Jena. *http://jena.apache.org/*. Retrieved April 25, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 557–564

Exploring inexperienced user performance of a mobile tablet application through usability testing.

Chrysoula Gatsou
School of Applied Arts
Hellenic Open University,
Patra , Greece
Email: cgatsou@teiath.gr

Anastasios Politis
Graphic Arts Technology
Faculty of Fine Arts and Design,
TEI of Athens
Athens, Greece
Email: politisresearch@techlink.gr

Dimitrios Zevgolis
School of Applied Arts
Hellenic Open University,
Patra, Greece
Email: zevgolis@eap.gr

Abstract—This paper explores inexperienced user performance through a usability testing of three alternative prototypes of a mobile tablet application. One key factor in inexperienced users adopting mobile technology is the ease of use of mobile devices. The interface layout one of the three prototypes was built on the basis of previous research conducted in collaboration with users. More specifically, our study involves five navigation tasks which novice users were required to complete with each of the three prototypes. Our results showed that participants displayed better task performance with the prototype F1, which was created in collaboration with participants, in contrast to prototypes F2 and F3, which both caused navigation problems.

I. Introduction

THE rapid growth of mobile tablet technologies has lead to exponential growth in numbers of novice users, that is, in ordinary people who lack skills in computer science and who are drawn from a wide range of backgrounds. According to Hassenzahl [1], there is no guarantee that users will actually perceive and appreciate the product in the way designers desire it to be perceived and appreciated. For example, a product with a specific screen layout intended to be clear and simple will not necessarily be perceived as such. Despite the best efforts on the part of designers, new technologies often fail to meet basic human needs and desires [2]. The difficulties concerned in designing an interface that will deal effectively with individual preferences and experience, while minimizing frustration on the part of the user, transfer errors and learning effort, is widely recognized as a persistent problem of Human Computer Interaction [3]. Making things more usable and accessible is part of the larger discipline of user-centered design (UCD), which includes a number of methods and techniques [4]. Usability testing is a method used to evaluate a product by testing it on representative users. Greenberg and Buxton point out that *"Usability evaluation is valuable for many situations, as it often helps validate both research ideas and products at varying stages in its lifecycle"* [5].

Prototyping is an essential part of usability testing, as it confirms whether users can effectively complete tasks by means of the prototypes that are being tested and allows us to deal with various types of problems. Furthermore, prototypes can also be useful in dealing with the more subjective aspects of an interface. A previous study by the present authors has shown that inexperienced users structure content information in a mobile tablet application differently from experienced users, when the former interact with mobile devices [6]. Carroll argues that an effective way of dealing with system complexity for the novice user is to provide a functionally simple system [7]. In order to create more affordable mobile interactive artifacts for inexperienced users, we have focused on the interface design of a mobile tablet application and tested it on real users. The goal of this study is to investigate the effect of different interfaces in usability testing with regard to inexperienced user performance and the perceived usability of a tablet mobile application. To present the results of our study, we start our paper with a review of the literature, which establishes the theoretical background for our study. We then describe the research methodology employed. We analyse the data and give our results, which we discuss, before offering some conclusions.

II. Background

A. Prototyping

Prototyping is an essential procedure that structures design innovation. The translation of user needs into a system specification was facilitated in our study by iteratively refined prototypes validated by users. Beaudouin-Lafon and Mackay define any prototype as a concrete representation of part or all of an interactive system [8]. A prototype is, in their view, a tangible artifact, rather than an abstract description that requires interpretation. In Moggridge's, view a prototype is a representation of a design made before the final solution exists [9]. By offering different prototypes of a mobile application to users and requesting feedback, we can be sure that we are designing for those who will actually use our designs. Prototyping serves various purposes in a human-centered design process. With a view to improving this process, we developed, as a continuation of two previous studies of card sorting and creative session , three interactive prototypes in order to explore more intuitive navigation methods for inexperienced users who are to interact with

978-83-60810-53-8/$25.00

mobile tablet devices. Our decision to construct three, rather than one prototype, rests on work by Tohidi et al. and Dow et al., who argue that multiple prototypes are more helpful than merely one in aiding users to formulate negative or positive comments [10],[11]. Moreover, we were able to verify whether or not the prototype that we created on the basis of user requirements has responded effectively to these. Houde & Hill classify the ways in which the prototypes can be of value to designers [12]. Prototypes, in their view, include any representational design idea, regardless of the medium involved. Their model defines three types of issues that a prototype may affect, namely, the role of a product in the context in which it is used, the look and feel of the product and its technical implementation. Floyd, in an earlier laboratory prototyping of complex software systems, describes two primary objectives of prototypes, namely, 1) to act as a vehicle for learning and 2) to enhance communication between designers and users, as developer introspection of user needs often leads to inadequate products [13]. Buchenou & Suri note that the prototypes perform an additional function. They strengthen empathy, *"an original experience every kind of representation, in any medium, that is designed to understand, to explore or communicate what you could work with the product, space or system design"*[14]. Lim et al., stress the role of prototypes as a vehicle for learning, "*prototypes are the means by which designers organically and evolutionary learn, discover, generate, and refine designs*" [15].

B. Usability testing

The term *"usability"* is frequently employed in the field of human-computer interaction (HCI). Nielsen describes usability as an issue related to the broader issue of acceptability [16]. In his view, *"Usability is a quality attribute that assesses how easy user interfaces are to use"*. Usability is a significant part of the user experience and therefore of user satisfaction. A formal definition of usability is given in the ISO standard 9241–11 : *"...the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction, in a specified context of use"*. Effectiveness is defined as the accuracy and completeness with which users achieve specified goals and efficiency as the resources expended in relation to the accuracy and completeness with which users achieve goals. Satisfaction is defined as the freedom from discomfort, and positive attitude to the use of the product, whilst the context of use is defined as users, tasks, equipment and the physical and social environments in which a product is used [17].

Usability testing is a method employed in user-centered design to evaluate product design by testing it on representative users. Such users thus yield quantitative and qualitative data in that they are real users performing real tasks. Usability testing requires an artifact that is fairly complete and rationally designed, which means that the appropriate place for usability testing is at a stage quite late in the design cycle [18].

Dumas & Redish argue that usability testing is a *"a systematic way of observing actual users trying out a product and collecting information about the specific ways in which the product is easy or difficult for them"* [19]. They also recommend that usability test possess the following five features:

1. The primary goal is to improve the usability of a product. For each test, you also have more specific goals and concerns that you articulate when planning the test.
2. The participants represent real users.
3. The participants do real tasks.
4. You observe and record what participants do and say.
5. You analyze the data, diagnose the real problems, and recommend changes to fix those problems.

III. Research Methodology

To examine how novice users conceptualize a mobile tablet application, we created a user test involving three prototypes of a mobile tablet application themed around the topic of 'first aid' (Fig.1).

Fig. 1 Participant during the usability testing.

All three interfaces had the same look and feel, in order to standardize the visual appeal and the emotional impact made by the various alternative versions employed in the test. These versions vary in terms of conceptual models and menu navigation, one of them F1 having been created on the basis of the participant collaboration in previous studies by the present authors [6], [20].

A. Participants

The literature gives no clear optimum number of participants to be employed in usability testing.

Nielsen [21] argues that five participants will discover 80% of the problems in a system. In any case, a small amount of users, that is, generally fewer than 10 subjects, is sufficient for any formative evaluation of usability [22]. On the other hand, Spool and Schroeder [23] state that five

users identified only about 35% of the problems in a website. The research by Turner et al. implies that a group size of seven may be optimal, even when the study is fairly complex [24].

According to Sauro & Lewis *"the most important thing in user research, whether the data are qualitative or quantitative, is that the sample of users you measure represents the population about which you intend to make statements"* [25]. Our session was designed specifically to include a pool representative of potential users of the mobile application being tested. Twelve participants (N=12) ranged from 18 to 79 (mean age = 41,6, SD = 20.9, years), seven of whom were men and five women, all of whom had participated in one or more previous studies. All participants were novices in terms of computing. They had no visual or cognitive impairment and their education was of at least high school level. Given the evidence from our previous studies, the number of people in this experiment was sufficient to provide satisfactory evidence and depth for us to study. The age and gender of the participants is shown in Table I.

TABLE I.
AGE, GENDER AND NUMBER OF PARTICIPANTS.

ID	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
Age	50	38	26	27	57	79
Gender	M	F	M	F	M	M
ID	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**
Age	52	31	18	45	65	67
gender	M	M	F	F	F	M

B. Material

Usability testing was performed on a Dell Inspiron Duo, 10.1tablet computer with a touch screen. A Panasonic HDC-SD40 digital camera was used to create a complete record of all user interactions with the interface. Furthermore, Camtasia Studio software was used to record a video of the movements made by the user on the interface during the test. Camtasia studio software captures the action and the sound from any part of the desktop. Digital tape recorder was also used.

C. The three prototypes

To reproduce a realistic software environment, for a period of three months, three prototypes were developed in Adobe Flash and we used them as a tool for recording user behavior during interaction. Prototypes help designers to balance and resolve problems that occur in different dimensions of design. Each prototype allowed the user to interact with mobile application and to carry out some tasks.

Interface F1

The first screen of the interface consists of icons that offer easy accessibility to the topic. We settled on this layout after a participatory session with users involved in our previous study [6]. There we concluded that users preferred icons for main menu selection, rather than a representation of options in words arranged hierarchically.

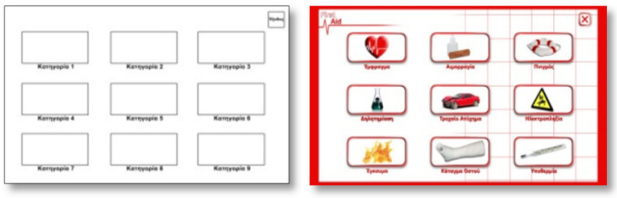


Fig. 2 The Interface F1

Interface F2

The colors remain the same in prototype F2, but the main menu has been moved to the left of the screen and now employs words, instead of icons. The options are the same in number as in the prototype F1. The subcategories are now placed in the middle of the screen. The aim of this layout was to explore whether a larger amount of text helps or hinders the inexperienced user to interact with a mobile application.

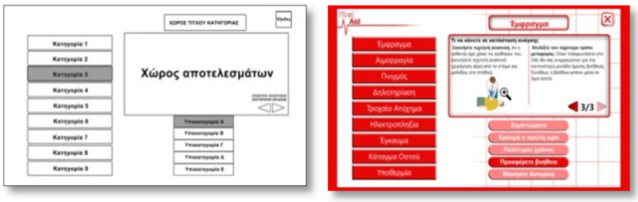


Fig. 3 The Interface F2

Interface F3

Prototype F3 is identical in basic design to prototype F2, except for a horizontal bar at the top of the screen, which enables the user to select subcategories. This layout resembles that of a website. The aim of this arrangement, which simulates the web environment, was to test the familiarity of users with little experience of surfing.

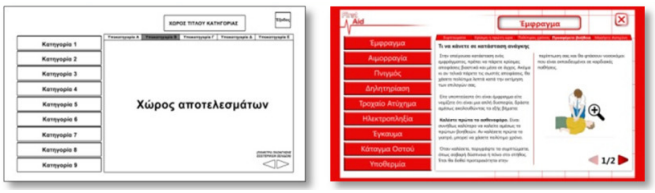


Fig. 4 The Interface F3

D. User Tasks

For the usability test, the participants were required to complete the five tasks given in Table II. The tasks were chosen as being representative and covered as many as possible of the features of the application.

TABLE II
PARTICIPANTS TASKS

Task 1	Turn on the mobile tablet device and select the icon "first aid".
Task 2	Find the information on Cardiopulmonary resuscitation (CPR.)
Task 3	Enlarge the image in order to see details.
Task 4	Select information on heart attacks.
Task 5	Find the information on symptoms of broken bones. Turn off the mobile device

E. User Performance

User performance was recorded in terms of the effectiveness, efficiency and ease of use of prototypes. In order to evaluate task effectiveness, we measured the percentage of tasks successfully completed within the set time limit. Task completion time refers to the time needed to accomplish the task. To evaluate efficiency, we recorded the time needed to process a task. To measure user satisfaction, we asked users to complete a post-test questionnaire.

F. Post -test Questionnaire

The main aim of administering written questionnaire after the test (post-test questionnaire) is to record participants' preference, in order to identify potential problems with the product. Information collected usually includes opinions and feelings regarding any difficulties encountered in using the product. Our questionnaire was based on System Usability Scale (SUS) developed by Brooke [26], since this is the most precise type of questionnaire for a small number of participants, as is shown by Tullis and Stetson's study [27]. SUS employs a "quick and dirty" approach in evaluating the overall subjective usability of a system (Appendix A). While SUS was originally intended to be used for measuring perceived usability, i.e. measuring a single dimension, recent research shows that this provides an overall measure of satisfaction of the system [27],[28],[29]. In addition to these advantages over other systems, the SUS is a powerful and multifunctional instrument [30].

G. Test protocol

Participation in the study lasted approximately one hour and 20 minutes and was conducted in an isolated room in our department. It consisted of the series of tasks that we mention above. All participants were tested individually.

After being welcomed by the experimenter, participants were told that they were to take part in a usability test and were to work with a prototype of a mobile tablet application. All participants gave their permission to be recorded on video. Subsequently participants completed the five tasks. The process of user testing is illustrated in Fig. 5. To minimize the potential for learning bias, the presentation order of the prototypes was counterbalanced.

IV. Results And Discussion

The main factors to be examined when testing usability are effectiveness, efficiency and user satisfaction. Effectiveness refers to how "well" a system does what it supposed to do. In order to evaluate task effectiveness, we measured the percentage of steps successfully solved within the time limit (7min). Efficiency refers to how quickly a system supports the user in what he wants to do. To evaluate efficiency, we recorded the time needed to process the task. Satisfaction refers to the subjective view of the system on the part of the user [4]. Qualitative and quantitative data were collected from each participant. Qualitative data included the participants' verbal protocol as recorded in videotapes.

Problems of usability were identified and categorized. We also collected comments on the prototypes and preference

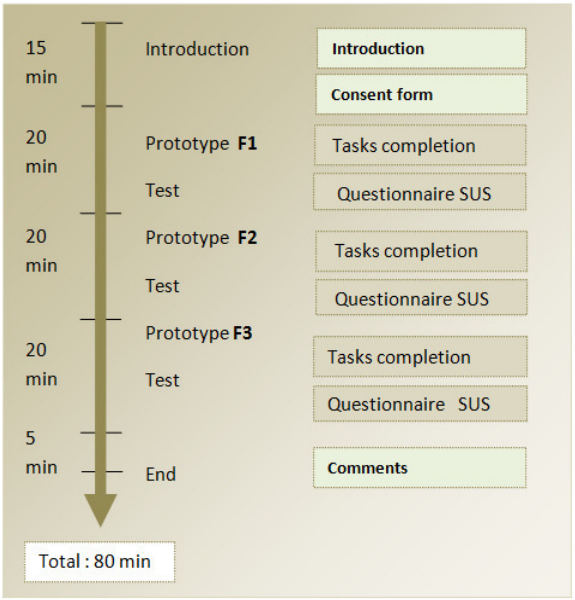


Fig. 5 User test process.

data and evaluations in the form of the SUS data questionnaire completed by the users after the test. Any user action that did not lead to the successful completion of a task we defined as error.

A. Effectiveness.

The percentage of users that manage to complete a task successfully thus becomes a measure of the effectiveness of the design. The number of errors made on the way to completing a task is an example of a performance measure [4]. An interaction effect is noticeable in the results, suggesting that the approach employed in the interface F1 may well have a marked impact on reducing the number of errors made. Tables III, IV show the user tasks and the error rate.

TABLE III
TASKS COMPLETION RATES

	Prototype F1	Prototype F2	Prototype F3
Task1	12/12	11/12	12/12
	100%	**91%**	**100%**
Task2	11/12	9/12	6/12
	91%	**75%**	**50%**
Task3	9/12	8/12	7/12
	75%	**66%**	**58%**
Task4	9/12	7/12	4/12
	75%	**58%**	**33%**
Task5	11/12	7/12	6/12
	91%	**58%**	**33%**

Errors were classified into two main categories, navigation errors and comprehension errors. Navigation errors occurred when partcipants didn't move as expected. Comprehension errors occured when participants didn't understand the design of the interface.

B. .Efficiency - Task Completion Time

We recorded the total amount of time required to complete each task in prototypes F1,F2 and F3, starting from turning

TABLE IV
TYPES OF ERRORS BY PROTOTYPE

Type of error	Prototype F1	Prototype F1	Prototype F1
Navigation	3	7	12
Comprehension	3	6	6
Total	6	13	18

the device on to turning it off. The mean amount of time required by participants in each age group is shown in Fig.7. Participants P6, P11, P12 failed to complete their tasks in prototype F2 within the time set (7min). In prototype F3, participants P2, P6, P11, P10, P12 failed to complete their tasks . Table V shows the results of the mean completion time and standard deviation for the participants for prototypes F1,F2 and F3. Data regarding time taken by each participant for each task is given in Appendix B.

Participant	Age	F1	F2	F3
P1	50	0:02:41	0:03:52	0:05:43
P2	38	0:02:49	0:03:35	
P3	26	0:02:10	0:04:11	0:04:26
P4	27	0:02:53	0:03:28	0:04:42
P5	57	0:02:15	0:02:40	0:05:07
P6	79	0:05:47		
P7	52	0:02:55	0:04:22	0:06:06
P8	31	0:02:45	0:03:33	0:04:30
P9	18	0:02:10	0:03:02	0:04:11
P10	45	0:02:23	0:04:53	
P11	65	0:03:58		
P12	67	0:04:26		

Fig.6 The tasks completion mean time (seconds).

For users testing the prototype F1, the time needed to complete tasks ranged between 2:10 min and 2:53 min up to the age of 57. For participants aged 57 years or older, task completion time increased. This affected mean task completion time and standard deviation. For prototype F2 tasks, completion times were clearly higher. Participants older than 57 failed to complete their tasks within the specified time. The mean completion time of those who did finish their tasks was 20.43% greater than the corresponding figure in prototype F1. For prototype F3, the mean completion time for those who succeeded in finishing was 33.04% greater than the corresponding figure in prototype F1.

Elderly users were thus not able to complete all the tasks in prototype F2 and F3 and specifically in prototype F3, where the layout of the prototype was slightly different, in that it resembled a web site. They had more information to process located on the left and at the top of the screen. These users found the interaction difficult to understand and to activate. On the whole, all users were more comfortable when interacting with prototype F1.

TABLE V
TASKS COMPLETION TIME (MEAN, SD)

	Prototype F1	Prototype F2	Prototype F3
Task completion time (mean)	03:06	03:44	04:58
Standard Deviation (SD)	01:06	00:41	00:43

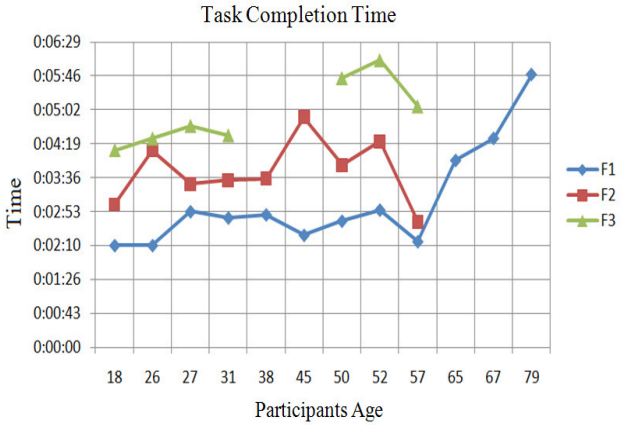


Fig.7 Task completion time per interface type.

C. Post test Questionnaire

We realised that time-on-task measures can be useful for collecting data on the efficiency of a system. On the other hand, such data does not give any information on overall satisfaction on the part of the user. User satisfaction may be an important factor in motivating people to use a product and may affect user performance. So, as a final point we decided participants were to complete an SUS questionnaire, so as to explore their experiences when interacting with the prototypes. A crucial feature of the SUS lies in the fact that asks the user to evaluate the system as a whole, rather than specific aspects.

All 10 questionnaire statements having been processed, the overall SUS score for each prototype turned out to be that given in Table VI. To calculate the SUS score, first we summed the score contributions of the items 1, 3, 5, 7 and 9 (Appendix A). The score contribution of these items are their scale position minus one. We then summed the score contributions of the other items: five minus their scale position. Finally, we multiplied the sum of the scores by 2.5, to obtain the overall score with a range between 0 to 100.

The survey results showed the overall satisfaction. Sauro [31] reports that a mean value over 74 is level B, value above 80.3 is level A. An average value of below 51 is level F (fail). The prototype F1 with an average value of 80.6 passes the threshold of 80.3 and are to be placed on level A, F2, with an average value of 63.3, belong to level B and F3, with a value of 48.1, is to be placed at Level F, which is regarded as failure.

However, with respect to F1, nearly all participants preferred the interface with the icons over the other two interfaces (F2, F3), in which there was a large amount of text. Some of the participants simply misunderstood the graphics keys that depicted a lens and whose purpose was to increase the photographs on the screen and the arrows that represented the act of selecting the next screen. If perhaps users had understood the graphics more fully, the error rate for prototype F1 may perhaps have been as low as zero.

TABLE VI
OVERALL SUS SCORE

Participants	F1	F2	F3
P1	80.0	70.0	70.0
P2	82.5	70.0	25.0
P3	90.0	82.5	60.0
P4	95.0	82.5	72.5
P5	87.5	80.0	72.5
P6	65.0	25.0	25.0
P7	77.5	70.0	27.5
P8	75.0	75.0	75.0
P9	92.5	82.5	75.0
P10	82.5	72.5	25.0
P11	70.0	25.0	25.0
P12	70.0	25.0	25.0
Mean	**80.6**	**63.3**	**48.1**

Overall users liked the process and regarded their interaction with the prototypes positively. Nevertheless, in some cases, the participants were apprehensive. Uncertain in their selections, they demanded greater confirmation and reassurance about the actions they were to take. In such cases, it is important for the researcher to motivate participants, encouraging them discreetly to investigate alternative directions, while simultaneously recording any mistakes made. As for individual prototypes, participants prefered the design of the first interface, which contained icons (F1). This was to be expected and users commented positively on its simplicity, ease of use and intuitiveness.

V. CONCLUSION

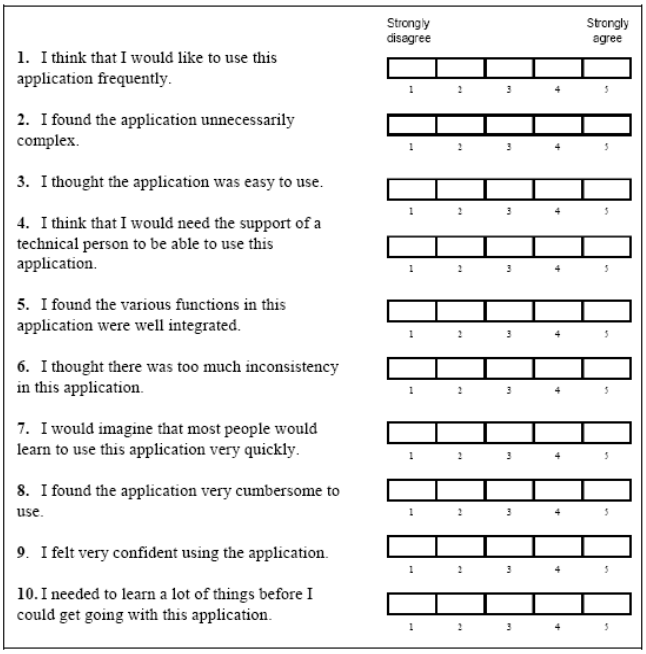
The aim of our study was to examine whether an interface design approach could improve performance of tasks by inexperienced users during interaction. To do this, we employed three different prototypes of the same application. We tested our empirical methodology on twelve individuals, all of them novices in terms of computer use.

One of the most remarkable discoveries we made is the large degree of difference in performance among the three different prototypes with regard to user effectiveness and the number of errors. The effectiveness and efficiency of the F1 prototype is evident in the fact that users made fewer errors and took less time to complete their tasks. Participants reported that the icon menu of the F1 prototype facilitated the execution of their tasks, as did the absence of text in menu selections. This confirms what emerged from a previous study by the present authors. Our findings imply that the users did not understand the basic conceptual models informing prototypes F2 and F3 [2].

The usability test performed on each of the prototypes showed that most users considered the prototype easy to use and intuitive. When evaluated by SUS, the same prototype received an overall score which placed it on level A. The test also helped in locating various issues regarding the other two prototypes F2 and F3 and, in particular, regarding what is to be corrected, so as to improve its usability for the elderly. However, we believe that our paper, which focuses more on the users and their cognitive abilities, offers a new insight into how inexperienced users perform tasks on mobile tablets.

APPENDIX

Appendix **A** *System Usability Scale*



ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. We would also like to thank the participants in our study.

REFERENCES

[1] M. Hassenzahl, The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, 13(4), 2002, pp. 479-497

[2] D. Norman, *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution.* MIT Press, Cambridge, MA, USA, 1999.

[3] D. Benyon and D. Murray: Applying user modeling to human-computer interaction design. *Artificial Intelligence Review,* 7(3-4): pp.199-225, 1993.

[4] J.Rubin, and D. Chisnell, *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests* (2nd Ed.). Indianapolis, IN: Wiley Publishing, 2008.

[5] S. Greenberg and B. Buxton, Usability evaluation considered harmful (some of the time). Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems, Florence, 2008, 111-120.

[6] C. Gatsou, A, Politis and D. Zevgolis, Novice User involvement in information architecture for a mobile tablet application through card

sorting. In proceedings *FEDSIS-MMAP* Wroclaw, 2012, pp. 711–718.

[7] J.M.Carroll, *The Nurnberg Funnel,* Cambridge, Mass.: MIT Press), 1999.

[8] M. Beaudouin-Lafon, and W.Mackay, "Prototyping Tools And Techniques". In: J. A. Jacko and A. Sears (Eds) *The Human-Computer Interaction Handbook.* Lawrence Erlbaum Associates, 2003.

[9] B. Moggridge, *Designing Interactions*. The MIT Press, 2007.

[10] Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. (2006). Getting the right design and the design right. In proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '06. ACM, New York, NY, pp. 1243-1252.

[11] S.P. Dow, S. P, A.Glassco, J. Kass, M. Schwarz, and S.R Klemmer,. The effect of parallel prototyping on design performance, learning, and self-efficacy. Tech. Rep. CSTR-2009-02, Stanford University.

[12] S.Houde, and C. Hill, "What do prototypes prototype?," What do prototypes prototype? In M. Helander, T. Landauer, & P. Prabhu (Eds.), Handbook of human-computer interaction ,2nd ed., Amsterdam: Elsevier Science, 1997, pp. 367-381.

[13] C.A.Floyd, Systematic Look at Prototyping. In *Budde, ed., Approaches to Prototyping.* Springer Verlag, pp. 105-122, 1984.

[14] M. Buchenau, & J. Fulton Suri, "Experience prototyping". In Proceedings *of Design of Interactive Systems* New York: ACM Press, 2000, pp. 424-433.

[15] Y.Lim, E. Stolterman and J.Tenenberg, "The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas". *ACM Trans. Comput.-Hum. Interact.* 15, 2 , pp1-27, 2008.

[16] J. Nielsen, Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. In *R. G. Bias & D. J. Mayher (Eds.), Cost-Justifying Usability Boston*, MA: Academic Press.1994 pp. 242-272.

[17] ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11, Guidance on usability, ISO.

[18] A.Cooper, R. Reimann, and D. Cronin, *About Face 3: The Essentials of User Interface Design.* John Wiley & Sons, Inc. 2007.

[19] J.S Dumas and J.C Redish, *A Practical Guide to Usability Testing* (revised Ed.). Portland, Oregon: Intellect Books,1999.

[20] C. Gatsou, A. Politis and D, Zevgolis, "Text vs visual metaphor in mobile interfaces for novice user interaction" *In Proc. of 16th International Conference on Electronic Publishing,* 2012 pp.271-279

[21] J.Nielsen, "Why You Only Need to Test With 5 Users". Jakob Nielsen's Alertbox, March 19, 2000.

[22] H. Petrie and N. Bevan. The evaluation of accessibility , usability and user experience *In: The Universal Access Handbook, C Stepanidis (ed), CRC Press*, 2009, pages 299–315.

[23] J. Spool and W. Schroeder " Testing web sites: Five users is nowhere near enough", In: Proceedings of the Conference extended abstracts on Human Factors in Computing Systems, CHI'2001. New York: ACM Press; 2001.

[24] C. W. Turner, J. R. Lewis, and J. Nielsen, "Determining usability test sample size". In *W. Karwowski (ed.), International Encyclopedia of Ergonomics and Human Factors* Boca Raton, FL: CRC Press, 2006, pp. 3084-3088.

[25] J. Sauro and J.R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Burlington, MA: Morgan Kaufmann, 2012.

[26] J. Brooke, SUS: a "quick and dirty" usability scale. In *P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), Usability Evaluation in Industry* (S. 189 -194). London: Taylor and Francis,1996.

[27] T. Tullis, and J. Stetson, "A comparison of questionnaires for assessing website usability," In *Proc. of the Usability Professionals Association (UPA)* 2004, pp. 7–11.

[28] J. R. Lewis, and J. Sauro, "The factor structure of the system usability scale", *Proc. Human Computer Interaction International Conference (HCII 2009),* San Diego, CA, 2009, pp. 94–103.

[29] J. Sauro, "Does prior experience affect perceptions of usability?" Available:http://www.measuringusability.com/blog/prior-exposure .php, January 19, 2011 [Nov.15, 2012]

[30] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human-Computer Interaction,* vol. 24, issue 6, 2008, pp. 574–594.

[31] J. Sauro, *A practical guide to the System Usability Scale (SUS): Background, benchmarks & best practices*. Denver, CO: Measuring Usability LLC.2011

Appendix **B** Data regarding time taken by each participant for each task.

"F1" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	20	25	19	18	22	33	24	25	17	18	24	27	23	5
E2	25	31	37	42	29	110	55	35	29	20	49	35	41	24
E3	60	56	28	41	35	122	38	52	39	25	55	60	51	25
E4	35	30	24	43	25	46	21	29	21	38	71	111	41	26
E5	21	27	22	29	24	36	37	24	24	42	39	33	30	7
SUM (sec)	161	169	130	173	135	347	175	165	130	143	238	266	186	
SUM (min:sec)	0:02:41	0:02:49	0:02:10	0:02:53	0:02:15	0:05:47	0:02:55	0:02:45	0:02:10	0:02:23	0:03:58	0:04:26	0:03:06	

"F2" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	33	34	38	32	27		30	33	25	42			33	5
E2	40	39	60	38	34		53	40	37	50			43	9
E3	73	67	54	65	40		59	65	56	67			61	10
E4	47	40	53	39	30		77	39	34	86			49	19
E5	39	35	46	34	29		43	36	30	48			38	7
SUM (sec)	232	215	251	208	160	0	262	213	182	293	0	0	224	
SUM (min:sec)	0:03:52	0:03:35	0:04:11	0:03:28	0:02:40	0:00:00	0:04:22	0:03:33	0:03:02	0:04:53	0:00:00	0:00:00	0:03:44	

"F3" Task	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	mean time	st.dev.
E1	52		37	41	43		54	39	34				43	8
E2	60		45	52	57		73	46	42				54	11
E3	92		60	62	65		89	63	67				71	13
E4	81		81	79	85		92	80	55				79	11
E5	58		43	48	57		58	42	53				51	7
SUM (sec)	343	0	266	282	307	0	366	270	251	0	0	0	298	
SUM (min:sec)	0:05:43	0:00:00	0:04:26	0:04:42	0:05:07	0:00:00	0:06:06	0:04:30	0:04:11	0:00:00	0:00:00	0:00:00	0:04:57	

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 565–569

Universal approach for sequential audio pattern search

Róbert Gubka, Michal Kuba, Roman Jarina
Faculty of Electrical Engineering, University of Žilina
Univerzitná 1, 010 01 Žilina, Slovak Republic
Email: robert.gubka@fel.uniza.sk

***Abstract*—This article deals with universal sequential audio pattern search and sound recognition method. Inspired by classical phoneme-based speech recognition and word spotting systems, where longer speech patterns are formed by sequences of basic speech units, we propose a methodology for creating a finite database of elementary sound models. These models can be arbitrary concatenated into sequences, thus forming a model of the required acoustical pattern or sound event.**

I. Introduction

AUTOMATIC speech recognition and word spotting systems are nowadays getting to the forefront in daily use. Intelligent human-computer communication interfaces allow us to take up the control over electronic equipment using our voice, making their usage more native and comfortable. All this is possible thanks to rigorous research in the field of human speech production and recognition. However, voice operated devices can only work with human speech and react only to specific spoken keywords or phrases in certain language, mostly English.

Universal intelligent system should be versatile, easily expandable for new commands in different languages, have ability to learn and operate also with non-speech sounds and acoustical events, e.g. for better evaluation of content and context of a situation. Such system can be adopted in different application areas. Audio is useful especially in situations when other sensors fails to reliably detect an event. For example in the context of surveillance systems, in weakly illuminated places, public transport areas, public halls or streets where it is not possible to evaluate video, or visual information alone is unreliable, the audio events spotting system can be very useful as burglar or violence alarm by detecting sound events, like glass brake, shout, footsteps, or any other significant sound defined by user [1], [2].

Another topic in the field of audio processing is information retrieval over multimedia content. Currently, huge amount of multimedia data such as music recordings, broadcast news, dialogs and conversations, etc., are available in large audio databases. However, most of these data are unstructured and have limited or no tag information about the content, hence, it is not easy for user to locate desired audio samples or segments.

Most of the existing applications perform the content-based analysis by segmentation of the audio data and subsequently classification of audio sequences into one of the specified sound classes, represented mostly by a statistical model [3]–[5]. Disadvantage of this approach is, that a sufficient amount of representative data is needed for each sound class to be created and user has only limited possibility to adding new queries into search. Moreover, with growing number of sound classes or queries in search also the computation and memory demands grow.

”Query-By-Example” (QBE) paradigm is an alternative approach to multimedia information retrieval. In the context of audio information, the user provides a short sound clip as a query and the system returns audio samples that are similar to the query. For example, the user provides a short utterance spoken by a particular person and expects that the system returns all the samples from the audio (video) database that contain the voice of the same person. Or the user gives a sample of an applause sound and the system should return all clips from the audio/video content that contain applause.QBE approach is also very popular in music information retrieval [6], [7]. The challenge of the QBE approach is that only very limited amount of training/reference data are available in advance and sound classes are apriori not known, thus conventional statistical model-based approaches and learning algorithms cannot be straightforward adopted.

Among various approaches to example-based audio event detection and retrieval, the most popular ones are based on similarity measure in audio feature vector space [8], [9] or hidden Markov models (HMM) [10], [11]. The approach in [11] was based on feature-based segmentation of audio using a dynamic Bayesian network. The inherent similarity or difference between sounds was determined by the corresponding similarity or difference between the audio features trajectory represented by HMM that approximates a general trend in query time behaviour.

In this paper we adopt different strategy to HMM modeling of an audio pattern. We propose the HMM approach that is inspired by methods very well explored in automatic speech recognition (ASR) and ASR-based spoken term retrieval. Every spoken word or sentence of speech of given language can be formed as a combination of basic acoustical-linguistic units, which are called phonemes. In every languages, there is a finite number of phonemes, e.g. there are approximately 44 phonemes in English language (may differ with particular dialect). In ASR-based keyword spotting systems, the search space is created only by statistical models of phonemes and

978-83-60810-53-8/$25.00

keywords or speech patterns are added into search as logical sequences of these models.

Our effort was to adopt the concept of elementary units for general sound and language independent speech recognition task. This approach requires to define a basic unit of sound (hereinafter called the elementary sound) and create a finite, but sufficiently large database of these units. For this purpose, we have adopted methods of unsupervised cluster analysis applied over huge amount of various audio data to create a database of elementary sound models. Major contribution of our proposal is that the elementary sound models can be treated as analogy to phoneme-based models in speech processing. Thus, statistical methods for speech recognition can be applied for general audio. As it is known to the authors, no similar method has been proposed in literature.

II. Concept of Elementary Sound Units

A. Elementary sounds

The idea of elementary sound units is based on the assumption that, in general, any acoustical pattern can be synthesized as a concatenation of short-time sound elements, stored in finite (but sufficiently large) sound inventory. This idea is derived from speech recognition and keyword spotting systems based on concatenation of sub-word phoneme units from finite database, to represent the sentences of speech. These units are predominantly represented by their parametric statistical models, created and estimated according to corresponding acoustical training examples.

Of course there are many challenges to successfully implement such approach. The biggest problem is the fact that unlike speech, which can be easily modeled since its physical production is known, general sounds are produced by unlimited number of ways and thus can be of infinite variance in temporal-spectral behavior. Unlike the phonemes, we can not practically create a database of the acoustical examples for general sound basic units. Instead, we have performed a cluster analysis over statistical models and created a database of the elementary sound models. The process of the creation is described below.

B. Parametric modeling

First step to create the database of the elementary sound models is to select a suitable method for statistical modeling and model parameters. For this purpose, we choose modeling via hidden Markov models (HMM), which is suitable for statistical description of time series of observations and therefore is commonly used in speech and general sound recognition tasks. A continuous-density HMM with N states consists of a set of parameters that generally comprises the matrix of transition probabilities, the initial state distribution, and the parameters of the state output density function, which is mainly approximated by a mixture of Gaussian components that can be expressed as follows:

$$P(\boldsymbol{x}|S) = \sum_k w_k \mathcal{N}_k(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (1)$$

where $\mathcal{N}_k$ is the Gaussian component with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$, and w_k is the weighting coefficient of this component in the state of the model. However, as the number of states in the search space grows, the system becomes more computationally and memory demanding. For this reason, we adopt a semi-continuous-density models, where all states share the same Gaussian components. Output density function of particular state is then determined only by the weighting vector $\boldsymbol{w}$.

Furthermore, assuming that only one of the components has major contribution to the resulting likelihood of the state, the summation operation in (1) can be replaced by the selection of maximal value. This modification reduce the processing time with minimal impact on resulting likelihood (mean difference less than 5%).

Similar to phoneme models, we choose a 3-state left-to-right model structure with equal transition probabilities, which can therefore be omitted.

C. Unsupervised clustering

One of the problems of statistical modeling is the determination of optimal number of Gaussian mixture components for output density function, which is usually determined by experiments [3], [4]. More sophisticated approach is based on Bayesian or Akaike Information Criterion [12], [13], Kullback-Leibler divergence [13], and unsupervised clustering methods [14], [15]. In [15], unsupervised *K-Variable K-means* clustering algorithm was proposed. This algorithm was adopted in our work to determine the optimal number of mixture components and also for derivation of the elementary sound models. The clustering algorithm is described below.

Algorithm 1 K-Variable K-means

1: Compute m and s as the mean and standard deviation of the distances between any pair of frames.
2: Set threshold distances: $T_1 = m - C.s$; $T_2 = m + C.s$ where $C \in < 0.5; 1.5 >$
3: Create 1st centroid as $\boldsymbol{c}_1 = \arg\max_i(\|\boldsymbol{x}_i\|)$
4: **for** $\forall \boldsymbol{x}_i$ **do**
5: $d_i = \min_j(d_{ij}(\boldsymbol{x}_i, \boldsymbol{c}_j))$; Find the distance to the closest centroid.
6: **if** $d_i < T_1$, Make $\boldsymbol{x}_i$ member of cluster j;
7: **if** $d_i > T_2$, Make $\boldsymbol{x}_i$ the center of a new cluster;
8: **end for**
9: Make all the remaining unclassified vectors members of their closest cluster.

D. Database of the Elementary Sound Models

The most important part of our system for general audio pattern searching is the database of the elementary sound models. Due to the gargantuan number of different sounds that may generally occur, it is impossible to create a finite database of acoustical training examples for elementary sound models. Therefore we have adopted the cluster analysis do derive and group sounds with similar statistical characteristics.

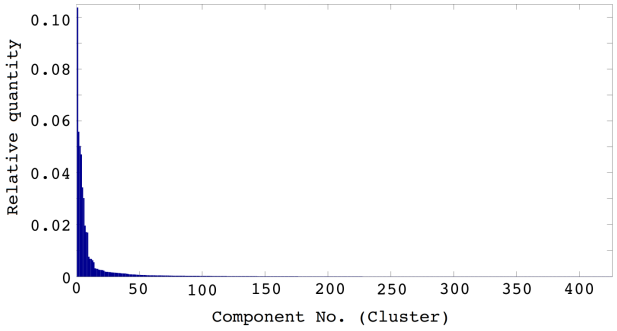


Fig. 1. Distribution of acoustical observations among clusters.

First, a sound database was collected, that consist of more than 30 hours of various short audio clips. This database involve different types of environmental and machinery sounds and noises, animal sound, human produced sounds and speech of different languages, music of different genres, etc.

The database was processed and the audio features described in section III were extracted. To obtain the Gaussian mixture components, all feature vectors were divided into clusters. Before the actual clustering, confusing data vectors were eliminated. The centroid of the whole data set was computed and the mean m and standard deviation s of distances of data vectors from this centroid were computed. Vectors with distance $d > m + 3.s$ were discarded.

Remaining data vectors were clustered using the unsupervised clustering and the means $\boldsymbol{\mu}_k$ and diagonal covariance matrices $\boldsymbol{\Sigma}_k$ occurred in (1) were computed for each cluster respectively. Clusters with less than 30 observations were discarded. As the result of clustering, state output density function consist of 426 Gaussian mixture components. Fig. 1 shows the distribution of acoustical observations among clusters.

Next, audio stream from all sound clips was formed and divided into 1 second long segments with 0.5 second overlap. From each segment a 3-state model was estimated. Because the models are defined by their weighting coefficient vector in each state, estimated models were clustered by unsupervised clustering. A histogram of distribution of the models within the clusters was computed. The clusters were assorted from the biggest to the smallest according to the number of members within the cluster. The elementary sound models were then derived as the means of the clusters. The clustering results in more than 10^4 clusters, although most of them comprise only one member. Therefore, only first 500 models were adopted for experiments.

Each of these models describe the audio segment in that the acoustical observations are statistically very similar. Conversely, acoustical observations of different elementary sounds differ in their characteristics.

III. Audio Features Selection

In order to achieve the best performance for classification, we have selected features that can capture the temporal and spectral characteristics of audio. Following work in [16], in which the features were selected by optimization algorithm, we have selected the following features:

1) *Line spectral frequencies/pairs (LSF/LSP)* - used as an alternative to linear prediction coefficients. The LSF are obtained by decomposing the LP filter transfer function A(z) into pair of auxiliary polynomials:

$$\begin{aligned} P(z) &= A(z) + z^{-p+1}A(z^{-1}) \\ Q(z) &= A(z) - z^{-p+1}A(z^{-1}) \end{aligned} \tag{2}$$

where $P(z)$ is symmetrical and $Q(z)$ asymmetrical p+1-order polynomial, where the zeros of $A(z)$ are mapped onto the unit circle in the z–plane.

2) *Spectral flux (SFX)* - measures changes in the shape of magnitude spectrum by calculating the difference between magnitude spectra of successive frames. The spectral flux is computed for frame at discrete time t as follows:

$$SFX(t) = \frac{\sum_k [a_k(t) - a_k(t-1)]^2}{\sqrt{\sum_k a_k(t)^2}\sqrt{\sum_k a_k(t-1)^2}} \tag{3}$$

where a_k is the k-th element of magnitude spectrum of given frame. Spectral flux describes the temporal changes of magnitude spectrum, thus represents the dynamic coefficients of spectrum.

3) *Zero crossing rate (ZCR)* - the number of time-domain zero-crossings within a frame, computed as a number of sample sign changes.

These features were extracted using the Yaafe [17] extraction tool. The resulting feature vector consists of 10 LSF's, one SFX, and one ZCR coefficient.

IV. Audio Pattern Searching

The theoretical background for sequential audio pattern searching was taken from [18] where a decoder for keyword spotting was proposed. Its function is based on Viterbi algorithm with propagation of accumulated score through the

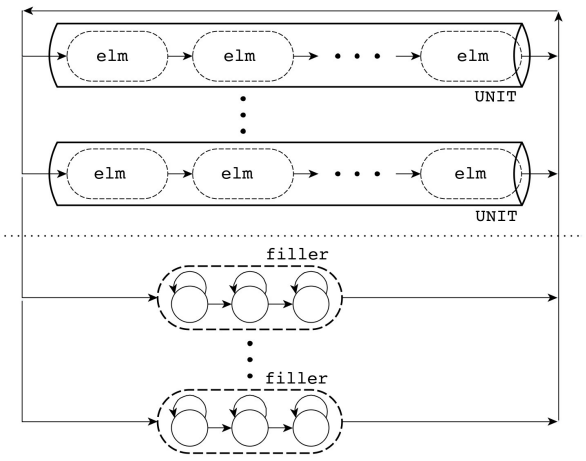


Fig. 2. Looped network of units and fillers.

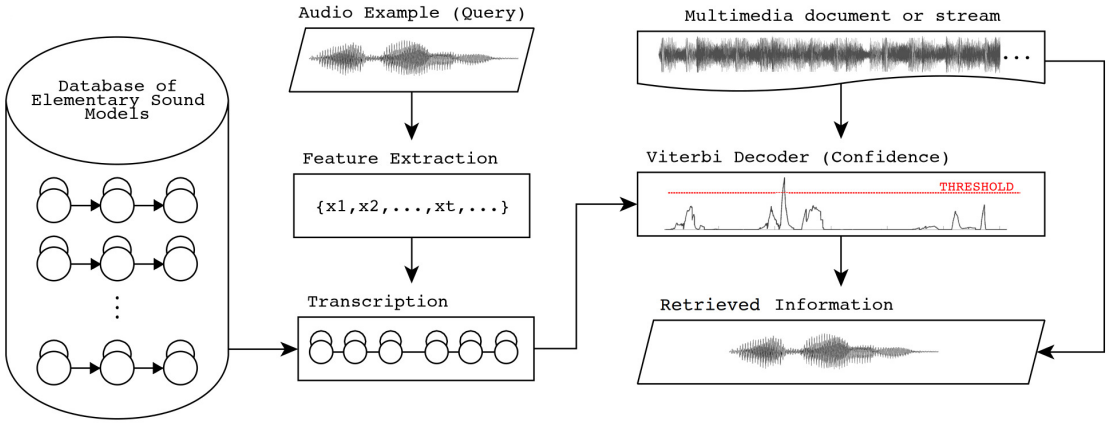


Fig. 3. Query-by-Example search.

looped network of units that represents searched keywords, and fillers.

In our implementation, the basic unit of the decoder is the elementary sound. In process of creating the model of demanded audio pattern, a representative audio sample is passed into decoder and transcribed into sequence of the elementary sound models. This transcription is found as the path through the elementary sound models with the highest score achieved for the training example. This logical sequence of elementary sounds is then added into search space as a unit, representing the searched audio pattern. Example of model network with units and fillers is shown in Fig.2.

In the process of decoding, each acoustical observation must be assigned to one of the states in model network. The acoustical observations between segments that correspond with searched patterns are assigned to the filler models, that represents any sound that may occur in background. This definition of filler offers us the possibility of using the proposed database of elementary sound models itself.

Another benefit of using the elementary sounds as fillers is that we can compute the confidence of particular unit as proposed in [18]. The confidence $C(u,t)$ of unit u at discrete time t is defined as normalized acoustic score as follows:

$$C(u,t) = S(u,t)/S(f_c,t) \tag{4}$$

where $S(u,t)$ it the acoustic score achieved for segment of audio by unit u, and $S(fc,t)$ is the acoustic score achieved by the best concatenation of fillers (in our case the elementary sound models) for the same segment. It follows that the confidence reaches the maximum value of 1 only when the score of the unit and the score of fillers concatenation are equal. In this case, the audio segment precisely correspond with the unit training example. In other cases when $C(u,t) < 1$ the probability of correct detection decreases and proper threshold must be set for experimental data. Fig. 3 shows the principal functions of the complex system for query-by-example general audio pattern recognition.

V. Experimental Results

A. Experiment setup

The experiments on the proposed database of the elementary sound models were performed in task of audio pattern search in recordings, which include acoustical patterns of five different sound types: *applause, crying, laughing, gunshot and speech (10 Slovak keywords)*. Ten artificial audio tracks were created by random concatenation of 20 audio examples from each sound class and various types of environmental background noises respectively. For each audio track, the examples were chosen randomly from the sound database.

Accuracy of the search was evaluated on the level of acoustical observations against the human annotation of records with common precision (P), recall (R), and F_1-measure (F_1) metrics defined as follows:

$$P = \frac{n_{correct}}{n_{total}}; R = \frac{n_{correct}}{n_{target}}; F_1 = \frac{2.P.R}{P+R}; \tag{5}$$

where $n_{correct}$ stands for correct positive detections, n_{total} for total positive detections and n_{target} for target positive detections.

B. Audio event detection

For each searched sound example, the unit was created, as described in Section IV. Each of these units was used as query for searching. Units were able to find corresponding training examples with practically 100 % accuracy and confidence close to 1. Although using simple correlation will be much more efficient in this case, this experiment proved our prior assumption that specific audio pattern can be modeled as a sequence of the elementary sound models.

In the next experiment, only one representative example for each class was used as query for search. By changing the confidence threshold, the balance between precision and recall was set, to obtain highest possible F-measure score. The decoder was able to find also other audio segments similar to the queries. Overall average precision and recall reaches 62.5 % and 58.5 % respectively.

TABLE I
EXPERIMENTAL RESULTS ON VARIOUS ACOUSTIC PATTERNS

Pattern (Class)	*	**Number of examples in query**		
		1	***2***	***3***
Applause	P	73.40 %	81.55 %	89.33 %
	R	76.78 %	97.81 %	96.75 %
	F_1	75.60 %	88.95 %	92.89 %
Crying	P	55.69 %	85.98 %	81.52 %
	R	55.50 %	52.49 %	57.71 %
	F_1	55.60 %	65.18 %	67.58 %
Laughing	P	52.88 %	98.73 %	83.76 %
	R	41.21 %	52.07 %	80.08 %
	F_1	46.32 %	68.18 %	81.88 %
Gunshot	P	87.18 %	84.35 %	91.00 %
	R	80.68 %	79.25 %	88.62 %
	F_1	78.07 %	81.72 %	89.80 %
Keywords	P	43.17 %	75.41 %	80.78 %
	R	48.38 %	63.17 %	70.22 %
	F_1	45.63 %	68.75%	75.13 %

In the next search run, one additional representative example of each class was added into the search space, so that two examples were in the query. By setting the confidence threshold for each example of representative pair, the overall average precision and recall increased to 85.2 % and 69.0 % respectively. Lastly, three representative examples were selected as queries and put into search. The confidence threshold was set for each examples of three. The average precision and recall again increased to 85.3 % and 78.7 % respectively.

The advantage of our system is that if a user has only one example of demanded audio pattern (query-by-example), new examples can be added into search as selected audio segments found in previous run. Thus, the system has ability to "learn" from users feedback after the search. Table I shows the average results achieved using 1, 2, and 3 examples in query for each sound class.

VI. CONCLUSION

The system for universal sequential audio pattern search has been proposed. Statistical model specifications were introduced and the database of elementary sound models was created, using the unsupervised clustering method. Experimental results show that it is possible to adopt the concept of elementary sound units for general audio pattern modeling and recognition. However, a proper confidence threshold must be set experimentally for each unit to obtain the best possible result. Experiments also show that adding more examples of particular audio pattern can significantly improve searching results. If these examples are selected by a user from previous search run results, the system is also able to learn according to user feedback.

In our future work, we aim to include expectation–maximization algorithm in the clustering and adopt Viterbi alignment and discriminative training for better discrimination of elementary sound models. We also plan to compare the performance of the system on larger set of audio features extracted from audio data.

ACKNOWLEDGMENT

This work has been supported by the Centre of excellence for systems and services of intelligent transport II., ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] Rouas, J. L., Louradour, J., & Ambellouis, S. (2006, September). Audio events detection in public transport vehicle. In Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE , Sept. 2006, pp. 733-738

[2] Vozarikova E.; Pleva, M.; Juhar, J.; Cizmar, A.,"Surveillance system based on the acoustic events detection", Journal of Electrical and Electronics Engineering. Vol. 4, no. 1, 2011, pp. 255-258.

[3] Lian-Hong Cai; Lie Lu; Hanjalic, A.; Hong-Jiang Zhang; Lian-Hong Cai, "A flexible framework for key audio effects detection and auditory context inference," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.3, pp.1026,1039, May 2006

[4] Heittola, T.; Mesaros, A.; Eronen, A.; Virtanen, T.; "Audio context recognition using audio event histograms," 18th European Signal Proc. Conf. (EUSIPCO-2010) Aalborg, Denmark, August 23-27, 2010

[5] Atrey, P.K.; Maddage, M.C.; Kankanhalli, M.S., "Audio Based Event Detection for Multimedia Surveillance," IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 , vol.5, no., pp.V,V, 14-19 May 2006

[6] Lemström, K., Tzanetakis, G.: Music Information Retrieval. In Bates, M.J. (Ed.): Understanding Information Retrieval Systems: Management, Types, and Standards, CRC Press, 2012.

[7] Chandrasekhar, V., Sharifi, M., & Ross, D. A. "Survey and Evaluation of Audio Fingerprinting Schemes for Mobile Query-by-Example Applications." In Int. conf. ISMIR, pp. 801-806. 2011.

[8] Stan Z. Li,"Content-Based Audio Classificafion and Retrieval Using the Naerest Feature Line Method",IEEE Transactions speech and audio processing, vo1.8, No.5, Sept. 2000.

[9] Marko Helen, Tuomas Virtanen . Audio query by example using similarity measures between probability density functions of features. EURASIP Journal on Audio, Speech, and Music Processing, 2010.

[10] Velivelli, A.; Zhai, C.X.; Huang, T.S., "Audio segment retrieval using a short duration example query," IEEE Int. Conf. on Multimedia and Expo, ICME '04., vol.3, pp.1603-1606, June 2004

[11] Wichern, G., Xue, J., Thornburg, H., Mechtley, B., & Spanias, A. Segmentation, indexing, and retrieval for environmental and natural sounds., IEEE Transactions on Audio, Speech, and Language Processing 18(3), 2010, pp. 688-707

[12] Tenmoto, H.; Kudo, M.; Shimbo, M., "Determination of the number of components based on class separability in mixture-based classifiers," Third International Conference Knowledge-Based Intelligent Information Engineering Systems, 1999. vol., no., pp.439,442, Dec 1999

[13] Xiao-Bing Li; Ren-Hua Wang, "State Divergence-Based Determination of The Number of Gaussian Components of Each State in HMM," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP 2006, vol. 1, no., pp.I,I, 14-19 May 2006

[14] Mucciardi, Anthony N.; Gose, Earl E., "An Automatic Clustering Algorithm and Its Properties in High-Dimensional Spaces," IEEE Trans. on Systems, Man and Cyber., vol.SMC-2, no.2, pp.247,254, April 1972

[15] Reyes-Gomez, M. J.; Ellis, D. P. W., "Selection, parameter estimation, and discriminative training of hidden Markov models for general audio modeling," International Conference on Multimedia and Expo, 2003. ICME '03. vol.1, no., pp.I,73-6 vol.1, 6-9 July 2003

[16] Chmulik, M.; Jarina, R., "Bio-inspired optimization of acoustic features for generic sound recognition," 19th Int. Conf. on Systems, Signals and Image Proc. (IWSSIP), 2012, pp.629,632, 11-13 April 2012

[17] B.Mathieu; S.Essid; T.Fillon; J.Prado; G.Richard; YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software, proceedings of the 11th ISMIR conference, Utrecht, Netherlands, 2010.

[18] J. Nouza, J. Silovsky, "Fast keyword spotting in telephone speech," Radioengeneering 2009, vol. 18, no. 4, Dec. 2009, pp. 665-670

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 571–575

Dependence of Kinect sensors number and position on gestures recognition with Gesture Description Language semantic classifier

Tomasz Hachaj
Pedagogical University of Krakow
2 Podchorazych Ave,
30-084 Krakow, Poland
Email: tomekhachaj@o2.pl

Marek R. Ogiela
AGH University of Science and Technology,
30 Mickiewicza Ave,
30-059 Krakow, Poland
Email: mogiela@agh.edu.pl

Marcin Piekarczyk
Pedagogical University of Krakow,
2 Podchorazych Ave,
30-084 Krakow, Poland
Email: marp@up.krakow.pl

***Abstract*—We have checked if it is possible to increase effectiveness of standard tracking library (Kinect Software Development Kit) by fusion of body joints gathered from different sensors positioned around the user. The proposed calibration procedure enables integration of skeleton data from set of tracking devices into one skeleton. That procedure eliminates many segmentation and tracking errors. The test set for our methodology was 700 recordings of seven various Okinawa Shorin-ryu Karate techniques performed by black belt instructor. In case when side Kinects were rotated in $\frac{\pi}{2}$ and -$\frac{\pi}{2}$ around vertical axis relatively to central one number of all not classified Karate techniques dropped by 48% while excessive misclassification error remained in the same level.**

***Index Terms*—Gesture recognition, Gesture Description Language, time sequence analysis, Kinect, pattern classification, semantic approach, Karate.**

I. Introduction

THE COMMON approach in gesture recognition is partitioning the movement sequence into sections that are represented by key frames. Those frames are than classified by different recognition techniques. For example in [1] authors propose an automatic learning method for gesture recognition. First, they apply the Self-Organizing Map to divide the sample data into phases and construct a state machine. Next, they apply the Support Vector Machine to learn the transition conditions between nodes. Nowadays multimedia devices that enable real-time tracking of observed users (like Microsoft Kinect controller) can be bought relatively cheaply. Because of that more and more researches apply them to record human body data (called body joints) that are automatically segmented from depth camera video data by dedicated software (like one implemented in Kinect SDK - Software Development Kit) like in [2], where a comparison of human gesture recognition using data mining classification methods in video streaming is proposed. The recognized gesture patterns of the study are stand, sit down, and lie down. Classification methods chosen for comparison study are back propagation neural network, support vector machine, decision tree, and naive Bayes. It has been proved that data acquired by Kinect device can be use not only to classify typical common - live gestures like waving or sitting but also to recognize high-speed gestures of martial arts sportsmen. In [3] authors aim at automatically recognizing sequences of complex Karate movements and giving a measure of the quality of the movements performed. The proposed system is constituted by four different modules: skeleton representation, pose classification, temporal alignment, and scoring. The proposed system is tested on a set of different punch, kick and defense Karate moves executed starting from the simplest case, i.e. fixed static stances up to sequences in which the starting stances is different from the ending one. All previously described methods uses many body features as an input for classification algorithm. However, in [4] the authors showed, that the majority of the information regarding the human motion resides in a lower dimensional space than one that can be obtained from all available features. These considerations further support the argument that human motion can be classified using a representation which considers a relatively low number of dimensions [5].

Knowing all of this we have proposed our new semantic classifier [6] called Gesture Description Language (GDL). The idea of GDL approach is to code the gesture sequences as the series of static key frames that appears in defined order. Those sequences are coded with context-free grammar called GDL script. GDL script consists of set of rules that creates together knowledge database similar to one an expert system has. The preliminary description of GDL architecture has been presented elsewhere [6], [7]. The basic assumption of GDL is:

- It is capable of classifying human body movements in real time.
- It can classify not only simple, real life gestures but also complicated movements like Karate techniques.
- It does not require large training dataset. Gestures are defined by user in GDL script. User can utilize as many body features as he or she needs in each rule definition.
- Gestures are split into key frames that appears in some order under given time restriction.
- The input data for classifier is set of body joints that arrive from tracking software in real - time (approximately with

978-83-60810-53-8/$25.00

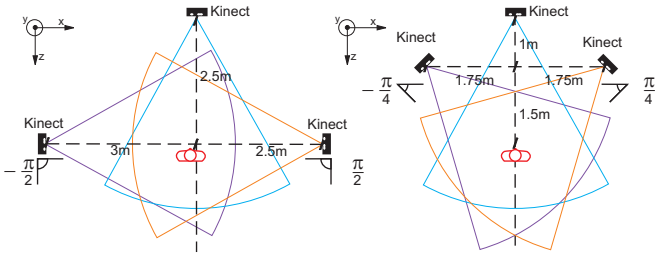


Fig. 1. Two tested multi-kinect environment configuration. Length of Kinect's view cone is 4 meters.

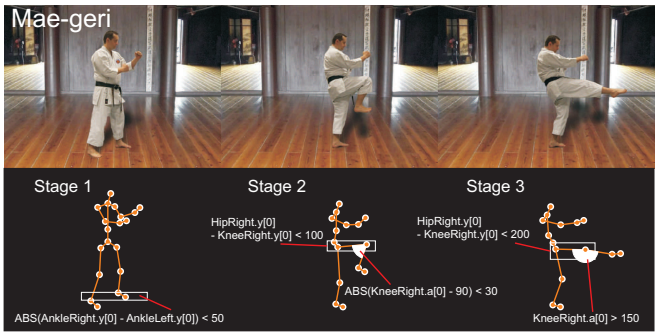


Fig. 2. Example Karate key-frames for GDL script. Movements are separated into stages, each stage is a key-frame used in semantic description. In this picture Mae-geri (front kick) begins with Moto-dachi (stance) but in our experimental recordings set it started from different, "neutral" position.

frequency 30 Hz). The set of tracked body joints is called Skelton (see Figure 2, bottom row).

- Our notation is invariant to user rotation around viewport of camera, because it can generates features regarding to angles measured between vectors defined by pairs of body joints (similarly to approach in [3]). Hovewer in opposite to [3] we can defines those angles dynamically while tailoring the GDL script description.

The idea of applying formal language to describe gestures is not new and was previously introduced for example in [8], [9]. However those papers describe only general framework of gesture description that might be potentially applicable for further recognition. The authors did not show how to use they annotations in pattern recognition tasks. They did not also validate they approach on any type of real-life data. Because of that those previous approaches were rather purely theoretical.

The novel contribution of this paper is test of GDL classifier performance on dataset that was acquired with one or three Kinect sensors that were positioned in two different configurations. We have checked if it is possible to increase effectiveness of standard tracking library (Kinect SDK) by fusion of body joints gathered from different sensors positioned around the user. The proposed calibration procedure enables integration of skeleton data from set of tracking devices into one skeleton. The test set was various Okinawa Shorin-ryu Karate techniques performed by black belt instructor. The whole solution runs in real-time and enables online and offline classification.

II. Material and Methods

In this section we will present experimental hardware setup, basis of GDL description and test dataset.

A. Multi - Kinect environment: setup and calibration

Figure 1 presents two tested multi-Kinect environment configuration. Each Kinect use its own tracking module (well know algorithm from Microsoft Kinect SDK which implementation can be used out of charges) that segments and tracks user skeleton in real time.

If front Kinect does not "see" particular body joint system checks if this joint is visible by another device. If yes our software takes coordinates measured by that second device. If more than two devices have detected same joint, coordinates are taken from camera that is closest to observed point. Each Kinect measure distance to observed point in its own right-handed Cartesian frame situated relatively to sensor orientation. Because of that same point V has different coordinates $\overline{v'} = [x', y', z', 1]$ and $\overline{v} = [x, y, z, 1]$ relatively to each pair of devices.

Our task now is to map all of those points to the same coordinate system. Let us assume that a Cartesian frame that represents orientation of each Kinect was translated and rotated around y (vertical) axis relatively to each other frame. That means there are four degrees of freedom (three for translation, one for rotation). Knowing that the linear transformation that maps coordinates of a point represented by vector $\overline{v'}$ in one coordinate system to coordinates $\overline{v}$ in another one has form of following matrix:

$$\overline{v'} \cdot \begin{bmatrix} cos(\beta) & 0 & -sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ sin(\beta) & 0 & cos(\beta) & 0 \\ t_x & t_y & t_z & 1 \end{bmatrix} = \overline{v} \tag{1}$$

In order to find unknown matrix coefficients following linear system has to be solved:

$$\begin{bmatrix} x'_1 & z'_1 & 1 & 0 \\ z'_1 & -x'_1 & 0 & 1 \\ x'_2 & z'_2 & 1 & 0 \\ z'_2 & -x'_2 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} cos(\beta) \\ sin(\beta) \\ t_x \\ t_z \end{bmatrix} = \begin{bmatrix} x_1 \\ z_1 \\ x_2 \\ z_2 \end{bmatrix} \tag{2}$$

$$y' + t_y = y \tag{3}$$

Where $\overline{v_1} = [x_1, y_1, z_1, 1]$, $\overline{v_2} = [x_2, y_2, z_2, 1]$ are points which coordinates are known in both frames. The linear system (2) and equation (3) is product of multiplication of matrix (1) by $\overline{v_1}$ and by $\overline{v_2}$. They multiplication by $\overline{v_1}$ produces the first and second equation in matrix (2) and equation (3). The multiplication by $\overline{v_2}$ produces third and fourth equation in matrix (2).

B. Bases of GDL scripts

The preliminary description of GDL architecture has been presented elsewhere [6], [7]. Because of that we will present only one example GDL script and its graphical explanation.

As we previously mentioned movement is separated into key frames. Each key frame is repressed by a rule that has a conclusion. If rule is satisfied for actual set of body joints positions (GDL uses forward chaining rezoning schema) its conclusion is memorized. It is possible to check with GDL script if some conclusion was satisfied in given time period. With this mechanism it is possible to generate key frames chains, which together create gesture. First row of Figure 2 presents three key frames of GDL script from Appendix that describes the Mae-geri kick. Second row explains body joints dependencies that are present in GDL script description. The last rule in script (the one with *sequenceexists* function) checks if all stages of movement have appeared in defined order under 1 second time restriction.

C. Test dataset

The dataset for our research are recordings of black belt Karate instructor[1] that performs seven different techniques: four static stances (Moto-dachi, Zenkutsu-dachi, Shiko-dachi and Naihanchi-dachi), two blocks (Gedan-uke and Age-uke) and one kick (Mae-geri). The instructor has indicated essential aspects of each technique (starting and ending positions of limbs and movement trajectory). The data was recorded during two sessions: one in which cameras was positioned as it was presented in Figure 1 on left, the second one as in Figure 1 on right. Second recording was done several weeks after the first one. Each gesture was partitioned into key-frames (Figure 3) that was later verified and accepted by instructor. Also expert was present during final validation of method. The same GDL script was used for all of recordings. The frame capture frequency was 30 Hz.

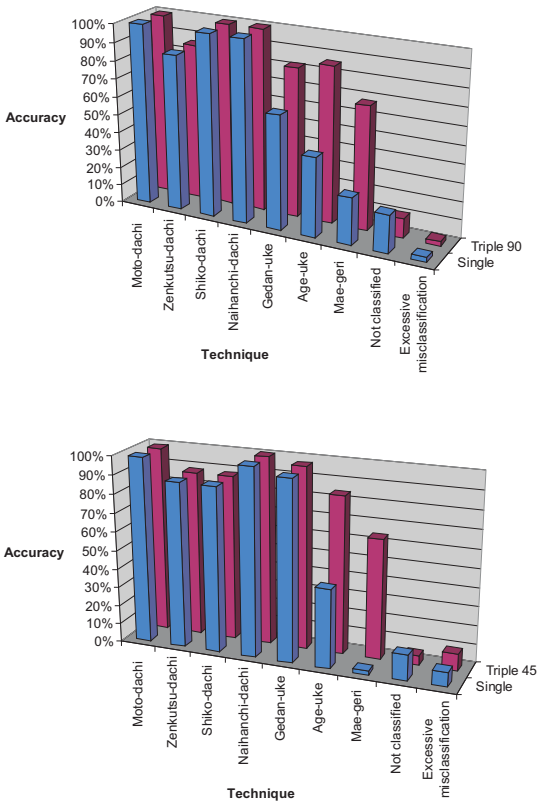


Fig. 3. Classification results from single and triple Kinect recordings. Triple 90 is left setup from Figure 1, Triple 45 is right setup from Figure 1.

III. Results

Tables 1-4 summarize the classification results of our experiment. The description in first column is the actual technique (or group of techniques) that is present in particular recording. Each technique (or group of techniques) was repeated 50 times. Symbol + means that particular recording consisted of more than one technique. Description in first row is classification results. Last but one row sums up percentage of correct classifications of particular technique. The last row sums up the percentage of correct classifications of techniques from first column. Summing up, we had 350 recordings of Karate techniques in each Kinect configuration (totally 700 recordings).

Because several Karate techniques can be present in same movement sequence we investigated if actual technique/techniques was/were classified. If yes that case was called correct classification. If technique was not classified and was not mistaken with similar one (like Moto-dachi which is similar to Zenkutsu-dachi) that case was called not classified. If technique was mistaken with similar one that case was called misclassified. Those three sums up to 100%. If technique was correctly classified but additional - actually not present - behavior was classified that case was called excessive misclassification. According to this terminology 90.4% of recordings from Table 4 was correctly classified, 5.2% was not classified and 4.4% was misclassified. Excessive misclassification was at the level of 9.0%. Figure 3 graphically presents results from Table 1-4.

IV. Discussion and conclusion

Our experiment has shown that integration of tracking data acquired by several Kinect devices with standard software increases the effectiveness of GDL classifier. This is due the fact that additional sensors that are situated at different angles than central one are capable of tracking body joints that in some situations might be covered by different body parts. This condition is especially visible in case of Mae-geri. Tracking of Karate kick is difficult task because of two factors: feet is moving with relatively high speed with large radius of path and in the last stage of Mae-geri feet is situated nearly at the same horizontal position as hip and knee. If the sportsman[2] is filmed only in front view knee and hip body joints are covered by feet and proper position of them have to be approximated by the software what, in practice,

[1]Karate instructor of Okinawa Shorin-ryu Karate with black belt degree (3 dan, sandan)

[2]In the meaning of Karate practitioner

TABLE I
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH SINGLE KINECT DEVICE (CENTRAL ONE) IN FIRST RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	6	0	0	0	7
Zenkutsu-dachi	1	37	0	0	1	0	0	12	1
Shiko-dachi +gedan-barai	0	0	50	0	27	0	0	0+23=23	0
Naihanchi-dachi	0	0	0	50	0	0	0	0	0
Gedan-barai +Zenkutsu-dachi	0	49	0	0	36	0	0	1+14=15	0
Age-uke +Moto-dachi	50	0	0	0	0	22	0	0+28=28	0
Mae-geri	4	0	11	0	0	0	13	26	4
%	100%	86.0%	100%	100%	63.0%	44.0%	26.0%	20.8%	2.4%

TABLE II
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH THREE KINECT DEVICES SITUATED AS SHOWN IN FIGURE 1 ON THE LEFT IN FIRST RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	6	0	0	0	7
Zenkutsu-dachi	1	37	0	0	1	0	0	12	1
Shiko-dachi +gedan-barai	0	0	50	0	46	0	0	0+4=4	0
Naihanchi-dachi	0	0	0	50	0	0	0	0	0
Gedan-barai +Zenkutsu-dachi	0	49	0	0	36	0	0	1+14=15	0
Age-uke +Moto-dachi	50	0	0	0	0	43	0	0+7=7	0
Mae-geri	4	0	0	0	0	0	34	16	4
%	100.0%	86.0%	100.0%	100.0%	82.0%	86.0%	68.0%	10.8%	2.4%

TABLE III
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH SINGLE KINECT DEVICE (CENTRAL ONE) IN SECOND RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	16	0	0	0	17
Zenkutsu-dachi	2	43	0	0	14	0	0	5	14
Shiko-dachi +gedan-barai	0	0	44	1	49	0	0	6+1=7	1
Naihanchi-dachi	0	0	0	50	7	0	0	0	7
Gedan-barai +Zenkutsu-dachi	2	45	0	0	47	0	0	3+3=6	0
Age-uke +Moto-dachi	49	0	0	0	0	21	0	1+29=30	0
Mae-geri	0	5	17	0	0	0	1	27	0
%	99.0%	88.0%	88.0%	100.0%	96.0%	42.0%	2.0%	15.0%	7.8%

TABLE IV
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH THREE KINECT DEVICES SITUATED AS SHOWN IN FIGURE 1 ON THE RIGHT IN SECOND RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	16	0	0	0	17
Zenkutsu-dachi	2	43	0	0	14	0	0	5	14
Shiko-dachi +gedan-barai	0	0	44	1	49	0	0	6+1=7	1
Naihanchi-dachi	0	0	0	50	7	0	0	0	7
Gedan-barai +Zenkutsu-dachi	2	45	0	0	48	0	0	3+2=5	0
Age-uke +Moto-dachi	49	0	0	0	0	42	0	1+8=9	0
Mae-geri	0	0	23	1	0	0	32	0	6
%	99.0%	88.0%	88.0%	100.0%	97.0%	84.0%	64.0%	5.2%	9.0%

generates serious positioning errors. Our results have showed that in both configurations of multi-Kinect environment the effectiveness of classification increases because of increasing of tracking accuracy. In case when side Kinects were rotated about $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ around vertical axis relatively to central one number all not classified techniques dropped by 48% while excessive misclassification error remained on the same level. In case when Kinects were rotated about $\frac{\pi}{4}$ and $-\frac{\pi}{4}$ around vertical axis relatively to central one number all not classified techniques dropped by 61.9% while excessive misclassification error increased by 15.4%. It can be concluded that if we want to increase the correct classification factor in case when excessive misclassification error is not critical second setup of Kinects is more profitable. Otherwise, one should apply first setup, which, in our experiment did not change excessive misclassification rate.

Our future goal will be development of GDL script for recognition of complete set of most popular Karate techniques. The completed classifier will be than utilized in self-training multimedia application. We also plan to apply our classifier as a part of touchless interface in our medical data visualization module [10]. This will allow medical personnel to personally access patient data during surgical interventions while their hands are sterile. We also consider to expand GDL script terminal symbols and to test its capability in recognition of sign language gestures [11].

APPENDIX

The GDL script for Mae-geri recognition.

```
//////////////////
//Mae-geri
//////////////////
//Both legs are in the same level above the ground
//Figure 2 Mae-geri stage 1
RULE ABS(AnkleRight.y[0] - AnkleLeft.y[0]) < 50
THEN MaeStart

//Right knee in the line with right hip, bended
//right knee
//Figure 2 Mae-geri stage 2
RULE (HipRight.y[0] - KneeRight.y[0]) < 100
& ABS(KneeRight.a[0] - 90) < 30
THEN MaeMiddleRight

//Kick with right foot - Figure 3 Mae-geri stage 3
RULE (HipRight.y[0] - KneeRight.y[0]) < 200
& KneeRight.a[0] > 150
THEN MaeEndRight

//Left knee in the line with left hip, bended left knee
//Figure 2 Mae-geri stage 2
RULE (HipLeft.y[0] - KneeLeft.y[0]) < 100
& ABS(KneeLeft.a[0] - 90) < 30
THEN MaeMiddleLeft

//Kick with left foot - Figure 3 Mae-geri stage 3
RULE (HipLeft.y[0] - KneeLeft.y[0]) < 200
& KneeLeft.a[0] > 150
THEN MaeEndLeft

//Proper sequence of Mae-geri stages
RULE (sequenceexists("[MaeMiddleRight,1][MaeStart,1]")
& MaeEndRight) |
(sequenceexists("[MaeMiddleLeft,1][MaeStart,1]")
& MaeEndLeft)
THEN Mae-geri
```

ACKNOWLEDGMENT

We kindly acknowledge the support of this study by a Pedagogical University of Krakow Statutory Research Grant.

REFERENCES

[1] M. Oshita, T. Matsunaga, Automatic Learning of Gesture Recognition Model Using SOM and SVM, Advances in Visual Computing, Lecture Notes in Computer Science, vol. 6453, 2010, pp. 751–759'

[2] O. Patsadu, C. Nukoolkit, B. Watanapa, Human gesture recognition using Kinect camera, Joint International Conference on Computer Science and Software Engineering (JCSSE), 2012, pp. 28–32.

[3] S. Bianco, F. Tisato, Karate moves recognition from skeletal motion, Proc. SPIE 8650, Three-Dimensional Image Processing (3DIP) and Applications 2013, 86500K (March 12, 2013); doi:10.1117/12.2006229.

[4] V. Ntouskos, P. Papadakis, F. Pirri, A Comprehensive Analysis of Human Motion Capture Data for Action Recognition, VISAPP 1, pp. 647–652. SciTePress, (2012).

[5] M. Trzupek, Semantic Interpretation of Heart Vessel Structures Based on Graph Grammars, Computer Vision and Graphics, Lecture Notes in Computer Science, vol. 6374, 2010, pp. 81–88.

[6] T. Hachaj, M. R. Ogiela, Recognition of human body poses and gesture sequences with gesture description language, Journal of medical informatics and technology, vol 20/2012, ISSN 1642-6037, October 2012, pp. 129–135.

[7] T. Hachaj, M. R. Ogiela, Semantic Description and Recognition of Human Body Poses and Movement Sequences with Gesture Description Language, Computer Applications for Bio-technology, Multimedia, and Ubiquitous City, Communications in Computer and Information Science, Vol. 353, 2012, pp. 1–8, Springer, Heidelberg.

[8] F. Echtler, G. Klinker, A. Butz, Towards a unified gesture description language, Proceeding HC '10 Proceedings of the 13th International Conference on Humans and Computers, pp. 177–182 University of Aizu Press Fukushima-ken, Japan, 2010.

[9] M. Kölsch, C. Martell, Towards a Common Human Gesture Description Language, Workshop on Mixed Reality User Interfaces, at VR 2006.

[10] T. Hachaj and M. R. Ogiela, Framework for cognitive analysis of dynamic perfusion computed tomography with visualization of large volumetric data, Journal of Electronic Imaging, vol. 21, issue 4, 10.1117/1.JEI.21.4.043017, 2012.

[11] W. Koziol, H. Wojtowicz, K. Sikora, W. Wajs, Analysis and Synthesis of the System for Processing of Sign Language Gestures and Translatation of Mimic Subcode in Communication with Deaf People, Knowledge Engineering, Machine Learning and Lattice Computing with Applications, Lecture Notes in Computer Science, vol. 7828, 2013, pp 61–70.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 577–584

Automatic Identification of Broadcast News Story Boundaries Using the Unification Method for Popular Nouns

Zainab Ali Khalaf[1,2]
[2]Department of Computer Science, College of Science, University of Basra, Iraq
Email: zainab_ali2004@yahoo.com

Tan Tien Ping
[1]School of Computer Sciences, Universiti Sains Malaysia USM, 11800 Penang, Malaysia
Email: tienping@cs.usm.my

Abstract—Herein we describe the latent semantic algorithm method for identifying broadcast news story boundaries. The proposed system uses the pronounced forms of words to identify story boundaries based on popular noun unification. Commonly used clustering methods use latent semantic analysis (LSA) because of its excellent performance and because it is based on deep semantic rather than shallow principles. In this study, the LSA algorithm with and without unification was used to identify boundaries of Malay spoken broadcast news stories. The LSA algorithm with the noun unification approach resulted in less error and better performance than the LSA algorithm without noun unification.

Keywords: spoken document; broadcast news; story boundary identification; latent semantic analysis

I. Introduction

Because nouns bear more semantic meaning than other parts of speech and because they are the main characteristics used to identify documents stories [1], natural language processing applications often focus on nouns as essential components of the documents being processed. Names of persons, for example, are useful noun components in natural language processing, especially during automatic sentence clustering. In recent years, spoken document processing has become a popular and interesting topic within the field of natural language processing. In general, spoken document processing adapts natural language processing applications using speech input rather than text input.

Processing spoken documents is challenging because of the word errors generated by the automatic speech recognition (ASR) process [2], [3]. Determining the boundaries of broadcast news stories is another obstacle to processing spoken documents. The lack of overt punctuation and formatting contributes to this problem. In order to retrieve information, the beginning and the end of the segments or paragraphs within a document must be determined [3]-[6]. The process of determining the boundaries of the segments in the text is not an easy process [3]-[7].

Word errors generated by the ASR process can occur when recordings are made in a noisy environment or when pronunciation is unclear. The latter is especially true for vowel letters. An example from a Malay broadcast news story is as follows: The name of a professional badminton player was written four different ways in four sentences when converted from spoken news to written news by the ASR system (lee chong wei, choong wei, chong wee, and chan wee). The conversion problem was related to the vowel sound, in that the [u:] sound can be written as “oo, o, ou, ew, ue, u, and ui” and the [i:] sound can be written as “ee, ea, ei, and ie.” Silent sounds (pronounced n+ unpronounced g) also pose problems for ASR [8], [9].

Identification of story boundaries with the added problem of pronunciation errors is a complicated task. It requires human knowledge of the rules of correct pronunciation of lexical items. To address these problems, we propose a new method to improve story boundary identification in spoken documents using the popular noun unification approach.

II. Related Work

The absence of punctuation and capitalization in spoken documents makes it challenging to automatically identify story boundaries in multimedia documents. Previous attempts have concentrated on three types of cues: visual cues, such as the presence of an anchor’s face [7] or motion changes [7]; audio cues, such as significant pauses or reset of pitch; and lexical cues, such as word similarity measures within speech recognition transcripts or closed captions of video [10], [11]. Cues from completely different modalities (audio, video, and text) are often consolidated to achieve better story boundary identification [7], [12].

Hearst et al. (1997) proposed the TextTiling approach to story boundary identification [10]. It is based on the straightforward observation that different topics usually employ different sets of words and that shifts in vocabulary usage are indicative of topic changes [10]. As a result, pairwise

978-83-60810-53-8/$25.00

sentence similarities are measured across the text and a local similarity minimum implies a story boundary. Stokes et al. (2004) evaluated word cohesion using a lexical chaining approach; in this method, related words in a text are linked into chains, and a high concentration of chain starting and ending points is an indication of a story boundary [11]. These two approaches were recently used to segment speech recognition transcripts of spoken documents such as broadcast news [12], [13] and meetings [14]. Rosenberg et al. (2006) presented results from a broadcast news story boundary identification system developed for the SRI NIGHTINGALE system, which was applied to English, Arabic, and Mandarin news show to provide input for subsequent question-answering processes [12]. Xie (2008) used word and subword multiple scales for story boundary identification and showed the robustness of subwords for reducing the impact of errors and improving identification of broadcast news story boundaries [13] . Wu (2009) used decision tree and maximum entropy methods to identify the positional story boundaries locally and then used a genetic algorithm to identify the final story boundaries [15].

III. Latent Semantic Analysis

The clustering of sentences can be used to find repeated information, and the clustering process is conducted by grouping similar sentences together. Previous studies have examined a number of different methods that can be used to identify similar sentences. Some of these methods use shallow techniques to detect the similarities in sentences (e.g., word or n-gram overlap), whereas other methods use a deep approach to examining the syntactic or semantic similarities. The latent semantic analysis (LSA) technique can be used to estimate both the similarity of word matching and semantic structures. Accordingly, the problem of synonymy is avoided [16], [17].

Spoken documents are typically scanned and split into sentences throughout the preparation process, and then term-by-sentence matrices (TSMs) are ultimately created. One of the payoffs of using LSA is that it reduces dimensionality and thus results in quicker clustering. When the matrix is prepared, it is subjected to singular value decomposition (SVD) (Figure 1) [16]. The SVD formula can be stated as follows:

$$A = L_{EV} * S * R^{T}_{EV}$$

Any rectangular matrix A (i.e., a TSM matrix) with order t×s is decomposed into three matrices (L_{EV}, S, R^{T}_{EV}). The matrix L_{EV} contains the left eigenvectors of A and describes the relationship between terms (rows) and sentences (columns), or it refer to a term-to-concept similarity matrix resulting from the equation $L_{EV} = A^{T}A$. The matrix S is an m × m diagonal matrix with the entries sorted in decreasing order. The entries of the S matrix are the singular values (eigenvalue), and the S matrix describes the relative strengths of each concept. The R^{T}_{EV} matrix, which is defined by the equation $R^{T}_{EV} = A\,A^{T}$, contains the left eigenvectors of A, and this matrix refers to a sentence-to-concept similarity matrix [16].

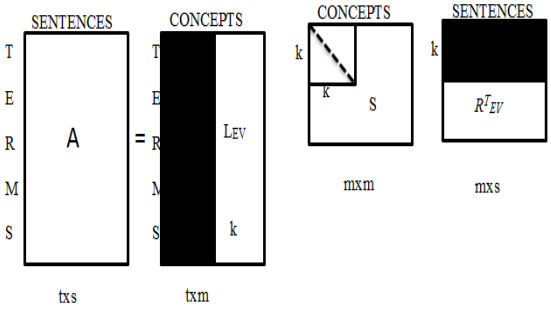


Fig. 1. Singular value decomposition (SVD)

The functionality of LSA will be explained using an example from the term similarity calculation. Consider Table I, which consists of four sentences from technical reports.

Table I. Example includes four sentences

S1	Shipment of gold damaged in a fire
S2	Delivery of silver arrived in a silver truck
S3	Shipment of gold arrived in a truck
S4	Gold silver truck

1. The TSM (Table II) is constructed as follows:

Table II. Term-by-sentence matrix (TSM)

	S1	S2	S3	S4
A	1	1	1	0
arrived	0	1	1	0
damaged	1	0	0	0
delivery	0	1	0	0
Fire	1	0	0	0
Gold	1	0	1	1
In	1	1	1	0
of	1	1	1	0
shipment	1	0	1	0
silver	0	2	0	1
truck	0	1	1	1

2. SVD is used to decompose the A Matrix into three matrices.

L_{EV} =

0.3966 -0.1282 -0.2349 0.0941
0.2860 0.1507 -0.0700 0.5212
0.1106 -0.2790 -0.1649 -0.4271
0.1523 0.2650 -0.2984 -0.0565
0.1106 -0.2790 -0.1649 -0.4271
0.3012 -0.2918 0.6468 -0.2252
0.3966 -0.1282 -0.2349 0.0941
0.3966 -0.1282 -0.2349 0.0941
0.2443 -0.3932 0.0635 0.1507
0.3615 0.6315 -0.0134 -0.4890
0.3428 0.2522 0.5134 0.1453

S =

4.2055 0.0000 0.0000 0.0000
0.0000 2.4155 0.0000 0.0000
0.0000 0.0000 1.4021 0.0000
0.0000 0.0000 0.0000 1.2302

R_{EV} =

0.4652 -0.6738 -0.2312 -0.5254
0.6406 0.6401 -0.4184 -0.0696
0.5622 -0.2760 0.3202 0.7108
0.2391 0.2450 0.8179 -0.4624

R^{T}_{EV} =

0.4652 0.6406 0.5622 0.2391
-0.6738 0.6401 -0.2760 0.2450
-0.2312 -0.4184 0.3202 0.8179
-0.5254 -0.0696 0.7108 -0.4624

The rank (r) of a matrix is the smaller of the number of linear independent rows and columns. SVD is used to reduce the rank and thereby the file size of the text. A reduced-rank SVD is performed on the matrix, in which the k largest singular values are retained, and the remainder is set to 0. The resulting representation is the best k-dimensional approximation of the original matrix in the least-squares sense [16]. Each sentence and term is now represented as a k-dimensional vector in the space derived by the SVD. In most applications the dimensionality k is much smaller than the number of terms in the TSM. In the above example, SVD ranks the concepts by importance for the text. By reducing the rank to 2, only the first two concepts are kept. Thus, the ranking matrices for the example are:

L'_{EV} =

0.3966 -0.1282
0.2860 0.1507
0.1106 -0.2790
0.1523 0.2650
0.1106 -0.2790
0.3012 -0.2918
0.3966 -0.1282
0.3966 -0.1282
0.2443 -0.3932
0.3615 0.6315
0.3428 0.2522

S' =

4.2055 0.0000
0.0000 2.4155

R'^{T}_{EV} =

0.4652 -0.6738
0.6406 0.6401
0.5622 -0.2760
0.2391 0.2450

Basically, to compute the similarity between the sentences the ranking matrix of R'^{T}_{EV} is used as input for the cosine distance equation. The cosine distance is a very popular way to measure the similarity and to compute the distance between any two sentences. Given two vectors of attributes, A and B, the cosine similarity, θ, is represented using a dot product as:

$$\text{Similarity} = \cos(\theta) = \frac{A.B}{\|A\|\|B\|} \dots (1)$$

To calculate cosine similarities for the example, the R'^{T}_{EV} matrix was used to calculate cosine similarities for each sentence as follows:

sim(Si, Sj) = (Si• Sj) / (|Si| |Sj|)

In our example, the similarity for S1 is calculated as:

sim(S1, S2) = (S1• S2) / (|S1| |S2|)
sim(S1, S3) = (S1• S3) / (|S1| |S3|)
sim(S1, S4) = (S1• S4) / (|S1| |S4|)

$$sim(S1,S2) = \frac{\left((0.4652 * 0.6406) + (-0.6738 * 0.6401)\right)}{\sqrt{\left((0.4652)^2 + (-0.6738)^2\right)} * \sqrt{\left((0.6406)^2 + (0.6401)^2\right)}} = -0.1797$$

$$sim(S1,S3) = \frac{\left((0.4652 * 0.5622) + (-0.6738 * -0.2760)\right)}{\sqrt{\left((0.4652)^2 + (-0.6738)^2\right)} * \sqrt{\left((0.5622)^2 + (-0.2760)^2\right)}} = 0.8727$$

$$sim(S1,S4) = \frac{\left((0.4652 * 0.2391) + (-0.6738 * 0.2450)\right)}{\sqrt{\left((0.4652)^2 + (-0.6738)^2\right)} * \sqrt{\left((0.2391)^2 + (0.2450)^2\right)}} = -0.1921$$

S3 returns the highest value; pair S1 with S3. The same method then is used to compute the similarity for S2, S3, and S3, S4. Consequently, similar sentences (cosine distance > threshold) are placed together to create a new sentence cluster. Then, a new matrix is created from this cluster and from the rest of the sentences. After applying SVD, all sentences are compared in a pairwise manner. This process is repeated until the distance of the similarity between the document sentences is larger than the previously indicated threshold.

IV. Proposed System

Previously developed systems for identifying news story boundaries depend on the dictation form of words. In contrast, the proposed framework uses the pronounced form. Table III shows some examples of the differences between the dictation and pronounced forms for some popular nouns.

TABLE III. EXAMPLES OF THE DIFFERENCES BETWEEN THE DICTATION AND PRONOUNCED FORMS OF SOME POPULAR NOUNS

Dictation form	Pronounced form with syllables
Abdul Rahman Abdulrahman Abdurrahman Abdulrrahman	ab/dur/rah/ma/n abdurrahman
Yassin,Yasin Yassen, Yasen Yasain, Yassain	Yas/si/n Yassin
Mohammed Mohamed	Moham/ma/d

Mohammad Mohamad Mohamat	Mohammad
Noor, Nour Nur, Nor	No/or Noor

The pronounced forms are more difficult in writing than in reading, as can be seen in the following examples:

1. Most of the sun letters or solar letters (t, v, d, r, z, s, l, and n) can be written with or without a duplicate letter, such as "s" in Yasain or "ss" in Yassain (both are correct). Duplicate consonants in popular nouns are considered to be a common diacritic, with the first being a consonant and the second a vowel [18], [19].

2. Dummy letters have no relation to neighboring letters and no correspondence to pronunciation; in other words, they are empty letters that have no sound (e.g., /h/ in Sarah, Fatimah, John, and Johnny) [8], [9].

3. Auxiliary letters with another letter constitute a diphthong (i.e., two letters combined to represent a single phoneme). These may be further categorized as a standard single-letter representation that uses another letter, as with "oo, ou, u, o in noor, nour, nur, and nor." These are irregular in dictation form. Table IV shows some examples of diphthongs and other ambiguous sounds [8], [9].

TABLE IV. THE DIFFICULTIES IN WRITING POPULAR NOUNS

Combination sound	Example
ai, ay, ei, y	Maitham, Maytham, Meitham, Mytham
oo, ou, o, u	Noor, Nour, Nor, Nur Fong, Foong Choy, Chooy
dh,z	Nadhem, Nazem
s, z	Asman, Azman
ee, ei	Swee, Swei
(ss,s), (dd, d), (mm, m), (rr, r), (tt, t), (vv,v), (zz,z), (ll, l), (nn, n)	Yassain, Yasain Aladdin, Aladin Mohammed, Mohamed Abdurrahman, Abdulrahman Abdultawab, Abduttwab Razzaq, Razaq, Razzak, Razak Abudllah, Abdulah Alnoor, Annoor

Avoiding the problem of writing popular nouns in different ways and thus reducing their impact on story boundary identification involves writing them in generalized and unified ways. This process requires use of an edit distance algorithm (Figure 2). This algorithm controls weights for the characters added and deleted and for the sun letters and dummy letters that are written but not pronounced in popular nouns.

The new system proposed herein proceeds in six stages:

Stage 1: Decode the spoken broadcast news to text using the sphinx 3 ASR system.
Stage 2: Use the maximum a posteriori (MAP) and maximum likelihood linear regression (MLLR) algorithms to improve the ASR acoustic model.
Stage 3: Apply the part of speech tagger (POS) to tag each word with its corresponding part of speech.
Stage 4: Use the generalized noun algorithm to unify the popular nouns (see section V).
Stage 5: Apply preparation processing (see section VI).
Stage 6: Identify the story boundaries using the LSA algorithm (see section III).

V. GENERALIZED NOUN ALGORITHM

The noun unification approach depends on phonetics (i.e., on the pronunciation of popular nouns rather than on the written form). The pronounced form of a word is based on the principle that "only the pronounced sounds are written down, even if they have no corresponding letters in dictation form. Also, what is not pronounced is left unprinted, even if it has a corresponding letter in dictation form"[1][8], [9]. Accordingly, some letters are either inserted or deleted in the pronounced form. There are many reasons for un-standard popular nouns, including the following:

1. The vowel combination makes it more difficult to find one form for the same noun.
2. Sun letters may or may not be duplicated.

[1] Alabbas, pp 5

```
Procedure LevenshteinDistance(S, T)
{
        Str1 ←Char( S) // Split S to array of characters
        Str2 ←Char( T) // Split T to array of characters
        m ←ArrayLen( Str1) // m=length of Str1 array
        n←ArrayLen( Str2) // n=length of Str2 array
        D[m,n] ← 0 // set initial values to Distance matrix D

         For (i←1 to m) // iterates until all char are validated

                D[i,0] ← i

         For (j←1 to n) // iterates until all words are validated

                D[0,j] ← j

        For (i←1 to m)
                For (i←1 to m)
                {
                 if Str1[i] = Str1[j] then
                        D[i, j] := D[i-1, j-1] // no operation required
                 else
                        {
                        if ( Str1[i] and Str1[j] == vowel) then weight=0.3; // substitution vowel letter
                        else // sound that have some relation like (“z/s”, “d/t”)
                         if ( relation(Str1[i], Str1[j] )==true) then weight=0.5;
                        else
                         weight=1;
                         D[i, j] := minimum
                         (
                 D[i-1, j] + 1, // a deletion
                 D[i, j-1] + 1, // an insertion
                 D[i-1, j-1] +weight // a substitution
                 )
                        }
                 }

}
```

Fig. 2. Edit distance algorithm

3. No two phones are exactly identical; within the same language, people pronounce things differently, and between different languages, no two sounds are ever exactly identical.
4. The distinction between vowels and consonants is not always clear cut, and there is a fuzzy boundary region between them in both human pronunciation and automatic recognition.

5. Some sounds are silent (dummy sounds) and are written in dictation form but are not pronounced [3].
6. Some foreign letters can be pronounced and written down using different letters. For instance, consider “Nazem” and “Nadhem:” The letters “z” and “dh” are used to represent the same name.

Converting popular nouns in a document to the general form using the generalized noun algorithm proceeds as follows:

```
Procedure Generalized Noun (Doc)
{
W ← Split (Doc , “ ”) // splits the text on every space
For (i←0 to N) // iterates until all words are validated
{ // check is W[i] noun or not using Part-of-Speech (POS) Tagger
If POS(W[i])= Noun then
        {
                //Find vowel combination and generalized it
            Loop
                Case (Ch← W[i]) :
                        // Group A include (vowel+h letter)
                Group A: replace Ch with μ;
                 // Group B include [ei] sound
                 Group B: replace Ch with β;
                 // Group C include (sun letteers)
                 Group C: replace Ch with λ;
                 // Group D include [ai] sound
                 Group D: replace Ch with Ω;

                                ...
        }  // End case }  // End Loop }  // End if
        } }
```

Poplar Nouns (PN)	Unified (PN)	Levenshtein Distance (LD)		Remark
		Before Unified	After UniFied	
Zainab Zaynab Zeyab	Zβab	LD(Zainab,Zaynab)=1 LD(Zainab,Zeynab)=2 LD(Zaynab,Zeynab)=1	LD(Zβab, Zβab)=0	β refers to [ei] sound group {ai,ay,ey,ei,ea}
Razzaq Razzak Razak Razaq	Raλaq Raλak	LD(Razzaq,Razzak)=1 LD(Razzaq,Razak)=2 LD(Razzaq,Razaq)=1 LD(Razzak,Razak)=1 LD(Razzak,Razaq)=2 LD(Razak,Razaq)=1	LD(Raλaq, Raλaq)=0 LD(Raλak, Raλak)=0 LD(Raλaq, Raλak)=0.5	λ refers to duplicated letters group "sun letters" {t, v, d, b, r, z, s, l and n}
Mohammed Mohamed Mohammad Mohamad Mohamat Mohammet	Mohλad Mohλed Mohλat Mohλet	LD(Mohammed,Mohamed)=1 LD(Mohammed,Mohammad)=1 LD(Mohammed,Mohamad)=2 LD(Mohammed,Mohamat)=3 LD(Mohammed,Mohammet)=1 LD(Mohamed,Mohammad)=2 LD(Mohamed,Mohamad)=1 LD(Mohamed,Mohamat)=2 LD(Mohamed,Mohammet)=2 LD(Mohammad,Mohamad)=1 LD(Mohammad,Mohamat)=2 LD(Mohammad,Mohammat)=1 LD(Mohamad,Mohamat)=1 LD(Mohamad,Mohammet)=3 LD(Mohamat,Mohammet)=3	LD(Mohλad, Mohλad)=0 LD(Mohλad, Mohλed)=0.5 LD Mohλad, Mohλat)=0.5 LD(Mohλad, Mohλet)=0.8 LD(Mohλed, Mohλed)=0 LD(Mohλed, Mohλat)=0.8 LD(Mohλed, Mohλet)=0.5 LD(Mohλat, Mohλat)=0 LD(Mohλat, Mohλet)=0.8 LD(Mohλet, Mohλet)=0	λ refers to duplicated letters group "sun letters" {t, v, d, b, r, z, s, l and n}
Johnny Johnnie Jonny Jonnie	JμλΩ	LD(Johnny,Johnnie)=2 LD(Johnny,Jonny)=1 LD(Johnny,Jonnie)=3 LD(Johnnie,Jonny)=3 LD(Johnnie,Jonnie)=1 LD(Jonny,Jonnie)=2	LD(JμλΩ, JμλΩ)=0	Ω refers to [ai] sound group {ie,y,uy}

TABLE V. MEASUREMENT OF THE SIMILARITY BETWEEN TWO STRINGS USING LEVENSHTEIN DISTANCE

Table V illustrates the measurement of the similarity between two strings using Levenshtein distance for several examples. The number of transformations (deletions, insertions, or substitutions) required to transform one string into another were measured before and after the generalized noun algorithm was applied.

VI. PREPROCESSING STAGE

The preprocessing module performs tagging, removal of stopping words, stemming, feature selection, and TSM creation. During the tagging process, each word is tagged with its corresponding POS. For example, the sentence "ali pergi ke sekolah" (Ali goes to school) is tagged as "ali/noun pergi/verb ke/preposition sekolah/noun." In this study we used the Qtag POS tagger. The next step is to remove stopping words, which removes all of the frequent and common words that do not carry important information. This step reduces the size of the spoken document. Such words include auxiliary verbs and prepositions (e.g., adalah/(is, are), akan/will, was/ialah, ke/to, pada/at). The removal of such words helps to improve the quality of the story boundary identification results by retaining only the words that contain significant information. This step can be performed using the stopping word list, which includes 1312 common Malay stopping words.

Stemming refers to reducing morphological variants of words to their stem, base, or root form, and it is used to improve the effectiveness of information retrieval (IR). The effect of stemming depends on the nature of the language vocabulary, and in some cases stemming may degrade retrieval performance [20]. Thus, a stemmer can improve the effectiveness of IR for some text corpora more than others [16], [17], [20]. In the system proposed here, an affixation stemmer for the Malay dataset was used. The words permakanan (diet) and makanannya (his/her food), for example, contain the base word "makan," and the common stem of the various forms of the word was weighted using the tf-idf (term frequency-inverse document frequency) weighting approach in the term-by-document matrix (TDM). The use of the stemming algorithm can increase retrieval performance by reducing morphological

variants of words and the time required for processing; at the same time, use of the roots of the words increases the similarity probability between the words in the clustering module.

After stemming is completed, feature selection is performed. One of the major challenges facing artificial intelligence applications is how to reduce the number of high dimensional data spaces. Dimensionality reduction is the process of reducing the number of random variables (words here) under consideration (for instance, retaining the significant words or the high frequency words) [16]. The efficiency of the relevant algorithms can be improved by decreasing the dimensionality of the size of the effective vocabulary and data spaces. In such cases, feature selection can be applied. Feature selection chooses an effective subset from a huge set of features. In this study, we used the open source library "weka" to select the useful features, and only the selected keywords (words) were used in the subsequent building of the TSM.

In the TSM, each row defines the terms contained in a sentence. Each cell entry contains the frequencies of occurrence of a term in a sentence. This TSM can be used to calculate the similarity between terms using story boundary identification methods. To illustrate, suppose we have the following set of five sentences:

S1 = w1 w2 w3 w4 w5 w6
S2 = w7 w2 w3 w4 w5 w6
S3 = w6 w8 w4 w5
S4 = w1 w9 w10 w4 w6
S5 = w10 w2 w2 w4 w11 w5

A data set can be represented by the TSM using the frequency weight matrix shown in Table VI.

TABLE VI. TERM-BY-SENTENCE MATRIX (TSM)

	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11
S1	1	1	1	1	1	1	0	0	0	0	0
S2	0	1	1	1	1	1	1	0	0	0	0
S3	0	0	0	1	1	1	0	1	0	0	0
S4	1	0	0	1	0	1	0	0	1	1	0
S5	0	2	0	1	1	0	0	0	0	1	1

VII. EXPERIMENTS AND RESULTS

To demonstrate the performance of the proposed algorithms, a transcript produced manually from spoken broadcast news was used [21] to identify the story boundaries. The databases used for this are called the mass-news corpus, and they consist of Malay broadcast news documents that were collected at Universiti Sains Malaysia as the output of the Malay ASR system [21]. The news stories used for this evaluation were collected in March 2011. The data set includes ~25 hours of transcribed speech. The ASR system was trained using a ~15 hour portion of the database, and the test sets included the remaining ~10 hours. The broadcast news stories included multiple speakers and recording in noisy environments. None of the test sets overlapped with the ASR training set. Table VII shows the Malay data source details that were used in this study.

TABLE VII. DATA SOURCE DETAILS

Size of language model	150 MB
Size of dictionary	1.74 MB
Number of news shows	18
Number of news stories in all news shows	379
Number of sentences in news database	4698
Number of words in the news	81116
Number of popular nouns in the news	39698 (49%)
Word error rate before adaptation	34.5%
Word error rate after adaptation	33.9%
Story length	Around 1 to 167
The rate of the audio signal extract	10 ms

Errors resulting from the process of story boundary identification were measured using the F-score, precision, and accuracy. To evaluate[2] the effectiveness of the story boundary identification module, we tested it using Malay spoken documents that contained ~380 stories in different domains (e.g., politics, economics, sports, local news, and international news). We evaluated two corpora. The first corpus was a gold standard file (GSF) corpus that represents the manual transcription of the Malay broadcast news. The second corpus was the ASR result (Hypothesis Result (HR)) for the Malay broadcast news. The GSF corpus was segmented into stories by human experts. In this experiment different k-dimensional clustering spaces were built where k ϵ [32, 50, 80, 100, 125, 150, 200]. This paper reports only the best results.

Table VIII shows the results of the story boundary identification for LSA with and without the noun unification process.

[2]The tools that used in this study
Java, Python, Apache Lucene package (java package) was used for compute F-measure and the other Measurement, Sphinx 3 as ASR system, WEKA package for feature selection, SPSS for Wilcoxon signed-ranks test, Qtag tool for part of speech tagger, an affixation stemmer tool, Jama package for Matrix computations.

TABLE VIII. STORY BOUNDARY IDENTIFICATION MODULE PERFORMANCE

	LSA			
	Without		With	
	GSF	HR	GSF	HR
Precision	0.759	0.680	0.895	0.814
Recall	0.617	0.559	0.914	0.769
F-Measure	0.681	0.613	0.904	0.791

The results of the story boundary identification algorithms were evaluated statistically using the Wilcoxon signed-ranks test. By applying the same statistical significance test, the results of the proposed algorithm were compared statistically with the results of the baseline algorithm (i.e., LSA without the noun unification process). The proposed system using the popular noun unification algorithm achieved an F-measure of 0.791, whereas the value was 0.613 for the baseline system when tested on the same set of Malay broadcast news stories.

VIII. CONCLUSIONS

Identifying broadcast news story boundaries plays an important role in many natural language processing applications, such as topic identification and story classification. The proposed system uses the pronounced forms to identify story boundaries based on popular noun unification. LSA is commonly used in clustering methods because of its excellent performance and because it is based on deep semantic rather than shallow principles. In this study, the LSA algorithm with popular noun unification achieved a better result than the general LSA algorithm in identifying news story boundaries for the same test set. The LSA algorithm with popular noun unification achieved an overall F-measure of 0.791 versus 0.613 for the general LSA algorithm when identifying news story boundaries for the same test set.

We predict that further work by adding ASR confidence measure to distinguish between correct and incorrect words in ASR result before any processing, e.g. story boundaries identification. In future work, we will apply this algorithm for English language.

ACKNOWLEDGMENT

ZAK owes her deepest gratitude to USM for its support of her PhD research. She also would like to extend thanks to Basra University for their helpful support.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern information retrieval vol. 463: ACM press New York, 1999.
[2] C. Chelba, et al., "Retrieval and Browsing of Spoken Content," IEEE signal Processing Magazine, vol. 25, pp. 39-49, 2008.
[3] M. Ostendorf, et al., "Speech Segmentation and its Impact on Spoken Document Processing," 2007.
[4] M. Abbas, et al., "Evaluation of Topic Identification Methods on Arabic Corpora," Journal of Digital Information Management, vol. 9, pp. 185-192, October, 2011.
[5] D. Li, et al., "Initial Experiments on Automatic Story Segmentation in Chinese Spoken Documents Using Lexical Cohesion of Extracted Named Entities" in ISCSLP, 2006, pp. 693-703.
[6] M.-m. LU, et al., "Multi-Modal Feature Integration for Story Boundary Detection in Broadcast News," IEEE, pp. 420-425, 2010.
[7] W. Hsu, et al., "Discovery and fusion of salient multimodal features toward news story segmentation," in Proceedings of SPIE, 2004, pp. 244-258.
[8] Wikipedia. (2013). Silent letter - http://en.wikipedia.org/wiki/Silent_letter.
[9] A. a. Galina. (2007-2013). English Vowel Sounds- http://usefulenglish.ru/phonetics/english-vowel-sounds.
[10] M. A. Hearst, "TextTiling: Segmenting text into multi-paragraph subtopic passages," Computational linguistics, vol. 23, pp. 33-64, 1997.
[11] N. Stokes, et al., "SeLeCT: a lexical cohesion based news story segmentation system," AI COMMUNICATIONS, vol. 17, pp. 3-12, 2004.
[12] A. Rosenberg and J. Hirschberg, "Story segmentation of broadcast news in English, Mandarin and Arabic," in Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, Stroudsburg, PA, USA, 2006.
[13] L. Xie, "Discovering salient prosodic cues and their interactions for automatic story segmentation in Mandarin broadcast news," Multimedia Systems, vol. 14, pp. 237-253, 2008.
[14] S. Banerjee and A. I. Rudnicky, "A TextTiling based approach to topic boundary detection in meetings," 2006.
[15] C. H. Wu and C. H. Hsieh, "Story segmentation and topic classification of broadcast news via a topic-based segmental model and a genetic algorithm," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 17, pp. 1612-1623, 2009.
[16] J. Geiß, "Latent semantic sentence clustering for multi-document summarization," Ph.D, University of Cambridge, 2011.
[17] Z. A. Khalaf and T. T. Ping, "Unsupervised Identification of Story Boundaries in Malay Spoken Broadcast News," Journal Of Emerging Technologies In Web Intelligence, vol. 5, pp. 28-34, 2013.
[18] Z. A. Khalaf, et al., "BASRAH: Arabic Verses Meters Identification System," in IALP, Penang-Malaysia, 2011, pp. 41-44.
[19] M. Alabbas, et al., "BASRAH: an automatic system to identify the meter of Arabic poetry," Natural Language Engineering-Cambridge University Press 2012, pp. 1-19, 2012.
[20] W. B. Frakes and R. Baeza-Yates, "CHAPTER 8: STEMMING ALGORITHMS Information Retrieval: Data Structures & Algorithms," ed. Englewood Cliffs, NJ: Prentice Hall, 1992, pp. 131-160.
[21] T. Tien-Ping, et al., "Mass: A Malay Language LVCSR Corpus Resource," Cocosda'09. Urumqi, China, 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 585–590

Fingerprinting System for Still Images Based on the Use of a Holographic Transform Domain

Valery Korzhik
(Member of IEEE)
State University
of Telecommunications
Saint-Petersburg, Russia
Email: val-korzhik@yandex.ru

Guilermo Morales-Luna
Computer Science
CINVESTAV-IPN
Mexico City, Mexico
Email: gmorales@cs.cinvestav.mx

Alexander Kochkarev and Ivan Shevchuk
State University of Telecommunications
Saint-Petersburg, Russia
Email: kochkareff@mail.ru, johan92@yandex.ru

***Abstract*—We consider the watermarking method based on a holographic transform domain image proposed by A. Bruckstein. Our testing showed that it is resistant not against all possible attacks declared by his inventor, under the condition of a very high image quality just after WM embedding. Only a small part among 120 bits embedding into the image has an acceptable error probability after extraction if some attacks hold. Therefore we propose to modify this system for fingerprinting where only fixed bits are embedded into the most reliable places of the frequency mask. Systematic linear binary codes with large minimal code distance are used in order to correct errors. Simulation showed that such system provides sufficiently reliable tracing "traitors" under the most types of attacks subjected to remove WM, while keeping a good quality of the image just after embedding.**

***Index Terms*—Watermarking, image processing, error correction codes, tracing traitors**

I. Introduction

DIGITAL watermarks effectively can be used for copyright protection of still images [1], [2], [3], [4]. However, intruders, the so-called "pirates", try to copy and spread these products illegally, and they attempt to remove the WM by performing different (sometimes very sophisticated) transforms over the watermarked products which, not impairing the product itself, should make impossible to extract them. In [3] there has been proposed an approach for watermarking insertion that is invariant to several transforms as rotation, scale and translation. But the use of the log-polar transforms results (confirmed at our experiments) to significant corruption of the cover images after WM embedding. Only a very restricted number of possible transforms are considered in [4]. A good robustness to practically all possible transforms has been obtained in [5] but unfortunately it works only for o-bit watermark. An extension of this method to multiple-bit watermark was presented in [6] but without the use of error correcting codes. Thus it can be concluded that although there were many proposals in the design of WM systems resistant to different attacks, this problem is so far not solved completely.

In [7], a WM system based on a "holographic" transform domain has been proposed, where the embedding procedure is performed in the area of the Fourier amplitude and then the message can be extracted even from cropped WM-ed image. This is why this method was called *holographic*, it is a metaphor of the physical hologram where the whole can be recovered from its small part.

The authors of [7] declare that this method allows to embed up to 120 message bits and to extract them correctly using an informed decoder even after several attacks as cropping, JPEG compression, changing of contrast and some other combinations of them. The embedding procedure is performed then as follows:

$$I^W = \mathcal{F}^{-1}\left(W_b \cdot \mathcal{F}\left(I\right)\right) \tag{1}$$

where $I = \left(I(x,y)\right)_{(x,y)}$ is a grey-level (8 bit) image in an (x,y)-pixel area, $W_b = \left(W_b(u,v)\right)_{(u,v)}$ is an embedding mask,

$$W_b(u,v) = 1 + (-1)^b \varepsilon \text{ whenever } (u,v) \in S^b_{ij}, \tag{2}$$

with $\left(S^0_{ij}\right)_{ij}$, $\left(S^1_{ij}\right)_{ij}$ being some collections of selected areas, corresponding to the chosen embedding mask in the frequency area for the (i,j)-th message bit 0 or 1, respectively, ε is a depth of embedding, $\mathcal{F}$, $\mathcal{F}^{-1}$ are, respectively, the direct and the inverse Fourier transforms, and $I^W = \left(I^W(x,y)\right)_{(x,y)}$ is the resulting watermarked image. An embedding mask can be chosen in different manners. In the paper [7] it is used the so called "equally radius" geometry shown on Fig. 1. For such mask it is possible to embed 120 message bits in the whole image. The extraction of each of the (i,j)-bits is performed by the following rule optimal in additive Gaussian noise attack channel:

$$b_{ij} = \frac{1}{2}\left[1 - \text{Sign}\left(B^1_{ij} - B^0_{ij}\right)\right] \tag{3}$$

where

$$B^b_{ij} = \sum_{(i,j)\in S^b_{ij}} \Re\left(\overline{q_{ij}}\, s_{ij}\right) \quad , \quad b \in \{0,1\},$$

$\left(s_{ij}\right)_{(i,j)} = \mathcal{F}(I)$ is the array of complex values obtained as the Fourier transform of the original image I, $\left(q_{ij}\right)_{(i,j)} = \mathcal{F}(I^W)$ is the array of complex values obtained as the Fourier transform of the watermarked image I^W, $\Re$ is the "real part" operator and the overline denotes complex conjugation.

Since the knowledge of original image $\left(I(x,y)\right)_{(x,y)}$ is necessary for the extraction procedure, this method is called

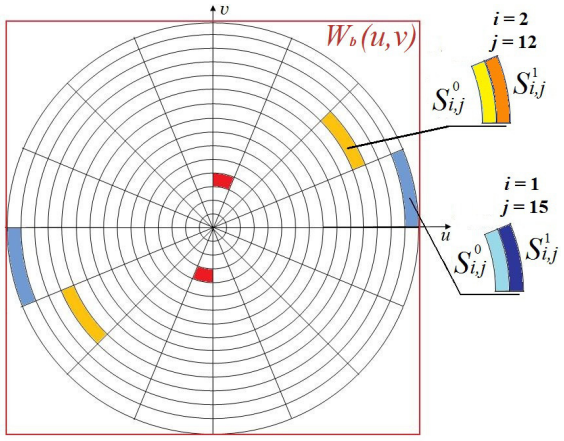


Fig. 1. Equally radius geometry embedding mask.

an *informed decoder*. Moreover, if the WM-ed image has suffered some attack, say cropping of windows or the removal of some rows and columns, it is necessary to know the changed version of the original image after such attacks. This means that the decoder should know the exact place of the window or the locations of rows and columns removed after the attack. Such problem is called the *registration problem*. Sometimes it can be solved very easily (because both attacked and original image are available for the decoder) but sometimes it requires a solution of an additional problem, known as the *registration one*. But we leave the registration problem outside our investigation.

II. Abounding on testing a commonly used method

Let us present the tests realized in accordance with the method proposed at [7].

The quality of the watermarked image is determined by the depth of embedding ε. In the Fig. 2, an original image and its watermarked images with $\varepsilon = 0.05$ and $\varepsilon = 0.2$ are presented.

We can see that the quality of the WM-ed image is still acceptable for $\varepsilon = 0.05$ but indeed unacceptable if $\varepsilon > 0.2$. (This claim has been confirmed after a testing of many typical images on computer screens.)

The results of message extraction for different images are presented at Table I, once a given image has been attacked through several transforms, while keeping good image quality after the embedding and attacks.

This testing shows that although a cropping of small "windows" gives excellent results as well as JPEG compression with quality factor $Q \geq 60\%$, further decreasing of the window's sizes and a quality of the JPEG compression results in a degradation of the WM system as well as an addition of a Gaussian noise with variance larger 25. Thus the claim [7] that such WM system satisfies the required conditions for being resistant against any attacks is only partly correct. (It is true only for some specific images). But in order to maintain a good idea proposal in [7] regarding the holographic transform domain and portioning of decision bit area into two subareas

(a)

(b)

(c)

Fig. 2. Image before and after watermarking. (a) Original image, (b) WM-ed image with $\varepsilon = 0.05$, (c) WM-ed image with $\varepsilon = 0.2$.

in line with the decoding rule (3) we suggest to modify WM system in some manner to adopt it in a modified form.

TABLE I
THE RESULTS OF ERROR PROBABILITY IN EXTRACTION PROCEDURE AFTER DIFFERENT ATTACKS.

Name of attack	PC
Cropping of window 200×200 pixels	4
Cropping of window 170×170 pixels	8
Saving in JPEG format with $Q = 60\%$	3
Saving in JPEG format with $Q = 50\%$	6
Saving in JPEG format with $Q = 20\%$	25
Saving in JPEG format with $Q = 10\%$	30
Addition of Gaussian noise with a variance $d = 25$	15

PC: Percent of corrupted bits on average of several images.

A description of the extraction procedure results, by simulation after different attacks, within this new approach and and the original method are presented in Section III and IV, respectively.

III. DESCRIPTION OF THE MODIFIED WM SYSTEM

Firstly, the results of our simulations, which show the probabilities of errors after extraction of bits on different places into the frequency mask and after different attacks, are presented in the Tables II-V.

By observing these tables, we can conclude that there are some bit locations where the probabilities of errors are unacceptable even if we would use some error correction codes, while there are some other bit locations where the probabilities of errors approach to zero. Then the following natural idea arises – let us embed message bits only in such "cells" of the mask where there appears a moderate number of errors.

We could try of course to execute a diversity concept. This means that the same bit is embedded in several cells. But experiments show that a soft decoding occurs useless in this case because the values q_{ij} in eq. (3) are falsely increased for some cells after the JPEG transforms and make worse the result of decision in a comparison with hard decision. One can use the hard (majority) decoding rule but it requires a large multiplicity of diversity and to find the gain to remove bits from "bad" cells which are providing the negligible effect. The amount of bits which have the acceptable probability of error is about 64 and they are displayed at columns 2–9 at Tables III–V. This value is not sufficient in order to embed reasonable information but it may be enough for a scenario of fingerprinting.

Let us consider a situation in which the owner of some image sales it legally to a set of M users without a permission to distribute this product further outside of this buyer set. But some members of the set did illegal redistribution of the product. Fortunately, the owner has access to the illegally redistributed copies. The owner of the product wants to recognize who was the illegal distributor (the "pirate" in other words). It is worth to note that such digital fingerprinting (FP) has very important role in enabling an early-release HD movie window for VOD [8].

In order to solve this problem, the owner can proceed in the following manner: he embeds an unique bit string in every copy sold to legal users, he extracts the embedded WM (which is called usually the *fingerprint*) from illegally redistributed copy and trace the pirate. We propose to select unique strings of the length equal to the number of practically error-moderate bits (in our case it is 64). Let us denote by R the area consisting of the columns labelled 2-9 at each of the Tables II-V (emphasized at their displays). The other bits, displayed at columns 10-15 are free of embedding.

Since there may occur errors even among the specially selected 64 bits, it is reasonable to use error-correction codes.

First of all we consider the use of BCH codes of length 63 with a hard decoding on Hamming distance [9], namely the codes $(63, 7)$, $(63, 10)$ and $(63, 16)$. But since the probabilities of bit errors, even among the columns 2-9, depend (as it can be seen from Tables II-V) on the positions of these bits, it is reasonable to use a more effective maximum likelihood decoding algorithm, namely, let:

$$\tilde{j} = \arg\max_{j} \left[\prod_{i \in I(e_{j1})} P_i \cdot \prod_{i \in I(e_{j0})} (1 - P_i) \right] \qquad (4)$$

where $\tilde{j}$ is the number of codeword after decoding, P_i is the error probability at the i-th bit, $I(e_{j1})$ is the set of components with value one at the vector e_j (the *support* of e_j), $I(e_{j0})$ is the set of components with value zero at the vector e_j (the *null set* of e_j), and $e_j = u \oplus v_j$ where u is the received binary vector after demodulation by (3), and v_j is the j-th code word at the BCH code.

In the next Section we consider the results of simulation of the proposed approach after attacks by different transforms.

IV. RESULTS OF FINGERPRINTING SYSTEM SIMULATION AFTER DIFFERENT TRANSFORMS

We use the embedding according to the relations (1), (2) into the area R, the bit extraction by the rule (3) and the decoding of the code words of the BCH codes $(63, 7)$, $(63, 10)$ and $(63, 16)$, by the minimal Hemming distance and maximum likelihood algorithm (4). As image transforms we apply the following ones:

- cropping of windows;
- removal of rows and columns;
- JPEG compression with different quality;
- addition of Gaussian noise.

We selected 1000 grey scaled images from the bank of images [10] and each of these images was tested 10 times with randomly chosen bit embedding. In reality we try BCH codes of the length 63 for more variants on the number of information bits, other than 3, but we show now only those cases in order to justify that there is no sense to take $k > 10$, because it results in poor probability of correct decoding even after the use of an optimal decoding algorithm.

In the Fig. 3 a fingerprinted image and its cover image after different transforms are presented.

TABLE II
THE PROBABILITY (IN PERCENTS) OF THE (i, j)-TH BIT ERROR AFTER A JPEG TRANSFORM WITH QUALITY FACTOR $Q = 10\%$.

$i\backslash j$	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	10	11	12	13	14	15
1	3	1	4	3	7	21	39	34	32	42	48	38	52	46	47
2	3	2	5	16	19	31	44	35	49	47	49	41	56	44	38
3	2	0	4	12	19	36	29	45	42	42	45	56	50	46	44
4	3	0	1	8	6	15	25	40	43	50	55	48	38	47	46
5	2	2	2	5	10	15	32	35	41	48	51	43	48	48	39
6	2	3	4	7	21	28	43	53	44	45	50	44	57	51	45
7	0	1	4	15	27	36	46	36	45	42	53	44	50	45	53
8	0	1	1	5	8	28	35	40	41	40	38	47	44	51	50

TABLE III
THE PROBABILITY (IN PERCENTS) OF THE (i, j)-TH BIT ERROR AFTER A JPEG TRANSFORM WITH QUALITY FACTOR $Q = 20\%$.

$i\backslash j$	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	10	11	12	13	14	15
1	0	1	1	2	2	8	7	24	30	41	42	38	43	46	35
2	2	0	2	2	10	16	34	32	55	44	44	54	52	44	38
3	2	2	2	1	6	15	33	40	36	41	38	51	49	42	48
4	2	0	1	3	3	7	7	13	38	40	38	49	57	51	41
5	0	0	1	0	2	5	13	14	42	51	47	52	51	44	38
6	0	1	1	2	3	9	33	45	43	42	44	57	52	47	45
7	0	1	2	2	2	17	27	41	38	50	40	42	48	47	49
8	1	1	2	0	2	7	10	30	42	33	45	51	35	45	42

TABLE IV
THE PROBABILITY (IN PERCENTS) OF THE (i, j)-TH BIT ERROR AFTER CROPPING OF WINDOW WITH SIZE 200×200 PIXELS.

$i\backslash j$	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	10	11	12	13	14	15
1	43	9	2	0	1	0	0	1	0	0	0	0	1	1	3
2	38	1	2	0	0	0	0	0	0	0	0	0	0	1	0
3	43	13	0	1	1	0	1	0	1	0	0	0	1	0	0
4	32	3	1	1	0	1	2	3	2	2	5	5	6	3	5
5	46	2	1	1	1	1	1	1	1	1	2	2	3	2	3
6	25	3	1	1	1	0	1	0	0	2	0	2	2	1	0
7	23	2	1	1	1	0	1	0	0	0	0	1	0	0	0
8	45	3	1	1	0	0	1	0	0	0	1	0	0	0	2

TABLE V
THE PROBABILITY (IN PERCENTS) OF THE (i, j)-TH BIT ERROR AFTER AN ADDITION OF GAUSSIAN NOISE WITH VARIANCE $d = 25$.

$i\backslash j$	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	10	11	12	13	14	15
1	1	1	1	3	4	6	10	11	10	20	19	20	21	25	24
2	0	1	3	12	11	14	18	15	15	17	21	20	40	31	29
3	5	4	3	9	9	11	11	14	13	22	22	28	33	25	30
4	0	3	1	4	3	10	6	13	13	16	13	22	16	29	29
5	1	2	2	1	5	11	14	10	15	14	26	21	22	28	39
6	1	4	3	7	13	8	20	14	16	24	23	23	23	28	25
7	1	0	4	7	8	11	20	20	22	22	20	24	26	35	33
8	1	0	5	3	4	10	9	7	13	21	18	23	29	23	28

In all cases we assume that the original image is known during the extraction procedure. Sometimes this condition can be provided very easily, whereas sometimes it requires to solve an additional problem for the original image registration, in this last case we refer to cropping (where it is necessary to know the window) or to row and column removal (where it is necessary to know which of them have been removed).

It is worth to note that the pirates can remove rows or columns in two different ways.

Consider the first way. A pirate selects some rows (or columns) and changes them to another ones, which can be obtained by interpolation of neighboring lines. In this case the image size and places of another rows are not changed. Hence it is the case when it is not necessary to solve a problem of original image registration.

Another case arises where the pirate deletes the lines and then he shifts the remaining lines to make invisible the removal place. In this case, the image size and places of several lines

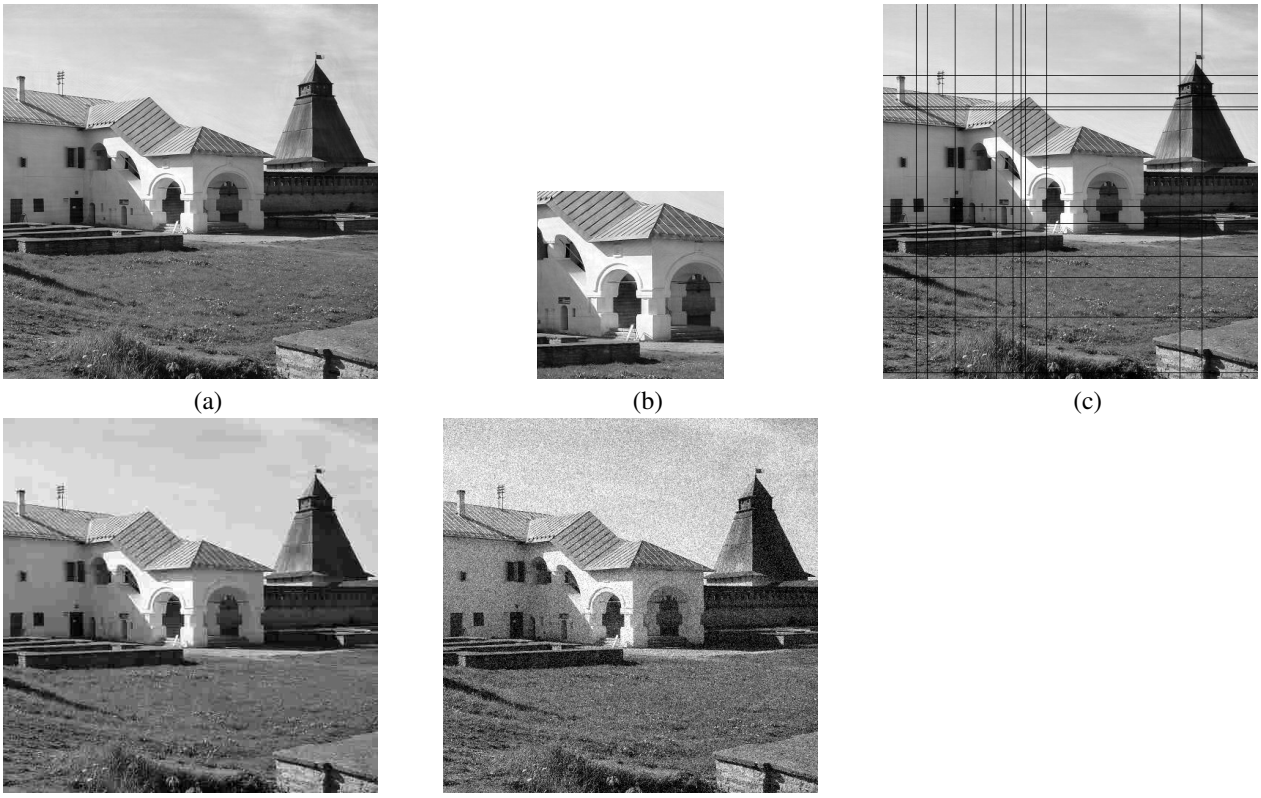


Fig. 3. Fingerprinted image (a) and the same image after different transforms: (b) cropping of window 200×200 pixels, (c) removal of rows and columns, (d) JPEG compression with factor $Q = 20$, (e) addition of Gaussian noise with $d = 25$.

are changed. Therefore for extraction procedure it is necessary to solve a registration problem.

The results of simulation (in terms of the incorrect decoding probabilities) depending on the type of code and different attack transforms are presented in Table VI.

Similar results for optimal decoding algorithm are shown in Table VII.

We note that for those attacks that can be easily recognized by a legal user (removal of rows and columns, cropping, addition of noise) the values of the symbol probabilities P_i, that are necessary for optimal decoding by the algorithm (4) can be taken from Tables IV–V, whereas it is hard to establish the quality factor Q used in the JPEG transform. The probabilities for the worst case ($Q = 20\%$) can be used because we proved that it results in the minimal probability of incorrect decoding on average.

From Tables VI–VII it can be seen that the maximum number of information bits k, that can still provide the acceptable probability of incorrect decoding after all attack transforms is 10 and the optimal decoding algorithm given by (4) is superior to the minimal Hamming distance algorithm.

V. Conclusion

Traitor tracing is a very important problem in the case of an early release of HD movie window for VOD. In the current paper we adopt a general idea to embed WM using the so called "holographic" concept [7] when the embedding procedure is performed in the Fourier domain. However we showed that such WM system is vulnerable to different image transforms which provide still a good image quality after them. Therefore we propose a modification of the WM system considered in [7] to a fingerprinting system, where it is sufficient to provide only a limited number of identification code words corresponding to different users that can be potential pirates. In a particular case we have selected 64 bits which survive after most of the transforms and we propose to execute a binary $(63, 10)$-BCH code to correct errors. We propose also to use a maximum likelihood decoding algorithm instead of the minimum Hamming distance algorithm since that is more effective

The simulation results showed that for many typical images the proposed fingerprinting scheme is resistant to such transforms as cropping, removal of rows and columns, JPEG compression and addition of Gaussian noise. Therefore we are rather sure that the proposed scheme can be recommended for practical applications to copyright protection within fingerprinting procedures.

But the problem of original image registration arises.

Sometimes it is easy to solve because the original images always are at the disposition of their owners. But sometimes it requires to know some parameters of transforms (as numbers

TABLE VI
THE PROBABILITIES OF INCORRECT DECODING BY MINIMUM HAMMING DISTANCE FOR DIFFERENT BCH CODES, DIFFERENT ATTACK TRANSFORMS AND DIFFERENT EMBEDDING DEPTHS ε.

BCH codes	(63, 7)	(63, 10)	(63, 16)	(63, 7)	(63, 10)	(63, 16)
(1)\(2)	0.05			0.1		
Saving in JPEG format with Q=20%	9.0×10^{-2}	1.6×10^{-1}	2.5×10^{-1}	2.3×10^{-2}	4.7×10^{-2}	7.9×10^{-2}
Saving in JPEG format with Q=30%	2.8×10^{-2}	5.7×10^{-2}	9.7×10^{-2}	5.7×10^{-3}	1.4×10^{-2}	2.5×10^{-2}
Saving in JPEG format with Q=60%	3.4×10^{-3}	6.6×10^{-3}	1.4×10^{-2}	1.2×10^{-3}	1.7×10^{-3}	3.8×10^{-3}
Cropping of window 200×200 pixels	1.8×10^{-2}	2.7×10^{-2}	3.6×10^{-2}	1.5×10^{-2}	2.0×10^{-2}	2.8×10^{-2}
Cropping of window 250×250 pixels	5.5×10^{-3}	8.3×10^{-3}	1.0×10^{-2}	4.1×10^{-3}	5.5×10^{-3}	7.9×10^{-3}
20 rows and 20 columns removal	2.7×10^{-2}	5.4×10^{-2}	1.1×10^{-1}	5.0×10^{-3}	1.2×10^{-2}	2.5×10^{-2}
Addition of Gaussian noise with $d = 25$	8.4×10^{-2}	1.4×10^{-1}	2.1×10^{-1}	1.3×10^{-2}	2.3×10^{-2}	3.9×10^{-2}

(1) Attack transform. (2) Embedding depth (ε).

TABLE VII
THE PROBABILITIES OF INCORRECT DECODING BY OPTIMAL DECODING ALGORITHM FOR DIFFERENT BCH CODES, DIFFERENT ATTACK TRANSFORMS AND DIFFERENT EMBEDDING DEPTHS ε.

BCH codes	(63, 7)	(63, 10)	(63, 16)	(63, 7)	(63, 10)	(63, 16)
(1)\(2)	0.05			0.1		
Saving in JPEG format with Q=20%	1.0×10^{-2}	1.4×10^{-1}	3.9×10^{-1}	1.5×10^{-3}	4.1×10^{-3}	1.0×10^{-2}
Saving in JPEG format with Q=30%	2.1×10^{-3}	3.8×10^{-3}	1.1×10^{-2}	1.0×10^{-4}	1.0×10^{-3}	3.5×10^{-3}
Saving in JPEG format with Q=60%	4.0×10^{-4}	9.0×10^{-4}	2.1×10^{-3}	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}
Cropping of window 200×200 pixels	8.5×10^{-3}	1.2×10^{-2}	2.2×10^{-3}	4.3×10^{-3}	9.1×10^{-3}	1.6×10^{-2}
Cropping of window 250×250 pixels	1.9×10^{-3}	3.8×10^{-3}	6.1×10^{-3}	2.1×10^{-3}	3.0×10^{-3}	4.7×10^{-3}
20 rows and 20 columns removal	2.8×10^{-3}	6.5×10^{-3}	1.4×10^{-2}	4.0×10^{-4}	1.3×10^{-3}	3.9×10^{-3}
Addition of Gaussian noise with $d = 25$	1.5×10^{-2}	3.1×10^{-2}	7.5×10^{-2}	2.0×10^{-3}	4.3×10^{-3}	1.2×10^{-2}

(1) Attack transform. (2) Embedding depth (ε).

of the removed rows and columns and their places) executed by pirates. This is still an open problem in general. Another problem is to change the equality radius geometry embedding mask (see Fig. 1) to another one in order to try to use the area with columns 10-15 (or maybe areas structured by another manner) to embed more than 10 bits with good enough probability of correct decoding. We are going to investigate these problems in the near future.

REFERENCES

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. Morgan Kaufman Publishers, 2002.

[2] M. Barni and F. Bartolini, *Watermarking systems engineering: enabling digital assets security and other applications*, ser. Signal processing and communications. Marcel Dekker, 2004. [Online]. Available: http://books.google.co.uk/books?id=DUuyektSYH0C

[3] J. Ó Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking," in *IEEE Int. Conf. on Image Processing ICIP1997*, 1997, pp. 536–539.

[4] C.-S. Woo, J. Du, and B. Pham, "Geometric invariant domain for image watermarking," in *IWDW*, ser. Lecture Notes in Computer Science, Y.-Q. Shi and B. Jeon, Eds., vol. 4283. Springer, 2006, pp. 294–307.

[5] S. Anfinogenov, V. I. Korzhik, and G. Morales-Luna, "Robust digital watermarking system for still images," in *FedCSIS*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2011, pp. 685–689.

[6] ——, "A multiple robust digital watermarking systems fro still images," *International Journal of Computer Science and Application*, vol. 9, no. 3, pp. 37–46, 2012.

[7] A. Bruckstein and T. Richardson, "A holographic transform domain image watermarking method," *Circuits, Systems, and Signal Processing Journal Special Issue*, vol. 17, no. 3, pp. 361–389, 1998.

[8] D. W. Alliance, "The digital watermarking alliance overview presentation," http://www.digitalwatermarkingalliance.org/docs/presentations/dwa_presentation.pdf, 2006–2011.

[9] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, ser. North-Holland Mathematical Library. North Holland, January 1983. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/0444851933

[10] P. Bas, T. Filler, and T. Pevný, ""Break our steganographic system": the ins and outs of organizing BOSS," in *Proceedings of the 13th international conference on Information hiding*, ser. IH'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 59–70. [Online]. Available: http://dl.acm.org/citation.cfm?id=2042445.2042452

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 591–596

Real-time Implementation of the ViBe Foreground Object Segmentation Algorithm

Tomasz Kryjak
AGH University of Science and Technology,
Krakow, Poland
e-mail: kryjak@agh.edu.pl

Marek Gorgon
AGH University of Science and Technology,
Krakow, Poland
e-mail: mago@agh.edu.pl

***Abstract*—This paper presents a novel real-time hardware implementation of the ViBe (VIsual Background Extractor) background generation algorithm in reconfigurable FPGA device. This novel method combines the advantages of typical recursive and non-recursive approaches and achieves very good foreground object segmentation results. In this work the issue of porting ViBe to a FPGA hardware platform is discussed, two modification to the original approach are proposed and a detailed description of the implemented system is presented. This is the first, known to the authors, FPGA implementation of this algorithm.**

I. Introduction

DETECTION of moving objects (foreground objects) is one of the most important issues in video sequences processing and analysis. It is used in advanced, automated video surveillance systems and traffic monitoring systems, where the robust segmentation of peoples or vehicles is essential to perform reliable tracking or recognition. Intensive work on the mentioned systems can be observed in the image processing researcher community as well as in the industry.

Foreground object detection methods can be divided into three categories: simple successive frame differencing, the so-called background modelling approach followed by background subtraction and optical flow.

In this paper, a hardware implementation of a method belonging to the second of these categories is presented. An extensive review of different approaches to foreground object detection can be found in [3].

The concept of background subtraction involves object detection based on the difference between the current video frame and the background, where the background is understood as an empty scene, i.e. without objects of interest (people, cars). It is worth noting that foreground object detection is not just a simple moving object detection issue. The background may contain moving elements: flowing water, moving leaves and shrubs, which should not be detected. On the other hand, some objects (eg. a pedestrian) may remain still for a while and should be continuously detected. Another source of segmentation errors are objects that start to move (eg. a parked car). The left empty space is than usually misclassified as foreground (a so called "ghost"). That is why the background representation should be adaptive in order to compensate some normally occurring changes such as lighting or movement of certain objects (i.e. chair in an office), as well as handle difficult cases like ghosts. The process is referred to as background generation or modelling.

In the literature one can find many descriptions of FPGA implementations of background generation methods. An extensive discussion of this issue is presented in [6]. The most important and recent articles are:

- Mixture of Gaussian [4] – HD greyscale video stream processing (1920 × 1080 @ 20 fps),
- Horparsert method [9] — 1024 × 1024 @ 32.8 fps, video stream processing, the high level synthesis language Impulse-C was partially used,
- Codebook [8] — 768 × 576 @ 60 fps video stream processing,
- Clustering [6] – HD colour video stream processing.

An FPGA implementation of background generation algorithms can be used in hardware accelerators (e.g. frame-grabbers with an FPGA device, which perform some image pre-processing and analysis) or smart cameras, where all the image processing, analysis and recognition is performed in the camera and only the results are transmitted to the main processing unit of a surveillance system.

II. The ViBe Algorithm

The foreground object segmentation algorithm ViBe (VIsual Background Extractor) was proposed by O. Barnich and M. Van Droogrnbroeck and described in detail in [1], [2], [12]. It contains several innovative elements (the solution is patented) and allows to obtain very good results, which is confirmed by a high place in the object detection algorithms ranking [5].

The background model in ViBe consists of a set of observed pixel values. This is an important difference compared to the most common methods, where the background model is based on probability distribution function. The authors of ViBe justify this concept, pointing out the difficulties in selecting the appropriate probability distribution and the corresponding update mechanism.

Let $v(x)$ denotes the pixel value in a given colour space at the point x in the image, and v_i the i-th sample from the background model. Then the model for each pixel x is defined as a set of N samples:

$$M(x) = \{v_1, v_2..., v_N\} \tag{1}$$

978-83-60810-53-8/$25.00

In order to classify the pixel $v(x)$ a sphere $Sr(v(x))$ of radius R centred at the point $v(x)$ is defined. The analyzed pixel is considered as background, if at least $\#_{min}$ samples from the model $M(x)$ lie inside the sphere. The distance is defined as Euclidean, and the procedure requires, in the worst case, N distance calculations and N comparisons.

The authors proposed a method of initializing the background model using a single video frame. This results in fast initialization and re-initialization i.e. in case of a sudden lighting change or surveillance system reboot. In this approach, however, the temporal context (history of the pixel) is not available, therefore, the assumption has been made that the adjacent pixels should have similar values. The initialization procedure involves filling the buffer $M(x)$ with randomly selected samples from the pixel's spatial context (size 3×3).

The disadvantage of this approach is its susceptibility to artefacts in the form of "ghosts" — a collection of pixels classified as belonging to the foreground, but actually not related to any real object. The elimination of such interference provides the discussed below background model update mechanism.

The ViBe algorithm uses a conservative update approach — the background model is modified only in the case of classifying a pixel as part of the background. On one hand, it prevents the penetration of moving objects into the background model, but at the same time it can lead to irreparable segmentation errors (e.g. "empty" space left by a car which drove away is classified as an object).

Contrary to popular background generation algorithms that use a pixel buffer approach (average of the buffer, the median of the buffer) where the update process relays on replacing the oldest sample by a new value (FIFO scheme) in ViBe the temporal context is not considered. The sample, to be updated, is chosen at random. In conjunction with the conservative approach this results in an exponential lifespan of a given sample. To further extend the time interval, which is covered by the background model, the update is performed with a fixed probability (e.g. 1/16).

In order to counteract the negative effects of the assumed conservative approach, a mechanism of updating the adjacent background models was proposed. It can be described as follows. If the current pixel $v(x)$ is regarded as belonging to the background, two update procedures are executed: for the current and the neighbouring background models. First of all, in a random fashion, it is determined whether the update should be executed (the proposed by the authors likelihood equals 1/16). Then, in the first case the sample to be substituted is randomly selected (1 out of N). In the second, the neighbouring model (1 out of 8 assuming a 3×3 context) and the sample (1 out of N) are chosen. The selected samples are then replaced by the value $v(x)$.

It is worth noting that the ViBe method requires very few parameters. The authors of the paper [2] proposed the following values: $N = 20$ (number of samples in the model), $R = 20$ (the radius of the sphere, value for greyscale images), $\#_{min} = 2$ (the minimum number of samples, which must lie within the sphere) and the update probability (1/16) and they were used in the module.

III. Considerations About Implementing ViBe in Hardware

One of the main problems with implementing background generation algorithms in hardware is providing a quick access to the external memory resources, where the background model is stored [6]. In the case of the ViBe algorithm is necessary to ensure the following transfer rate:

$$T = N \times B \times PC \times 2 \tag{2}$$

where: N – model size (number of samples), B – number of bits per one sample (for greyscale images $B = 8$, for RGB $B = 24$, for CIE Lab $B = 23$), PC – pixel clock (for VGA resolution 640×480 PC is 25 MHz). The use of the multiplier two, results from the need to perform write and read operations. Substituting the appropriate values the following rates are obtained: $T \cong 690$ MB/s for greyscale and $T \cong 2070$ MB/s for RGB. In case of HD image processing (i.e. 1920×1080, pixel clock 148.5 MHz) the rates are respectively: $T \cong 4898$MB/s i $T \cong 12293$ MB/s.

Modern FPGA boards are usually equipped with an external DDR3 RAM module. In this study two platforms were analyzed: ML605 (Virtex 6 device) and VC707 (Virtex 7 device), both from Xilinx. The first of these is equipped with a memory with a maximum theoretical transfer rate of 6400 MB/s, and the other 12800 MB/s. Wherein, in the case of dynamic memory, it is impossible to obtain the maximum values, because of the necessity of refreshing and accessing individual banks, and columns.

Analysis of the presented numbers allows to draw the following conclusions. VGA-resolution algorithms can be implemented on both platforms. In the case of an HD video stream only the grey-scale version can be realized on the newer VC707 board. Also, there are a few possibilities to process a HD video stream without increasing the memory bandwidth: processing every n-th frame (lower FPS) or storing for example only one out of four pixels (3 from the 2×2 context are approximated).

The ViBe method can be quite easily implemented in hardware. The distance calculation between the current pixel and the samples in the model is possible to realize in parallel. Other operations, including the pseudo-random number generation are also feasible. Quite complex is only the propagation of the current pixel value to neighbouring models mechanism, which requires the generation of a very wide (more than $N \times B$ bits) context and therefore large number of delay lines - usually implemented in Block RAM memory resources.

IV. The Modifications Proposed to the Algorithm

In the first stage of the research the paper [12], in which the authors propose a series of improvements to the ViBe algorithm was examined in detail. Unfortunately, implementing most of the presented ideas in reconfigurable resources in

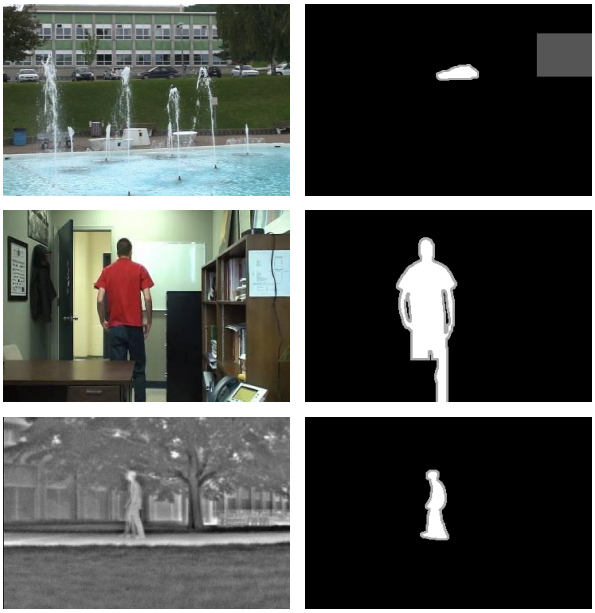


Fig. 1. Sample test images from the `changedetection.net` database. Left coloum - input images, right column - groundtruth. First row — *fountain* sequence (movement in background), second row — *office*, third row — *park* (thermal image)

a pipeline data processing scheme cause huge difficulties or seems to be impossible.

However, in order to enhance the algorithm, two ideas: changing the colour space and a false detection reduction mechanism in areas with background motion (e.g. flowing water) were examined in detail in this paper.

A software model of the algorithm was implemented in C++ using the OpenCV library [7] and examined on the IEEE Workshop on Change Detection (*changedetection.net*) [5] database to evaluate the proposed improvements. The database contains sequences divided into six categories: basic, dynamic background (e.g. flowing river), camera jitter, intermittent object motion, shadows and thermal images. In each of them 4 to 6 videos are included. It can be concluded that the database contains sequences which cover a large part of the situations occurring in surveillance system which are problematic to background generation algorithms. However, the main advantage of the database and a feature that distinguishes it from other collections (e.g. Wallflower [11]), is a large number of manually annotated reference frames with areas marked as: background, shadow, movement, slight blurring and motion (foreground objects). This allows for a reliable assessment of the algorithms performance in different situations. Furthermore, performance results for the most state of the art algorithms are avaliable online (`http://www.changedetection.net/`). Sample images are presented in Figure 1.

The methodology used in the experiments can be described as follows. The object mask computed by the algorithm was compared with the reference mask. Because the ViBe method does not contain a build-in shadow detection procedure, only the foreground and background classification were considered. The following rates were calculated:

- TP (true positive) – pixel belonging to a foreground object classified as a pixel belonging to the foreground,
- TN (true negative) — pixel belonging to the background classified as a background pixel,
- FP (false positive) — pixel belonging to the background classified as a pixel belonging to the foreground,
- FN (false negative) — pixel belonging to a foreground object classified as a background pixel.

Then, based on the calculated parameters two measures were determined: the percentage of wrong classifications:

$$PWC = \frac{FN + FP}{TP + FN + FP + TN} \times 100\% \qquad (3)$$

and precision:

$$P = \frac{TP}{TP + FP} \qquad (4)$$

In the first experiment three colour spaces were examined: greyscale, RGB and CIE Lab. In the first two cases the Manhattan (L1) distance metric was used. Additionally for RGB the Euclidean (L2) metric was calculated. In the case of CIE Lab the fo following formula was used:

$$d_{CIELab} = \alpha \cdot |L_I - L_B| + \beta \cdot (|a_I - a_B| + |b_I - b_B|) \qquad (5)$$

where: L_I, a_I, b_I – current pixel in CIE Lab colour space, L_B, a_B, b_B – background model sample in CIE Lab colour space, α, β - weights (in the experiments set to α = 1, β = 1.5). The parameter values for α and β were chosen after evaluation on several test images. The analysis of the CIE Lab colour space was performed due to good segmentation results obtained in a previous work [6]. The experimental results are summarized in Table I.

TABLE I
PERFORMANCE OF THE VIBE ALGORITHM DEPENDING ON THE USED COLOUR SPACE

Colour space	Distance	PWC [%]	P
Greyscale	L1	3.78	0.67 %
RGB	L1	2.71	0.62 %
RGB	L2	2.28	0.69 %
CIE Lab	eq. (5)	2.18	0.71 %

The presented values are the average rates for the entire dataset. The only modified algorithm parameter was the R threshold. It was set experimentally to obtain best PWC and P ratios.

The results indicate a slight advantage of the CIE Lab over the RGB (L2 metric) colour space. In addition, the hardware implementation of equation (5) is much easier than the Euclidean distance calculation (square and the square root operations require large amounts of FPGA logic resources). Therefore in the final hardware module it was decided to use the CIE Lab colour space, which is a modification to the original proposal from [2].

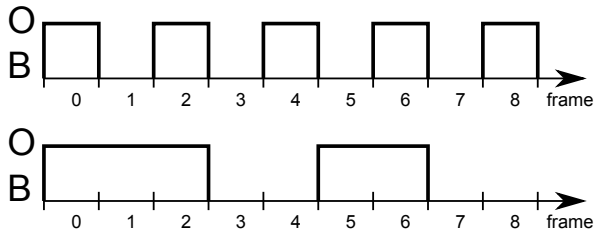


Fig. 2. Two kinds of blinking pixel. O – classification as foreground object, B – classification as background

In the paper [12] an extension to the ViBe method was proposed to detect pixels that are alternatively classified as object and background. They occur most often in cases where small background movement is present (flowing water, fountain, moving grass or leaves). The authors introduced the following mechanism to compensate these interferences. The pixels belonging to the inner boundary of the background that in the previous iteration were classified differently than in the current were detected. In this case the auxiliary variable "blink rate" was incremented by 15, otherwise decremented by 1. If the ratio exceeded the threshold value (set to 30), this pixel was removed from the object mask [1].

This paper proposes an extension to that analysis, which uses two counters: the consecutive classifications of a pixel as background and as an object. This made it possible not only to detect the pixels that change every single iteration (video frame), but also every few ones (Figure 2, bottom graph).

The proposed approach yielded slightly better results. For example, for the "`changedetection.net`" sequence "Canoe" (flowing water) the original approach obtained results: $PWC = 2.19$ and $P = 0.63$, and the proposed $PWC = 1.97$ and $P = 0.68$. It is worth noting that the modification only slightly complicates the algorithm, especially few the hardware implementation.

As post-processing the binary median filter (square window, size 7×7) was selected. It is worth noting thant adding the filter significantly improves the results obtained by the algorithm. An example is presented in Table II.

TABLE II
THE IMPACT OF THE POST PROCESSING MEDIAN FILTERING ON THE ALGORITHMS PERFORMANCE. MEAN RESULTS FOR THE WHOLE DATABASE

Post-processing	PWC [%]	P
none	2.18	0.71 %
median 7×7	1.76	0.88 %

[1] in [12] the mask has been divided into two categories (object and update) and the blinking pixel detection affected only the second. Due to the significant complexity of implementing in pipeline the filling holes operation, which was proposed as post-processing of the update mask, in the presented research this topic was omitted and only one mask was used. For the same reasons the determination of the inner boarder was omitted.

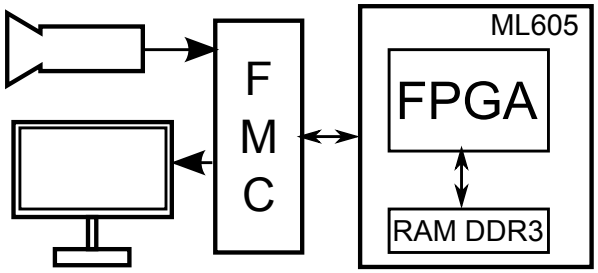


Fig. 3. Scheme of the proposed foreground object detection system

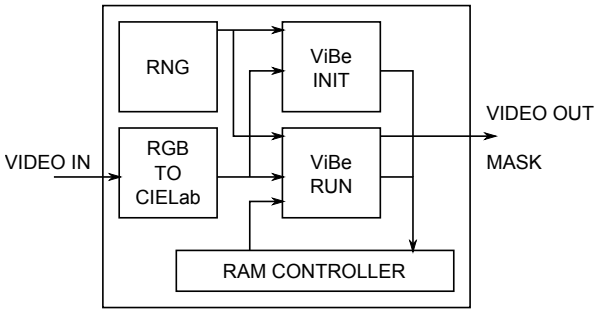


Fig. 4. Block diagram of the modules implemented in the FPGA device

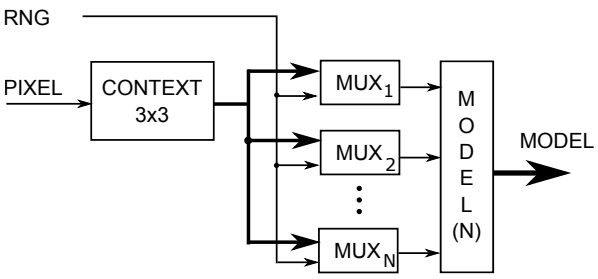


Fig. 5. Block diagram of the *ViBe INIT* module

V. HARDWARE IMPLEMENTATION

Schematically, the proposed system is presented in Figure 3. It consists of an HDMI source (camera or graphic card), HDMI display (LCD screen), Avnet FMC DVI IO module (FPGA Mezzanine Card) with HDMI input and output, and ML605 development board with Virtex 6 FPGA device (XC6VLX240T) from Xilinx. The board is also equipped with an external DDR3 RAM.

All modules were described in VHDL and Verilog hardware description languages. The block diagram of the FPGA design is shown in Figure 4. The *RGB TO CIELab* module is responsible for the colour space conversion [6]. Pseudo-random number generation (*RNG*) was carried out using the concept described in [10]. It is worth noting that the authors made the VHDL code of different RNG versions available, which easy integrates with the project. The used external RAM controller was very similar to the described in [6].

The *ViBe INIT* module is responsible for initializing the background model. The scheme is presented in Figure 5. It consists of a 3×3 context generation module, which uses

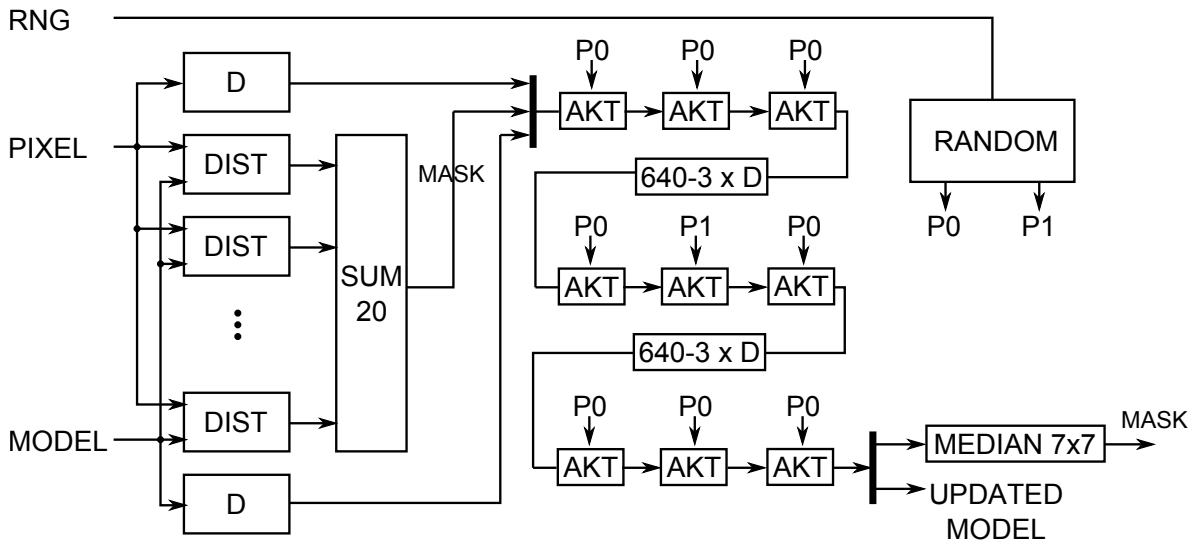


Fig. 6. Block diagram of the *ViBe RUN* module

a delay line approach and N ($N = 20$) multiplexers (*MUX*), which are responsible for the selection of the appropriate sample from the context (1 out of 9 for each *MUX*). The selected value is then stored in the background model. The multiplexers are controlled using a vector obtained from the *RNG* module, thus the model is randomly initialized.

The main module is *ViBe RUN*, which detailed diagram is presented in Figure 6. The inputs are: RNG (pseudo-random number vector), PIXEL (current pixel in the CIE Lab colour space), MODEL (background model read from the external RAM). In the first phase, the distances between the current pixel and the samples from the model are calculated and then compared with the value R (*DIST* - realization of equation (5)). Then it is checked whether the number of distances less than R exceeds the $\#_{min}$ threshold. In the next stage, the 3×3 context consisting of the following signals PIXEL, MODEL and MASK (foreground object mask) is generated. It is worth noting the significant resource usage of this solution - it requires the use of 28 block memory modules (Block RAM). The delay block *D* allows synchronizing the pipeline. The *ACT* module has both a function of a single delay and contains logic that implements the update procedure. The substitution of a background model sample with the current pixel is controlled by the variable P0 (for neighbouring pixels) or P1 (for the central pixel) and depends on the random factor (see Chapter 2) which is schematically illustrated in the form of the *RANDOM* module. The last stage of the process is the median filtering (*MEDIAN 7×7*). The updated model is stored in the DDR RAM and the foreground mask displayed. The blinking pixel detection logic is omitted for clarity reasons.

The presented system was integrated and synthesized for the Virtex 6 FPGA device using the Xilinx ISE Design Suite 14.4. The maximum operating frequency (reported after place & route) was 140 MHz, which is more than enough for processing a VGA colour stream in real-time. FPGA resource usage is summarized in Table III. It is worth noting that due to the large context used in the design and buffers required for the DDR RAM controller, the BRAM_18 (Block RAM) utilisation is quite high. The compatibility of the hardware module with the software C++ model was confirmed using the ISim simulation tool.

TABLE III
FPGA RESOURCE USAGE

Resource	Used	Available	Percentage
FF	12571	301440	3 %
LUT 6	9278	150720	6 %
DSP 48	13	768	1 %
BRAM_18	172	832	20 %

VI. CONCLUSION

This paper describes the implementation of the ViBe background generation algorithm in FPGA. Two modifications were proposed: the use of the CIE Lab colour space and improved detection of blinking pixels that have both increased the effectiveness of the method. The results show that, using an appropriate hardware platform, with a fast external RAM, allows implementing in a pipeline manner a quite complex video stream analysis algorithm in real-time. The proposed system enables processing of a colour video stream with a resolution of 640 × 480 and 60 frames per second. The module can be used in advanced, automated video surveillance systems and other application with require a reliable foreground mask and real-time image processing.

ACKNOWLEDGMENT

This work was supported by the AGH University of Science and Technology grant no 11.11.120.612. The authors would like to thank Mateusz Komorkiewicz for his support in redesigning the DDR RAM controller used in this project.

REFERENCES

[1] O. Barnich and M. Van Droogenbroeck, "Vibe: A powerful random technique to estimate the background in video sequences," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 945–948.

[2] O. Barnichsz and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *Image Processing, IEEE Transactions on*, vol. 20, no. 6, pp. 1709–1724, 2011.

[3] A. S. H. Elhabian S. Y., El-Sayed K. M., "Moving Object Detection in Spatial Domain using Background Removal Techniques - State-of-Art," *Recent Patents on Computer Science*, vol. 1, pp. 32–34, 2008.

[4] M. Genovese and E. Napoli, "FPGA-based architecture for real time segmentation and denoising of HD video," *Journal of Real-Time Image Processing*, pp. 1–13, 2011.

[5] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, June, pp. 1–8.

[6] T. Kryjak, M. Komorkiewicz, and M. Gorgon, "Real-time background generation and foreground object segmentation for high defnition colour video stream in FPGA device," *Journal of Real-Time Image Processing*, pp. 1–17, 2012.

[7] OpenCV, "Strona www: `http://http://opencv.org/` (last acess: May 2013)," 2013.

[8] R. Rodriguez-Gomez, E. Fernandez-Sanchez, J. Diaz, and E. Ros, "Codebook hardware implementation on FPGA for background subtraction," *Journal of Real-Time Image Processing*, pp. 1–15, April 2012.

[9] ——, "FPGA implementation for real-time background subtraction based on horprasert model," *Sensors*, vol. 12, no. 1, pp. 585–611, 2012.

[10] D. Thomas and W. Luk, "FPGA-optimised uniform random number generators using luts and shif registers," in *Field Programmable Logic and Applications (FPL), 2010 International Conference on*, 2010, pp. 77–82.

[11] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, 1999, pp. 255–261.

[12] M. Van Droogenbroeck and O. Paquot, "Background subtraction: Experiments and improvements for vibe," in *IEEE Change Detection Workshop*, 2012, pp. 32–37.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 597–601

Image Semantic Annotation using Fuzzy Decision Trees

Andreea Popescu*, Bogdan Popescu†, Marius Brezovan‡ and Eugen Ganea§
Faculty of Automation, Computers and Electronics
University of Craiova,
Bd. Decebal 107, Craiova Romania
* Email: andreea.popescu@itsix.com
† Email: bogdan.popescu@itsix.com
‡ Email: brezovan_marius@software.ucv.ro
§ Email: ganea_eugen@software.ucv.ro

***Abstract*—One of the methods most commonly used for learning and classification is using decision trees. The greatest advantages that decision trees offer is that, unlike classical trees, they provide a support for handling uncertain data sets. The paper introduces a new algorithm for building fuzzy decision trees and also offers some comparative results, by taking into account other methods. We will present a general overview of the fuzzy decision trees and focus afterwards on the newly introduced algorithm, pointing out that it can be a very useful tool in processing fuzzy data sets by offering good comparative results.**

I. Introduction

IN TODAY'S society there is a continuous process of improvement in the wide area of knowledge acquisition, as it has a direct impact on many areas of activity. Algorithms dealing with extracting knowledge from data have as a result the decision trees and the inference procedures. The classification methods offer different results, in terms of efficiency, domains they can be applied to or ease of use.

The ID3 algorithm was initially introduced by Quinlan [14]. This algorithm offers some restrictions in terms of applicability, as it offers good results for symbolic domains, but not for numerical domains as well. [15]

Fuzzy sets have developed as an extension of the neural networks, since decisions are easier to understand when using them. They provide support for knowledge comprehensibility by offering a symbolic framework. [16] [17]

The symbolic rules together with the fuzzy logic offer complementary support for ease of understanding and modeling fine knowledge details. The fuzzy methods are today's subject in many studies, undergoing continuous improvements in order to offers good results when dealing with inexact data.

The data extraction method we propose in this paper takes into account both a fuzzy approach and a classical decision tree, being able to handle inexact data in a way that is easy to understand.

Some known studies of the fuzzy decision trees present the automatic induction of binary fuzzy trees using new discrimination quality measures. [5] The present method uses for the construction of fuzzy sets an adapted version of the ID3 algorithm. One of the methods we use as a reference is the ID3 algorithm adapted by Janikow in order to be used with fuzzy sets. [2]

II. Fuzzy Decision Trees

Decision tree structures are used to classify data by sorting it from root to leaf nodes. From the well known common tree induction algorithms we mention ID3, C4.5 or CART as they consisted reference points for our work.

In classical decision trees, nodes make a data follow down only one branch since data satisfies a branch condition, and the data finally arrives at only a leaf node.

On the other hand, fuzzy decision trees allow data to follow down simultaneously multiple branches of a node with different satisfaction degrees ranged on [0,1]. To implement these characteristics, fuzzy decision trees usually use fuzzy linguistic terms to specify branch condition of nodes. Different fuzzy decision tree construction methods have been proposed so far.[15] [14] [24]

Different papers are considering the direct fuzzy rules generation without Fuzzy Decision Tree. [26] [27] Complex techniques are used including generation of fuzzy rules from numerical data pairs, collect these fuzzy rules and the linguistic fuzzy rule base, and, finally, design a control or signal processing system based on this combined fuzzy rule base.

In [13] decision tree construction methods are incorporated into fuzzy modeling. They use the decision tree building methods to determine effective branching attributes and their splitting intervals for classification of crisp data. These intervals are then used to determine fuzzy boundaries for input variables, which will be used to form fuzzy rules. As a matter of fact, they use the decision tree construction methods for preprocessing and not for building fuzzy decision tree.

Regarding the approach in [24], the discretization of attributes is made in linguistic terms, relying on the distribution of pattern points in the feature space. Opposite to other fuzzy decision trees, this discretization to boolean form helps in reducing the computational complexity while preserving the linguistic nature of the decision in rule form. In order to minimize noise it's used pruning, resulting in a smaller

978-83-60810-53-8/$25.00

decision tree with more efficient classification. The extracted rules are mapped onto a fuzzy knowledge-based network.

The rest of the paper contains the description of the proposed algorithm for fuzzy tree induction, the set of experiments and the conclusions to the current approach.

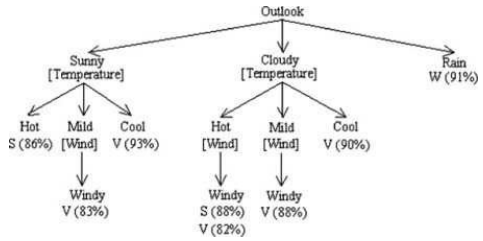


Fig. 1. Example of Generic Fuzzy Decision Tree

III. Proposed Method for Fuzzy Decision Tree Induction

A. Cluster Optimal Index Partitioning for Fuzzy Sets

A very common problem in clustering is finding the optimal set of clusters that best describe the data set. Many clustering algorithms generate a required set of clusters passed as input. In order to solve this problem, the solution would be to repetitively run the algorithm with a different set of inputs until the best schema is found.

In order to validate that, an auxiliary measure needs to be taken care of. We called this cluster optimal index.[1]

A number of cluster validity indices are described in the literature. A cluster validity index for crisp (non fuzzy) clustering is proposed by Dunn [18].

The implementation of most of these measures is very expensive computationally, especially when the number of clusters and the number of objects in the data set grow very large.

In regards to the clusters resulted by applying this mechanism, we have implemented a method of calculating the membership function of the numerical data obtained for each cluster.

The membership degree set is not a binary element from 0, 1 (as for classical decision trees), but is included in the interval [0, 1]. For each node, an attribute has a different membership degree to the current set, and this degree is calculated from the conjunctive combination of the membership degrees of the object to the fuzzy sets along the path to the node and its membership degrees to the classes, where different t-norms operators can be used for this combination.

The fuzzy decision tree induction has two major components: the procedure for fuzzy decision tree building and the generation of the fuzzy set of rules. The proposed fuzzy decision tree building procedure constructs decision tree by recursive partitioning of data set according to the values of selected attribute.

The following steps need to be implemented: attribute value space partitioning methods, branching attribute selection branching test method to determine with what degree data follows down branches of a node, and leaf node labeling methods to determine classes for which leaf nodes stand.

B. Algorithm notations and abbreviations

For better understanding of the described methodology, we have used specific notations as listed below:

- $L = \{L_1, \ldots, L_m\}$, represents the set of m classes of objects,
- $A = \{A_1, \ldots, A_n\}$, represents the set of n attributes we are taking into consideration for our analysis For each attribute we consider the following:
  - $dom(A_i)$ is the domain for the attribute A_i
  - $u^i \in dom(A_i)$, is a crisp value of attribute A_i
  - $FS_i = \{a^i_{p_1}, a^i_{p_2}, \ldots, a^i_{p_{i_k}}\}$, denotes the set of fuzzy numbers resulted after the fuzzy clustering of attribute A_i
  - we denoted FS the set of all fuzzy numbers for all attributes:
$$FS = \{FS_1, \ldots, FS_n\}.$$
- $T = \{t_1, t_2, \ldots, t_s\}$, represent the s training objects. Each element has the following format:
$$t_k = (u^1_k, \ldots, u^n_k, y^1_k, \ldots, y^m_k),$$
where:
  - $u^i_k \in dom(A_i)$ is the crisp value of attribute A_i from the training object t_k
  - a single value from y^i_k is 1, the rest of them are 0 (having 1 on the i^th position means that object t_k belongs to class L_i)
- Membership degree of value $u^i_k \in t_k$ to fuzzy number $a^l_j \in FS_l$ is denoted by $\mu_{a^l_j}(u^i_k)$. For simplicity, this matching operation is denoted by T_0 operator.
$$T_0(u^i_k, a^l_j) = \mu_{a^l_j}(u^i_k) \quad (1)$$
- $\chi = (\chi_1, \ldots, \chi_s)$ are the confidence factors of the objects from the training set ($\chi_i \in [0,1]$ represents the membership degree of object t_i from the training set T). Usually $\chi_i = 1$, $\forall i \in \{1, \ldots, s\}$.
- The Fuzzy set of the set of training objects in node N is denoted by
$$\chi^N = (\chi^N_1, \ldots, \chi^N_s), \quad (2)$$
where χ^N_i is the membership function of object t_i in node N.
- $I(\chi^N)$ represents the entropy of class distribution to set χ^N, in node N
- $I(\widehat{\chi}^N)$ represents the entropy of class distribution after the current node is split by attribute A_i
- $\chi^{N|a^i_j} = (\chi_1^{N|a^i_j}, \ldots, \chi_s^{N|a^i_j})$, denotes the membership degree of training objects from T to fuzzy numbers of attribute A_i, ($\chi_k^{N|a^i_j}$ represents the membership degree of

object t_k to fuzzy number $a_j^i \in FS_i$).
$\chi_k^{N|a_j^i}$ is calculated as follows:

$$\chi_k^{N|a_j^i} = T(T_0(u_k^i, a_j^i), \chi_k^N),$$

where:

- T_0 is defined in 1,
- T is a $T-norm$ operator that can be defined as follows: $T(a,b) = \min(a,b)$

- $Z_k^{N|a_j^i}$, represents the counter for examples in T belonging to class C_k and fuzzy number a_j^i of attribute A_i ($a_j^i \in D_i$).
$Z_k^{N|a_j^i}$ is calculated as follows:

$$Z_k^{N|a_j^i} = \sum_{l=1}^{s} T_1(\chi_l^{N|a_j^i}, y_l^k),$$

where T_1 is a $T-norm$ operator that can be used as follows:$T_1(a,b) = a \times b$.

C. Decision Tree Node Structure

We considered a custom node structure for the extended fuzzy decision tree.

Each node N from the fuzzy decision tree is described as follows:

- F is the set of restrictions on the path from N to the root node
- V is the set of splitting attributes on the path from N to the root node
- S is the set of children of node N, when splitting is done according to attribute A_i
- χ contains the membership degree to node N

D. Fuzzy Decision Tree Induction Algorithm

In what follows it's presented a recursive algorithm for fuzzy decision tree induction of the training objects associated to the dataset we used. It is supposed that the partitioning (or clustering) mechanism of the considered attribute data is already implemented and now further used.

As described above, the numeric partitioning is done using a modified version of C-Means algorithm with additional clustering logic. The algorithm is recursive, it returns the root node and it is called for each splitting phase. Basically, at each level, after attribute partitioning, a particular attribute is selected for further splitting and branching the tree.

As already mentioned, negative information gain can also result from the t-norm (min operator) that is used in the algorithm to compute the membership degrees of the samples in a node. A negative information gain, even if it hasn't a real meaning, can lead to a correct ranking of the candidate test attributes. Instead of that, if information gain ratio is used, a negative value for the information gain cannot produce a good result. answer.

Algorithm 1 Fuzzy decision tree induction

function FUZZYTREE(m, n, s, χ, T, FS, A)
  $N \leftarrow newNode(\chi)$
  $maxGain \leftarrow 0$
  $imax \leftarrow 0$
  for $i \leftarrow 1,n$ **do**
    $Z^N \leftarrow 0$
    for $k \leftarrow 1,m$ **do**
      $Z_k^N \leftarrow 0$
      ▷ For each attribute A_i we compute $Z_k^{N|a_j^i}$ matrix, when $k = 1,m$ and $j = 1,p_{i_k}$
      for $j \leftarrow 1,p_{i_k}$ **do**
        $Z_k^{N|a_j^i} \leftarrow 0$
        for $l \leftarrow 1,s$ **do**
          $\chi_l^{N|a_j^i} \leftarrow T_0(u_l^i, a_j^i)$
          $Z_k^{N|a_j^i} \leftarrow Z_k^{N|a_j^i} + \chi_l^{N|a_j^i} \times y_l^k$
        end for
        $Z_k^N \leftarrow Z_k^N + Z_k^{N|a_j^i}$
      end for
      $Z^N \leftarrow Z^N + Z_k^N$
      $I(\hat{\chi}^N) \leftarrow 0$
      for $k \leftarrow 1,m$ **do**
        $I(\hat{\chi}^N) \leftarrow I(\hat{\chi}^N) - \frac{Z_k^N}{Z^N} \times \log_2(\frac{Z_k^N}{Z^N})$
      end for
      $I(\chi^N|A_i) \leftarrow 0$
      for $j \leftarrow 1,p_{i_k}$ **do**
        $I(\chi^{N|a_j^i}) \leftarrow 0$
        for $k \leftarrow 1,m$ **do**
          $I(\chi^{N|a_j^i}) \leftarrow I(\chi^{N|a_j^i}) - \frac{Z_k^{N|a_j^i}}{Z^{N|a_j^i}} \times \log_2(\frac{Z_k^{N|a_j^i}}{Z^{N|a_j^i}})$
        end for
        $I(\chi^N|A_i) \leftarrow I(\chi^N|A_i) + \frac{Z^{N|a_j^i}}{Z^N} \times I(\chi^{N|a_j^i})$
      end for
      $\widehat{Gain}(\chi^N, A_i) \leftarrow I(\hat{\chi}^N) - I(\chi^N|A_i)$
      $SplitI(\chi^N|A_i) \leftarrow 0$
      for $j \leftarrow 1,p_{i_k}$ **do**
        $SplitI(\chi^N|A_i) \leftarrow SplitI(\chi^N|A_i) - \frac{Z^{N|a_j^i}}{Z^N} \times \log_2(\frac{Z^{N|a_j^i}}{Z^N})$
      end for
      $Gain(\chi^N, A_i) \leftarrow \frac{\widehat{Gain}(\chi^N, A_i)}{SplitI(\chi^N|A_i)}$
    end for
    if $Gain(\chi^N, A_i) > maxGain$ **then**
      $maxGain \leftarrow Gain(\chi^N, A_i)$
      $imax \leftarrow i$
    end if
  end for

Algorithm 2 Part 2

▷ We split node N according to attribute A_{imax}
for $j \leftarrow 1, p_{imax}$ **do**
  for $i \leftarrow 1, s$ **do**
    $\overline{\chi}_i \leftarrow T(T_0(u_i^{imax}, a_j^{imax}), \chi_i^N)$
  end for
  $N.S_j \leftarrow$ FUZZYTREE(m, n, s, $\overline{\chi}$, T, FS, $A - \{A_{imax}\}$)
  $N.S_j.F \leftarrow N.F \cup \{[A_{imax}\ is\ a_j^{imax}]\}$
  $N.S_j.V \leftarrow N.V \cup \{A_{imax}\}$
end for
return N
end function

IV. EXPERIMENTS

In order to demonstrate the applicability of the proposed framework we executed a wide set of experiments and verified the accuracy of the results. We have performed comparative results between the algorithm we developed, denoted here as *BFD* and two other well known similar approaches.

The other references we used were *C*4.5 [23], a well known decision tree learner based on neural networks and *NEFCLASS*, a fuzzy rule based classifier which combines fuzzy systems with neural networks. We analyzed precision and complexity for each of the 3 implementations.

For our tests we used four data sets from *UC Irvine Machine Learning Repository*[20]. You can see in Table I information related to the attributes used in the dataset we considered.

TABLE I
TEST DATASETS

data	size	attributes	classes	missing value
iris	150	4	3	no
glass	214	10 (incl. id)	7	no
thyroid	215	5	3	no
pima	768	8	2	no

The basic approach for testing was 10-fold cross validation. Data was broken into 10 sets of size n/10. We trained on 9 datasets and tested on 1. Performed this operation 10 times and considered the mean accuracy. For the algorithm we developed (*BFD*), we considered a threshold of 10% for the clustering mechanism, and since using Fuzzy C-Means, we also considered a maximum of 10 number of clusters as parameter.

In the Table II we present the average error rate $\bar{\varepsilon}$ after testing with each of the implementation.

TABLE II
ERROR RATE COMPARISON

model	iris	glass	thyroid	pima
BFD	5%	33%	5%	23%
C4.5	4%	30%	7%	31%
NEFCLASS	4%	35%	6%	29%

The precision analysis of the models considered, as seen in the table above, is good for the implementation we made, and certifies that the approach we had has good results and will offer similar results on other data sets, as part of our future work.

In terms of implementation, the algorithm was developed in *C#.NET*, and is part of a complex framework we continuously improve. The implementation decision was taken given the advantages and support that Microsoft offers for their products and the large community supporting, for best practices, performance and efficient problem solving.

V. CONCLUSION

This paper is aimed to introduce a new fuzzy method for handling inexact data. The approach is an extension of the classical decision trees by using fuzzy methods.

The comparative analysis we presented in this paper demonstrate that the considered approach is very solid and returned consistent results.

As observed in the above table, the precision of the proposed method is good enough and we can use it as a good reference and further integrate it in a framework we build among this approach.

The decision of considering algorithm implementation using fuzzy sets reported higher evaluation scores when focusing the training and tests on specific operational fields. We presented a novel and enhanced mechanism of image semantic annotation for segmented color images.

REFERENCES

[1] B. Popescu, A. Popescu, M. Brezovan, and E. Ganea, *Unsupervised Partitioning of Numerical Attributes Using Fuzzy Sets*, Computer Science and Information Systems (FedCSIS), 2012, pp.751-754.

[2] C. Z. Janikow, *Fuzzy Decision Trees: Issues and Methods*, IEEE Trans. on Systems, Man, and Cybernetics - Part B, Vo1.28, No.1, pp.1-14, 1998.

[3] J. Jang, *Structure determination in fuzzy modeling: A fuzzy CART approach*, Proceedings IEEE Conf on Fuzzy Systems, 1994.

[4] Y. Yuan, M. J. Shaw, *Induction of fuzzy decision trees*, Fuzzy Sets and Systems, Vol.69, pp.125-139, 1995.

[5] D. Burdescu, M. Brezovan, E. Ganea, and L. Stanescu, *A New Method for Segmentation of Images Represented in a HSV Color Space*, Advances Concepts for Intelligent Vision Systems - Lecture Notes in Computer Science, Vol.5807, pp.606-617, 2009.

[6] R. Weber, *Fuzzy-ID3: a class of methods for automatic knowledge acquisition*, Proc. of 2nd Int. Conf. on Fuzzy Logic and Neural Networks(lizuka), 1992, pp.265-268.

[7] A. Gyenesei, *Fuzzy Partitioning of Quantitative Attribute Domains by a Cluster Goodness Index*, TUCS Technical Reports, 2000.

[8] J. Zeidler and M. Schlosser, *Continuous-valued attributes in fuzzy decision trees*, Proc. of the 6-th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (Granada, Spain), pp. 395-400, 1996.

[9] R. Agrawal, T. Imielinski, and A. Swami, *Mining association rules between sets of items in large databases*, Proceedings of ACM SIGMOD, 1993.

[10] R. Agrawal and R. Srikant, *Fast algorithms for mining association rules in large databases*, Proceedings of the 20th VLDB Conference, 1994.

[11] B. Popescu, A. Iancu, D. Burdescu, M. Brezovan, E. Ganea, *Evaluation of Image Segmentation Algorithms from the Perspective of Salient Region Detection* Advances Concepts for Intelligent Vision Systems - Lecture Notes in Computer Science, Vol. 6915, 2011, pp 183-194.

[12] P.K. Das and R. Bogohain, *And application of fuzzy soft set in medical diagnosis using fuzzy arithmetic operations on fuzzy number* SIBCOLTEJO, Vol. 05, pp. 107-116, 2010.

[13] T. Tani and M. Sakoda, *Fuzzy modeling by ID3 algorithm and its application to prediction of heater outlet temperature* Proc. of IEEE Int. Conf. on Fuzzy Systems (San Diego), pp.923-930, 1992.
[14] J.R. Quinlan, *Induction on Decision Trees* Machine Learning, Vol. 1, 1986, pp. 81-106.
[15] T.G. Dietterich, H. Hild and G. Bakiri, *A Comparative Study of ID3 and Backpropagation for English Text-to-Speach Mapping* Proceedings of the Interantional Conference on Machine Learning, 1990.
[16] L.A. Zadeh, *A Theory of Approximate Reasoning*. In Hayes, Michie and Mikulich (eds) Machine Intelligence 9 (1979), pp. 149-194.
[17] L.A. Zadeh, *The Role of Fuzzy Logic in the Management of Uncertainity in Expert Systems*. Fuzzy Sets and Systems, 11, 1983, pp. 199-227.
[18] J.C. Dunn, *Well separated clusters and optimal fuzzy partitions*. J. Cybern. Vol. 4, 1974.
[19] Xiaomeng Wang and Christian Borgelt, *Information Measures in Fuzzy Decision Trees*. Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on (Volume:1)
[20] C.L. Blake and C.J. Merz. *UCI Repository of machine learning databases*. http://archive.ics.uci.edu/ml/
[21] C. Borgelt and R. Kruse, *Graphical Models Methods for Data Analysis and Mining*. J.Wiley and Sons, Chichester, England 2002
[22] D. Nauck and U. Nauck, *Nefclass neuro-fuzzy classification (computer software)*. http://fuzzy.cs.uni-magdeburg.de/nefclass/
[23] J.R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufman, San Mateo, CA, USA 1993.
[24] S. Mitra, K.M. Konwar and S.K. Pal, *Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation*. IEEE Transactions on Systems, Man, and Cybernetics, Part C, Vol. 32(4), pp. 328-339
[25] Dong-Mei Huang, *An algorithm for generating fuzzy decision tree with trapezoid fuzzy number-value attributes*. International Conference on Wavelet Analysis and Pattern Recognition, 2008, vol. 1, pp. 41-45
[26] L.-X. Wang and J.M. Mendel *Generating fuzzy rules by learning from examples*. IEEE Transactions on Systems, Man and Cybernetics, vol. 22(6), pp. 1414-1427
[27] F. Herrera, M. Lozano and J.L. Verdegay, *Generating Fuzzy Rules from Examples Using Genetic Algorithms*. Hybrid Intelligent Systems, 2007. HIS 2007. 7th International Conference on, pp. 340-343
[28] Tzung-Pei Hong, Kuei-Ying Lin and Shyue-Liang Wang, *Fuzzy data mining for interesting generalized association rules*. Fuzzy Sets and Systems, vol. 138(2), pp. 255-269

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 603–610

Architectural Redesign of a Distributed Execution Environment

Cosmin M. Poteraș
University of Craiova,
Department of Computers and Information Technology
Faculty of Automation, Computers and Electronics
Craiova, Romania
cpoteras@software.ucv.ro

Mihai Mocanu
University of Craiova,
Department of Computers and Information Technology
Faculty of Automation, Computers and Electronics
Craiova, Romania
mocanu_mihai@software.ucv.ro

Marian Cristian Mihăescu
University of Craiova,
Department of Computers and Information Technology
Faculty of Automation, Computers and Electronics
Craiova, Romania
mihaescu@software.ucv.ro

***Abstract*—This paper describes the architectural redesign of a distributed execution framework called State Machine Based Distributed System which uses a state machine-based representation of processes in order to reduce the applications development time while providing safety and reliability. Initially the system has been built on top of the .Net Framework employing static programming techniques and made use of a custom data storage. The new architecture is intended to take advantage of the fast growing technologies like dynamic languages and graph databases for speeding up even more the applications development and improve the dynamism of the execution model.**

I. Introduction

NOWADAYS, complex systems have become more and more demanding for online visualization and computational steering. Analyzing the outcome of a complex simulation as a post-processing phase is almost unacceptable, while interactivity is a must-have feature. Distributed computing as well as super fast networks, have come to rescue, offering a proper environment for achieving high performance simulation, online visualization and steering. Building complex systems is more a matter of integrating state of the art tools into execution platforms.

Online visualization and computational steering techniques play a very important role in speeding up simulations by allowing on-the-fly analysis and guidance of the ongoing process. Computational steering is nothing else than manual intervention against the ongoing process with the purpose of guiding it towards the space of interest. Computational steering occurs at three different levels: the program level (***program steering,***) which implies changes on the program's state (shared variables), data level (***data steering***), which allows interventions against the data space, and execution level (***dynamic steering***), which implies direct changes to the program's flow by injecting code or invoking routines.

Developing distributed simulation platforms able to provide the means for interactivity has not been easy. Many surveys have documented the task [1], [2], however only few of them get close to fulfilling production needs. It's worth mentioning few of them.

Collaborative Online Visualization and Steering framework, COVS [3] integrates tools for visualization (VTK), communication libraries (VISIT, PV3), steering tools (VISIT, ICENI, gViz).

RealityGrid [4]. [5] is a library which serves as an API which uses check-pointing techniques for steering commands.

CUMULVS (Collaborative User Migration, User Library for Visualization and Steering) [6], [7] developed at Oak Ridge National Laboratory, besides steering, it benefits of powerful recovery techniques and tasks migration.

CSE [8], [9] (Computational Steering Environment) carries out steering through a data manager which works closely with simulation processes (called satellites).

After examining all these platforms it became obvious that choosing the right tools for a distributed simulation and steering system is not an easy task. While ensuring scalability, portability, flexibility, extensibility, could be achieved with the proper tools, ensuring safety seems more like a design task. Writing safe code does not guarantee that the output of the application will be safe. Just imagine the effects of a malfunction of a medical software which assists a surgery. It is essential that at each moment in time, the application is in a consistent and expected state. When designing the application one must make sure that the application reacts accordingly no matter what. This led us to the idea of representing the tasks (processing units) as finite state machines. This way, we force the application developer to consider all states that the application may step into, prior to implementing them properly. Besides an execution model (transitioning states until a final state has been reached) finite state machines serve as packages which can be spread across a distributed environment leading our way towards load balancing and recovery algorithms.

A similar approach has been introduced in [10], and it made use of Statecharts as a conceptual model for a visual tool called Statemate. Statemate is a very powerful simulation environment for statecharts-based models. Statecharts imply a hierarchical and compositional structure increasing the complexity of the model which could lead to an increase of the risk of design errors.

Our previous researches have resulted in the implementation of a state machines-based distributed framework (SMBDS) [11] consisting of an execution engine as well as a class library. The class library reduces the development time considerably by including a set of interfaces and base classes. Developing new applications on top of the frame-

978-83-60810-53-8/$25.00

work only requires the implementation of these interfaces and classes (parameters and state machines). The execution engine loads them and runs them on the distributed environment. SMBDS has been implemented on top of the .Net Framework using C# language. In order to be able to dynamically load custom parameter/state machines classes, the .Net Reflection package has been used. However, C# still remains a static language and for this reason all nodes in the distributed system needed to have access to all code files which makes the distributed environment harder to maintain (whenever a new type of state machine is needed, all nodes need to be updated with the corresponding code file). To overcome this drawback we started considering the power of dynamic languages. There were arguments in favor and against dynamic languages, which will be discussed later in this paper. However we considered the advantages of using dynamic languages worth assuming the drawbacks. This led us to a redesign process which changed the framework almost entirely. At first, we considered only including the dynamic code as a string and simply run it. But then, as we dig into the load balancing and scheduling algorithms, we realized that the execution model is similar to a graph, so why not changing the execution engine so we can take advantage of the NoSQL graph databases. We will discuss the execution model later in this paper.

The rest of the paper is organized as follows. Section II shortly presents the technologies used for implementation pointing out the main benefits of using them. Section III shortly describes the architecture of the state machines-based distributed system (SMBDS) as it used to be prior to the redesign process and explains the reasons behind redesign. Section III presents a new execution model for the state machines-based distributed framework. Section V concludes the paper and presents our future development intentions.

II. Tools and Technologies

In this section we will argument our decisions regarding the infrastructure.

The first choice that we have had to make after deciding to use a dynamic language, was which language we should use. The most popular dynamic languages that we have considered were Python and Ruby. A 'versus' discussion between Ruby and Python is beyond the scope of this paper. After all they can both reach the scope of our framework. However we decided to go for JRuby, as it stands on top of the Java Virtual Machine which makes it possible to call java code from JRuby offering the possibility of including dynamically invoked java code in our state machines.

The second choice that we have had to make was related to the graph representation of our execution model. We have examined the most popular graph database infrastructures, namely Neo4j, Titan Server with three storage back ends (Apache Cassandra, Apache HBase, Oracle Berkley DB) together with the Tinkerpop stack [12].

Neo4j proved to be the most mature graph database. However it falls fast in distributed environments as it's storage is limited to only one machine.

Titan claims to bring a lot of performance boost. Even though it is a relatively new product and it has some limitations (for example, the most important: indexes need to be declared before first use of the key), Titan Server seems a good choice since it uses very powerful storages behind it.

Considering the CAP theorem (any practical database system can only provide two of the following three: scalability, availability and consistency), Titan can use three storage backends: Apache HBase, which provides consistency and scalability, Oracle Berkley DB which provides consistency and availability and Apache Cassandra which provides scalability and availability.

In our case HBase and Cassandra seem to be the best choices depending on the applications needs (distributed systems require scalability). There is a tradeoff between high availability (Cassandra) and consistency (HBase). Both of them are scalable therefore they are suitable for our framework's needs. Their performance and reliability have been successfully proven in production environments at Facebook or Twitter.

As HBase and Cassandra are open-source, distributed and column-oriented storages, and not really graph databases, it might feel strange to use them as graphs, however the Tinkerpop stack models successfully the graphs on top of these storages and they claim for very high performance results [13], [14], outperforming regular SQL relational databases.

The Tinkerpop stack includes a very powerful graph traversal language, Gremlin, built on top of a widely used java interface for graph databases, Blueprints. This makes Gremlin compatible with multiple existing graph engines and eventually with future engines that will implement the Blueprints interface. Being implemented on top of JVM, Gremlin can run java code, so one can actually traverse the graph with custom java code.

Putting all these pieces together we get a very powerful distributed platform for our framework.

Fig. 1 illustrates the software stack for our framework. JVM is where all tools met. On top of it runs JRuby together with all libraries that implement Blueprints. As a bridge between SMBDS and the graph database storages we are using the Tinkerpop's Gremlin language which complies to the Blueprints interface.

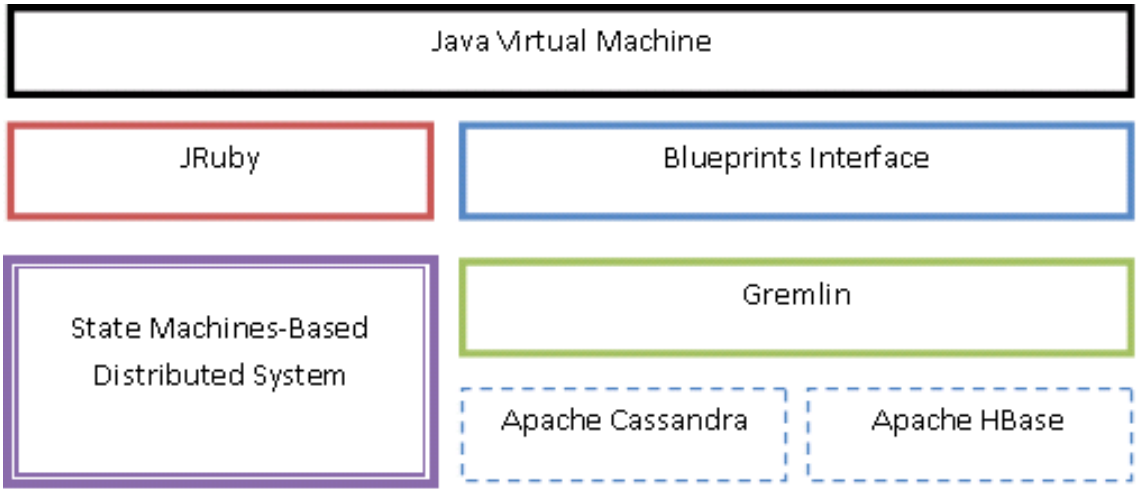


Fig. 1 Software stack

III. Framework's Architecture

This section is intended to reveal the motivation behind the redesign process. Let's examine the static architecture of the system before we proceed with the redesign process.

The challenge behind SMBDS was to offer safety and reliability while taking advantage of a distributed execution environment. SMBDS aims of ensuring safety at design time rather than code safety only, and it does that by representing all computational tasks as finite state machines. This representation forces the application developer to consider all states that the application may step into and react accordingly to each of them. This eliminates the erroneous states while offering a robust and traceable execution model.

The architecture of SMBDS is illustrated in Fig. 2.

SMBDS consists of five modules: Simulation Module (or Processing Module), Visualization Module, Control and Communication Module, Shared Memory and Client Application.

All the magic of SMBDS happens on the simulation level as this is where the simulation tasks, represented as state machines, are being executed.

Each node of the system owns a state machines manager which is the bridge between all modules, mainly because it is responsible for acquiring the state machines from the distributed environment, migrating them to other hosts when necessary, running the state machines and at the same time interacting with the client application for handling steering commands and providing visualization information.

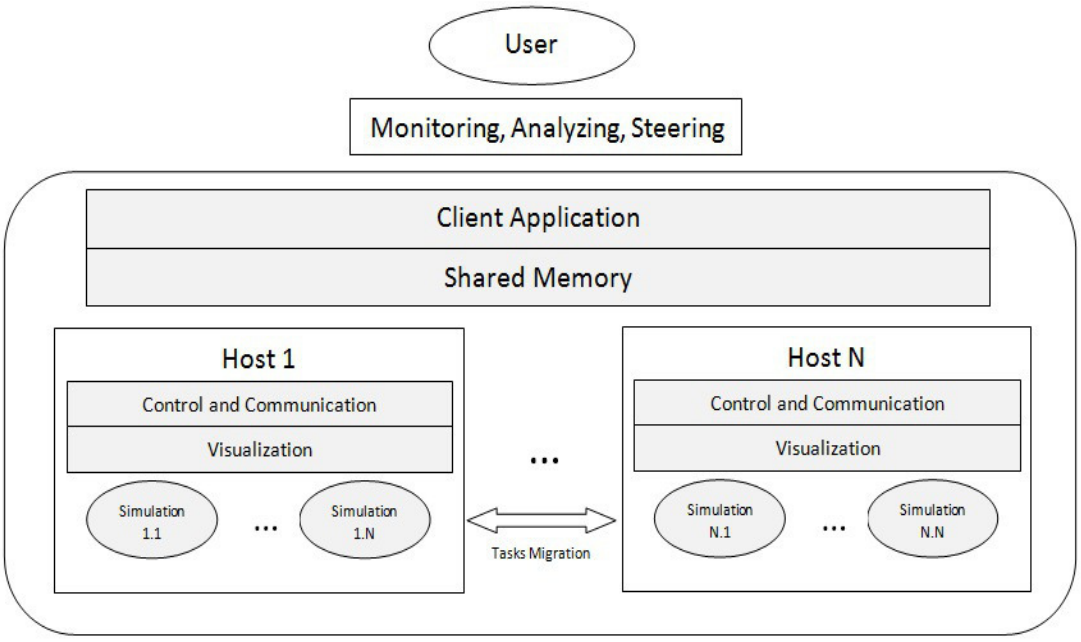


Fig. 2 State Machines-Based Distributed System's architecture.

The control and communication module is able to acquire data, forward output data towards the visualization pipeline while monitoring the available resources and executing steering commands.

As we are dealing with a decentralized architecture, the shared memory took the form of a distributed storage space and it holds the system's parameters (steerable variables).

The user analysis the output data filtered by the visualization pipeline, monitors the state of the computational resources as well as the execution distribution across nodes, and interacts with the system by launching new tasks (state machines), changing the execution parameters (stored in the shared memory), guiding the simulation towards the space of interest.

Concerning the applications development, SMBDS exposes a class library which facilitates the implementation of custom finite state machines which once created, will be passed to the state machines managers and executed.

Implementing new custom state machines requires the implementation of an interface (IParameter), representing the state machine's parameters, if not already implemented, and the extension of a state machine class (StateMachine). Based on these two, the framework's engine is able to manage the execution of state machines. Figure 3 illustrates the class diagram of the framework's execution engine.

The engine's main class is the StateMachineManager. Each host will launch a state machines manager which will run continuously until final states are reached. The main role of this class is to manage the execution of state machines but also migrate the state machines to (pack and send) and from (receive and unpack) other hosts (state machine managers). Packing a state machine consists of pausing the execution and extracting data from the state machine (extracting StateMachineData object), while unpacking restores the machine's execution by creating an instance of the corresponding state machine class, initializing the state machine with the StateMachineData object and lastly resuming the execution from the state where it has previously been paused. Any computations performed for the interrupted state, carried out before packing, will be discarded.

It becomes obvious that migrating a state machine consists of simply sending the StateMachineData object towards the destination host. The StateMachineData class holds execution data of a state machine: a list of parameters, a transition table, the current state, a unique identifier (needed for tracking the machine especially when dealing with migration), the type of the state machine (used to pick the correct state machine class when packing/unpacking the machine using .Net Reflection) and a list of final states.

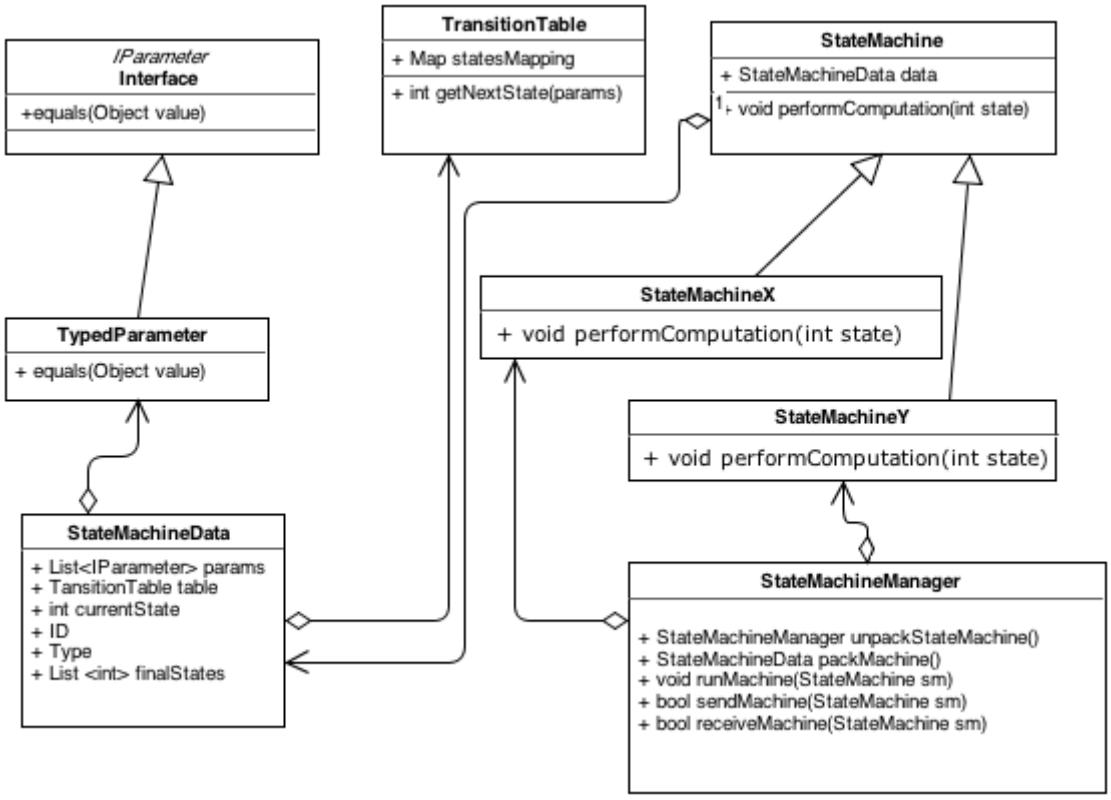


Fig. 3 SMBDS class diagram.

The StateMachine class exposed by the framework's library is an abstract class exposing an abstract method called performComputation which executes the code associated with the current state. This method is being invoked by the state machines manager after each transition. The computations are carried out taking as input the values of the machine's parameters. The parameters are usually altered by the computations of the current state. After the execution of the current state the machines manager invokes the getNextState() method of the machine's transition table which based on the current values of the parameters and the current state determines which is the next state to be transitioned. The manager loops again and invokes the performComputation method for the newly transitioned state unless the state is not final.

To make it more clear, we will include a sample template of how the performComputation method could look like:

```
void performComputation(List<IParameter>
  params,int currentState){
      switch(currentState){
       case: 1
           {//code for state 1}
           break;
       case: 2
           {//code for state 1}
           break;
       …..
       case: N
           {//code for state N}
           break;
       default:
           throw new Excetion
             ("Invalid state. Design time error")
      }
  }
```

It becomes obvious that the business logic of an application will reside in the performComputation method, in the parameter classes as well as in the transition table. This is why the application developer will be needed to derive from the StateMachine class and implement the performComputation method for every type of state machine.

The TranzitionTable class is nothing more than a mapping between <parameter values, current state> tuples and future states. The business logic in case of the TransitionTable class resides in the values it contains rather than in its implementation.

The parameter classes will implement the IParameter interface which exposes the *equals* method which is required to match the values of two parameters and return *true* if they are considered to be equal or *false* otherwise. The *equals* method's implementation might range from very simple equalities to very complex checkings against custom objects, depending on the business logic requirements.

That being said, we can resume the applications implementation requirements to three steps:

1. Implement the parameters classes
2. Implement a state machine class for each type of state machines needed by the application, by extending the StateMachine class
3. Define transition tables.

Here is where the main drawback of the system appears. Since the framework has been developed using a static language (C#) the newly developed state machine class needs to be available on all nodes in order for that node to be able to run state machines of that type. So we can easily identify three important drawbacks at this stage:

- Updating a running application may not be an easy task due to security restrictions.
- All state machines need to comply to one of the classes, which avoids building and executing custom state machines on-the-fly
- The development time increases by the fact that any change in code needs to be spread across the distributed environment in order to be well tested. Also if one needs to run a certain state machine only few times, the development time might exceed the usage time for that state machine.

IV. Dynamic Execution Model

To overcome the drawbacks identified in the previous section we have considered the use of dynamic languages. Instead of implementing classes every time we needed a new type of state machines, we can now include the code as string on the state machine object itself and invoke it dynamically when needed. This raises an important controversy specific to dynamic languages, namely code safety. As the code is not compiled prior to running it, it might contain erroneous code (typos, references to undefined variables or methods, etc) which can damage the execution Besides trying to overcome unsafe code at design time by considering all states that a machine can step into prior to writing the code, we could make use of an adequate development strategy like test driven development and we can increase code safety. However it is well known that dynamic programming requires a lot more attention when writing code than static programming. Assuming that we need to pay attention when writing code, and tediously test it before running it in production, this tradeoff gives us a lot of flexibility.

We can now run as many custom state machines as we want without having to implement classes and move code files across the running system, while reducing the development effort.

Extending the StateMachine class for each type of state machines is no longer needed. We simply make use of the StateMachine class and add a new attribute called *codeMapping* which will map states to their corresponding code represented as strings.

For example, in Ruby Language, the *codeMapping* attribute could be of type Hash and have as keys the state numbers, and as values string containing the code that needs to be run for the associated state:

```
codeMapping = {
   1 => "puts 'code for state 1'",
   2 => "puts 'code for state 2'",
   # …..
   N => "puts 'code for state N'",
}
```

In this context, the state machines manager, would simply invoke the code defined for the current state.
So, instead of calling:

```
performComputation(params, state)
```

it will simply invoke the code dynamically:

```
eval(machine.codeMapping[state])
```

At this stage, our state machines have become abstract collections of states which are nothing but objects that combine code and data, and define some kind of ordering between them resulted from the transition table.

Each state is very similar to a function which takes input parameters and input data, executes some custom code against them and outputs transition parameters and data needed by other states, and so on. If we were to represent the structure of states machines based on their states, data and transition table, it would look like a graph. The nodes of the graph would be the states or data objects, and the arrows would be the dependencies between states constrained by the value of the transition parameters (according to the transition table). Taking it further, we can represent all state machines running at a certain moment and we can conclude that we're dealing with a graph of states and data objects with dependencies between them and we no longer care to which state machine they belong to. Such a graph is illustrated in Fig. 4.

The figure includes the states of two finite state machines (S_{1i} for the former and S_{2i} for the latter) and their input/output data objects (d_i). The transitions between states are being conditioned by the machine's parameters (P) values. For example, machine 1 will move from state S_{10} to state S_{12} only if the value of P_1 matches the value on the arrow (relationship) between S_{10} and S_{12}, which is V_{12}. Each state can be launched only if all data dependencies are satisfied, and the state is considered to be executed completely only after all output data has been delivered. In case a state needs as input data, another state's output, the former state can't be launched unless the latter has been completed. The execution relies on the *execution pointers*. These are nodes in the graph and they point out the last executed state of each state machine. As the machine traverses throughout it's states the corresponding pointer will move to the new state.

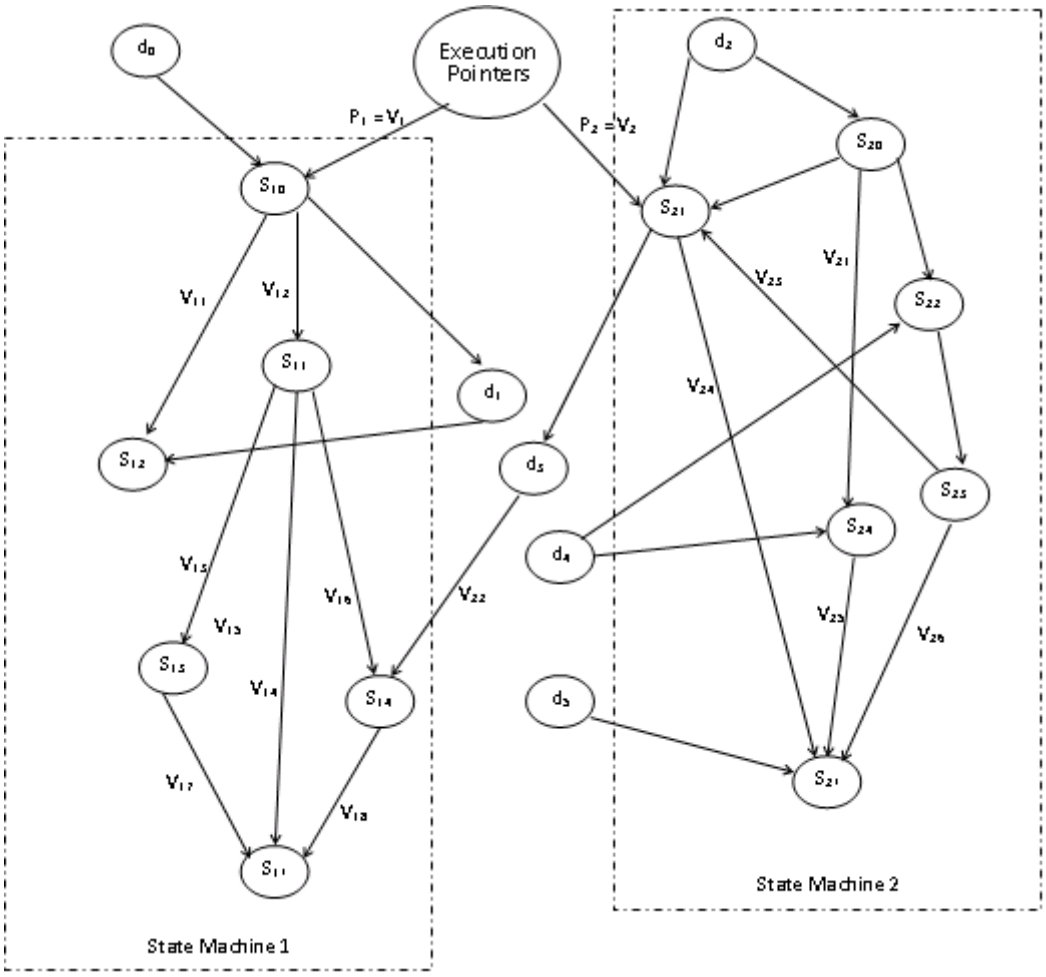


Fig. 4 Execution graph.

It becomes obvious that the execution graph has to be shared by all nodes (execution managers). Normally this would be an important loss for our architecture since the old architecture was not centralized. Keeping the graph in the shared memory area would not be very efficient since it has a distributed architecture and it would require important amount of synchronization communication which would slow down the system considerably.

Fortunately, graph databases and their spectacular evolution have allowed us to benefit from a very powerful infrastructure for distributed environments which has been discussed in section II.

Considering the new execution model we can easily observe that the simulation module of the old architecture no longer handles state machines entirely but states of machines.

For this reason the execution manager available on each host will be responsible for acquiring states that can be immediately executed, run them, save the output and then handle another state.

The algorithm can be resumed by presenting the approximate structure of the class that handles the database (GraphDB), and the main loop that keeps the engine running until the execution is finished.

```
class GraphDB

def exist_final_states
 execution_pointers.out.each do |state|
  return false if not state.is_final?
 end
 return true
end

def pick
 execution_pointers.out.out.filter
       {P = V && in(data).available}
end


def save_output(state)
 database.save(state.output_data)
 exec_pointer= state.in.in(exec_pointer)
 exec_pointer.save_params
    (state.output_params)
 exec_pointer.out = state
end

end
```

The exist_final_states method checks if all final states have been reached, in which moment the execution should stop, by traversing the graph starting from the execution pointer nodes, following the outgoing relationships and returning true if all states pointed by the execution pointers are final states, otherwise it returns false.

The pick method starts from all the execution_pointers, traverses down the tree through the last completed states, down to the next state by checking the current parameters' values and returns the first state found. The method can be easily adjusted to return more than one state depending on the load balancing policy.

The save_output method, saves the output data objects resulted after processing the current state into the graph, identifies the corresponding execution pointer and moves its position to the currently completed state.

The main loop run on every process (simulation process), is similar to the following piece of code:

```
while graphDB.exist_final_states
  current_state = graph_db.pick()
  //dynamically run the Ruby code
  eval(current_state.code)
  graph_db.save_output(current_state)
end
```

V. Conclusions and Future Work

The proposed architecture brings more flexibility to our distributed framework by allowing the developer to write dynamic code and at the same time reduces the development effort by not requiring the implementation of static code for each type of custom state machine. As a consequence the update process of a running application no longer requires sending code files across the distributed environment, but simply running finite state machines with embedded custom dynamic code.

The new graph execution model breaks the finite state machines into independent states and handles them all together. As new state machines arrive, their states get appended to the graph and dynamically allocated to computational resources (through each host's execution manager).

The new approach ensure fair load balancing by having the nodes acquire the processing tasks (machines' states) as their resources become available as opposed to static allocation (prior to launching the application) or centralized dynamic allocation algorithms.

Reducing the amount of work required by a task at the minimum of only one state also improves the load balancing.

Our near future research thoughts refer, to evaluating the performance of the new architecture and identify the performance gain as well as the weaknesses of the new platform architecture.

Dynamic load balancing algorithms have always been an important lead for us and they will be given special concern in the near future.

Data distribution plays a very important role when we deal with load balancing algorithms in distributed environments, as it is more efficient to move the processing towards data than the other way around, therefore special interest will be shown on this issue.

Creating tools for better monitoring and steering of the distributed environment, of both resources and execution is also on our to-do list.

REFERENCES

[1] W. Gu, J. Vetter and K. Schwann. An annotated Bibliography of Interactive Program Steering, SIGPLAN Notices 29 (1994), pp. 140-148 and Technical Report GIT-CC-94-15 (Georgia Institute of Technology)

[2] R.J. Allan and M. Ashworth. A Survey of Distributed Computing, Computational Grid, Meta-computing and Network Information Tools, available from http://www.ukhec.ac.uk/publications/reports/survey.pdf

[3] Morris Riedel, Wolfgang Frings, Sonja Habbinga, Thomas Eickermann, Daniel Mallmann, Achim Streit, Felix Wolf, Thomas Lippert, Andreas Ernst, Rainer Spurzem: Extending the collaborative online visualization and steering framework for computational Grids with attribute-based authorization. GRID 2008: 104-111

[4] S. Jha, S. Pickles, and A. Porter. A Computational Steering API for Scientific Grid Applications: Design, Implementation and Lessons. In Workshop on Grid Application Programming Interfaces, Brussels, Belgium, Sept. 2004.

[5] J. M. Brooke, P. V. Coveney, J. Harting, S. Jha, S. M. Pickles, R. L. Pinning and A. R. Porter, Computational Steering in RealityGrid, Proceedings of the UK e-Science All Hands Meeting, September 2-4, 2003

[6] J. A. Kohl and P. M. Papadopoulos. Efficient and Flexible Fault Tolerance and Migration of Scientific Simulations Using CUMULVS. In 2nd SIGMETRICS Symposium on Parallel and Distributed Tools, Welches, OR, Aug. 1998.

[7] G. A. Geist, J. A. Kohl, and P. M. Papadopoulos. CUMULVS: Providing Fault-Tolerance, Visualization and Steering of Parallel Applications. Intl. Journal of High Performance Computing Applications, 11(3):224-236, Aug. 1997.

[8] J.J. van Wijk and R. van Liere. An environment for computational steering. In G.M. Nielson, H. M¨uller, and H. Hagen, editors, Scientific Visualization: Overviews, Methodologies, and Techniques, pages 89–110. Computer Society Press, 1997.

[9] R. van Liere, J.D. Mulder, and J.J. van Wijk. Computational steering. Future Generation Computer Systems, 12(5):441–450, April 1997.

[10] David Harel, Michal Politi - Modeling Reactive Systems with Statecharts: The Statemate Approach, McGraw-Hill, Inc. New York, 1998, ISBN:0070262055

[11] Cosmin M. Poteras, Mihai L. Mocanu - A State Machine-Based Parallel Paradigm Applied in the Design of a Visualization and Steering Framework, Recent Researches in Applied Informatics, Proceedings of the 2nd International conference on Applied Informatics and Computing Theory (AICT '11), ISBN : 978-1-61804-034-3, pp232-236, WSEAS, Prague, Czech Republic, September 26-28, 2011

[12] www.tinkerpop.com

[13] Rodrigues, M.A., Broecheler, M., "Titan: The Rise of Big Graph Data", Public Lecture at Jive Software, Palo Alto, 2012

[14] Broecheler, M., LaRocque, D., Rodrigues, M.A., "Titan: A Highly Scalabe, Distributed Graph Database", GraphLab Workshop 2012, San Francisco, 2012

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 611–618

Color Classifiers for 2D Color Barcodes

Marco Querini and Giuseppe F. Italiano
University of Rome "Tor Vergata", 00133 Rome, Italy
marco.querini@uniroma2.it, italiano@disp.uniroma2.it

***Abstract*—2D color barcodes have been introduced to obtain larger storage capabilities than traditional black and white barcodes. Unfortunately, the data density of color barcodes is substantially limited by the redundancy needed for correcting errors, which are due not only to geometric but also to chromatic distortions introduced by the printing and scanning process. The higher the expected error rate, the more redundancy is needed for avoiding failures in barcode reading, and thus, the lower the actual data density. Our work addresses this trade-off between reliability and data density in 2D color barcodes and aims at identifying the most effective algorithms, in terms of byte error rate and computational overhead, for decoding 2D color barcodes. In particular, we perform a thorough experimental study to identify the most suitable color classifiers for converting analog barcode cells to digital bit streams.**

I. Introduction

BARCODES are optical machine-readable representations of data, capable of storing digital information. Barcode data are represented as a sequence of bytes, which are then mapped to analog signals (in this case, barcode elements) and transmitted over a printing and scanning (Print&Scan) channel which introduces noise, distortions and interference, corrupting the transmitted signal (in this case, the barcode image after scanning). At the receiver, the distorted barcode is mapped back to bytes. The received binary information is just an estimate of the transmitted binary information. Indeed, byte errors may result due to the amount of noise encountered in the transmission. Because noise and distortions always occur in practice, as a result barcode reading algorithms have to cope necessarily with errors, and the trade-off between reliability and data density of barcodes is a significant design consideration. To cope with errors, redundancy is added by channel coding, which is a viable method to increase reliability in a noisy communication channel (which in our case is represented by the Print&Scan channel) at the price of reducing the information rate. The higher the expected number of errors and the redundancy needed for coping with it, the lower the actual data rate (in our case, the barcode data density).

Traditional barcodes, referred to as one-dimensional (1D) barcodes, represent data by varying the widths and spacings of parallel lines. The amount of digital information stored in 1D barcodes is limited and can be only increased by laying out multiple barcodes. This approach has many negative effects, however, such as enlarged barcode areas, more complex reading operations, and increased printing costs. For this reason, the barcode technology has been deploying geometric patterns (such as squares, dots, triangles, hexagons) in two dimensions: such barcodes are referred to as bidimensional (2D) codes. Both the increasing demand for higher density barcodes and the wide availability of on-board cameras in mobile devices has motivated the need for 2D color barcodes, such as the colored DataGlyphs developed at Xerox Parc [1], the High Capacity Color Barcode (HCCB) developed at Microsoft Research [2], [3], the high capacity color barcode technique proposed in [4], and HCC2D, the High Capacity Colored 2-Dimensional code [5], [6]. Color barcodes generate each module of the data area with a color selected from 2^n-ary schemes (e.g., 4-ary color schemes encoding 2 $bit/module$ or 8-ary color schemes encoding 3 $bit/module$), where a module (or cell) is the atomic information unit of a 2D barcode.

Since black and white codes encode 1 $bit/module$, in principle the data density of a color barcode can be twice (4 colors) or three times (8 colors) as much as the data density of the corresponding black and white barcode. However, the actual capacity depends on the amount of redundancy added to the barcode data for correcting errors, which occur due to both geometric and chromatic distortions introduced by the Print&Scan channel. Since colors are more sensitive to the distorsions introduced by the channel, the measured error rate of color barcodes can be significantly larger than the measured error rate of black and white barcodes, all other conditions being equal (i.e., when all barcodes are generated, printed and scanned under same conditions, such as module size, amount of redundancy, printing and scanning resolutions). In our experiments, under the same operating conditions, black & white QR codes had an average byte error rate of roughly 2% while their 4-color counterpart (HCC2D codes) had an average byte error rate of roughly 10%. In this framework, the higher error rates of color barcodes can be mitigated by the use of larger redundancies in the coding, which in turn may reduce substantially the higher data densities potentially offered by color barcodes, thus reducing their benefits. For instance, in order to tolerate a byte error rate of 2%, we need to reserve at least 4% of the barcode area for an error correction code (such as Reed Solomon), thus obtaining less than 96% for its data density, while a 10% byte error rate implies that at least 20% of the barcode area must be used for error correction, reducing its data density to less than 80%.

In this paper we tackle this problem by designing and experimentally evaluating algorithms for retrieving digital data from color cells undergoing chromatic distorsions (due to printing and scanning), so as to minimize their error rates. In particular, we perform an experimental study of the practical performance of several color classifiers and clusterers for converting analog barcode cells to digital bit streams. This

978-83-60810-53-8/$25.00

allows to identify the most effective algorithms for decoding color barcodes in terms of their error rate and their total running times. Moreover, we investigate the trade-off between redundancy and data capacity for 2D color barcodes. This allows to optimize the data storage, addressing the need for high density barcodes (capable of storing as much information as possible in as small an area as possible).

To accomplish this task, we have developed a prototype capable of using different algorithms for color classification. We have chosen algorithms so that they are representative of general classes, such as minimum distance classifiers, decision trees, clustering, probabilistic classifiers and support vector machines. Our experimental findings show that the impact of different color classifiers on the error rate achieved in decoding can be significant. Furthermore, the use of more complex techniques, such as support vector machines, does not seem to pay off, as they do not achieve better accuracy in classifying color barcode cells. The lowest error rates are indeed obtained by means of clustering algorithms and probabilistic classifiers. From the computational viewpoint, classification with clustering seems to be the method of choice, since it is simple and it does not need time consuming training phases.

II. Related Work

To the best of our knowledge, there is little research in the literature on the color classification for 2D color barcodes. One of the first reported attempt to use color in a 2D barcode can be found in a patent by Han et al. [7], who used reference cells to provide standard colors for correct indexing. Bulan et al. [4] proposed to embed data in two different printer colorant channels via halftone-dot orientation modulation, that is, to print two colors at the same spatial location. This allows to nearly double the capacity of black and white barcodes, which is equivalent to use a 4-ary color scheme for encoding 2 $bit/module$. This work was extended in [8] by using three instead of two colorant layers and a interference mitigating design of the orientations of the three colorants to improve capacity. Microsoft HCCB uses a grid of colored triangles to encode data, using a palette of 4 or 8 colors (4-ary color or 8-ary color scheme). HCC2D, the High Capacity Colored 2-Dimensional Barcode [5] uses a grid of colored squares (using a palette of 4 or 8 colors) and has a symbol structure which builds upon a QR code [9] basis for preserving the QR robustness to distortions. Different color barcode technologies adopt different stategies for classifying colors, that is, for converting analog barcode cells to digital bit streams. For instance, the strategy adopted by the Microsoft HCCB decoder is to make use of a color palette, while Bagherinia and Manduchi [10] proposed an algorithm for decoding barcode elements in a color barcode that does not display its reference colors.

Despite the increasing interest in color barcodes, we are not aware of any previous attempt at performing a comparative analysis of the performance of different methods for color classification in this framework, which is the main contribution of our work. Experimentation in color classification has been previously addressed in other domains such as color recognition of objects in indoor and outdoor images [11], color recognition of license plates [12] or skin color detection in face localization and tracking [13] [14], where, similarly to our problem, there is the need to discriminate among different classes of color pixels. We emphasize that results from other domains (e.g., pixel classification into skin color and non-skin color) do not necessarily carry through the classification of color cells in barcodes, because the underlying conditions are rather different. Indeed, color classification in 2D barcodes mainly focuses on classifying color cells with minimal size (e.g., thousands cells per square inch) after undergoing a printing and scanning process, while in other domains color elements have other characteristics (e.g., colors in video frames representing natural images). Furthermore, the effective decoding of color barcodes requires much more accuracy and precision than the other applications considered for color classification.

III. 2D Color Barcodes

In this section, we introduce 2D color barcodes, which take advantage of colors for achieving higher data density than black and white barcodes. This is obtained at the price of coping with chromatic distortions during decoding. In order to introduce the typical decoding process of a color barcode, we describe next the HCC2D code, which will be used as a paradigmatic example throughout the rest of the paper. We remark that most of the findings reported in this paper about color classification for HCC2D codes apply to color classification for other 2D color barcodes as well.

A. The HCC2D code

In this section, we describe our HCC2D code, a 2D color barcode which is made of a matrix of square color cells, whose color is selected from a color palette. Figure 1 illustrates samples of HCC2D codes with 4 and 8 colors.

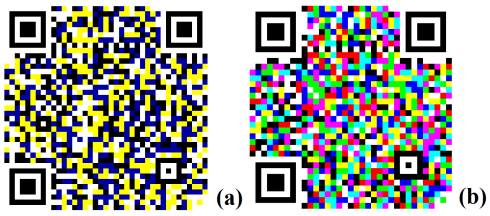


Fig. 1. Samples of the High Capacity Colored 2-Dimensional code (HCC2D): (a) 4 colors and (b) 8 colors. Figure taken from [5]. (Viewed better in color).

We have designed the HCC2D format with the main goal of increasing the data density while preserving the strong robustness to distortions of Quick Response (QR) codes. QR codes are black and white 2D barcode designed by the Japanese corporation Denso Wave which are quite widespread among

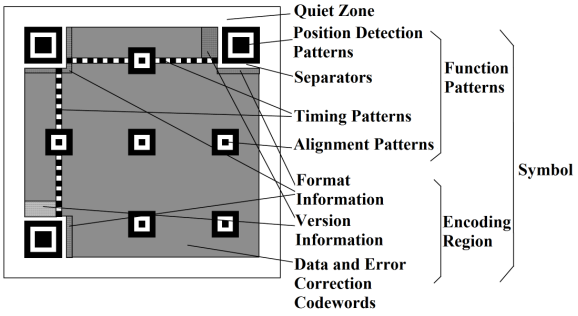


Fig. 2. Structure of generic QR codes and HCC2D codes, which inherit all function patterns of QR codes.

2D barcodes, because their acquisition process appears to be strongly reliable, and are suitable for mobile environments, where this technology is rapidly gaining popularity.

Structure of QR codes, and consequently, of HCC2D codes is illustrated in Figure 2, being composed of *Function Patterns* and *Encoding Regions*. The *Position Detection Patterns*, the *Alignment Patterns*, the *Timing Patterns*, and the *Separators for Position Detection Patterns* support the detection process in detecting the presence, the proper orientation and the correct slope of a code into an image. The *Format Information* describes the error correction level used in the code. As previosly introduced, the higher the correction level, the higher the redundancy and the reliability of the barcode reading process, but the lower the actual data density rate. The *Version Information* represents the code size, that is, the amount of cells (per side) making up the code. Note that the *Version Information* alone does not determine the final print out size (expressed in $inch^2$ or cm^2), which also depends on hardware parameters, that is, on the printing resolution and on how many printer dots make up each color cell. Finally the *Data and Error Correction Codewords* contains data plus redundancy.

We designed the HCC2D code preserving all the *Function Patterns*, the *Format Information* and the *Version Information* defined in the QR code. Maintaining the structure and the position of such patterns and critical information allows the HCC2D code to preserve the strong robustness to geometric distortions of QR code. Because the retrieval of the *Format Information* and of the *Version Information* is a crucial step during the decoding phase (it may led to reading failures) and its storage requirement is small, there is no significant advantage representing it by color cells. The most important changes are gathered in the *Data and Error Correction Codewords* area. The most noticeable difference with a QR code is that the modules belonging to the *Data and Error Correction Codewords* area are of different colors; in a HCC2D code with a palette composed of 4 colors each module is able to encode 2 $bit/module$, while 3 $bit/module$ are stored using 8 colors. Introducing colors in the *Data and Error Correction Codewords* area requires to address some issues, which we have described in details in [5]. Consider that during QR code reading only the brightness information is taken into account, while HCC2D codes have to cope with chromatic distortions during the decoding phase. Since the *Encoding Region* is made of color cells, the HCC2D decoder needs to know the complete color palette in order to decode the symbol. To consider the color palette as an *a priori* shared knowledge between encoding and decoding processes is not a reliable solution; this is because chromatic distortions would not be properly taken into account, arising differently in each printed and scanned image. The processing should be adaptive to each image for better performance. To make a parallelism with black and white barcodes, QR codes compute an adaptive threshold on each image for discriminating dark and light modules, rather than using static thresholds.

In order to ensure adaptation to chromatic distortions arisen in each scanned code, we have introduced in the HCC2D code an additional field, the *Color Palette Pattern*. This is because color cells of a *Color Palette Pattern* are supposed to be distorted in the same way color cells of the *Encoding Region* are. We make use of replicated color palettes either for cluster initialization or for training machine learning classifiers. Figure 3 illustrates *Color Palette Patterns* in HCC2D codes, located at the boundaries. Note that the *Color Palette Patterns* are not too close to the three *Position Detection Patterns* areas and are far away from each other, thus ensuring that they are robust to local distortion. Furthermore, *Color Palette Patterns* take only 2 rows and 2 columns from a symbol consisting of between 21 and 177 rows and columns, and thus, the overhead is small for high density barcodes.

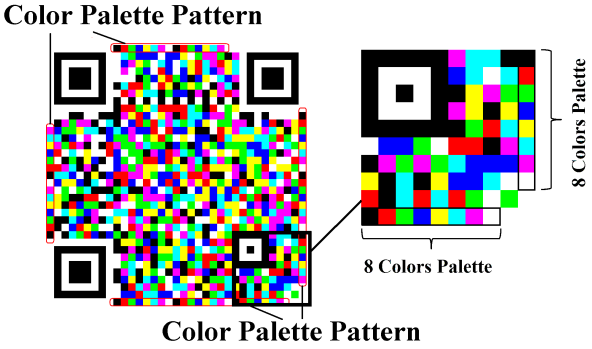


Fig. 3. The four *Color Palette Patterns* are pointed out in a HCC2D code using 8 colors. Figure taken from [5]. (Viewed better in color).

IV. Color Classifiers for 2D Color Barcodes

Since the printing and scanning processes introduce chromatic distortions in color barcodes, the decoding success rate depends on the capacity to correctly classify colors of barcode cells. A barcode cell is correctly classified if its original color (before printing) and the class assigned by the classifier to the cell (after scanning) corresponds to each other. A classifier is an algorithm that distinguishes between a fixed set of classes based on labeled training examples. Algorithms reading black and white barcodes may just use a threshold to separate the two classes (that is, dark and light elements), while cells of color barcodes need to be properly classified in many classes,

depending on the number of colors. We distinguish 4 or 8 classes (each representing a reference color) into which color pixels may fall, where each pixel is sampled from a cell of the 2D color barcode to decode. Each class reference color is associated with either a 2-bit sequence or a 3-bit sequence. The sequence length depends on how many bits are modulated into each barcode cell (as previously introduced, 4-ary color schemes encode 2 $bit/module$, while 8-ary color schemes encode 3 $bit/module$). Because no classifier is perfect, it is important to know whether a classifier is producing good results on real data sets.

A color classifier may have a training phase and a classifying phase. In the training phase, the classifier is provided with known samples. A known sample consists of a region in the barcode image containing the color to be learned and the corresponding label for that color. For every sample that is added during the training phase, the color classifier computes a color feature and assigns the associated class label to it. A color feature vector (to which a barcode cell is associated with) depends on the color space in which the image is encoded. Usually colors are defined in three dimensional color spaces. For instance, these could either be RGB (Red, Green and Blue) or YUV. The Y in YUV stands for "luma", that is brightness (for instance, black and white TVs decode only the Y channel of the signal). U and V provide color information and are "color difference" signals of blue minus luma (B-Y) and red minus luma (R-Y). Without loss of generality, assume that each color feature is a represented as a three-dimensional vector in the YUV color space, because the high correlation between RGB channels and the mixing of chrominance and luminance data does not make RGB a very favorable choice for color analysis and color-based recognition algorithms [13]. When all the trained samples (color feature with a label) are added to the classifier, we get a trained color classifier. After the training phase, barcode cells are classified into their corresponding color classes. In the classifying phase, the trained classifier is used on new observations (color features without labels). The classification engine calculates color features of unlabelled samples and classifies them, by associating a label (in our case, a color class) with each unlabelled color element. Once the classification is completed, the original bitstream (which was previously encoded in the 2D barcode) can be retrieved. This is made by concatenating bits from each bit sequence associated with a barcode cell, where the mapping between a barcode cell and a bit sequence is given by the classification output. For instance, without loss of generality, assume that a 2D color barcode is encoding 2 $bit/cell$ by using 4 different reference colors (e.g., black, cyan, magenta and white). Then, assume that each reference color is mapped to a binary sequence (e.g., black is mapped to $\{11\}$, cyan to $\{10\}$, magenta to $\{01\}$ and white to $\{00\}$). Under these assumptions, a dark cell carries the bit sequence $\{11\}$ whether the cell is labelled with the black class by the color classifier. Because a percentage of color cells is always misclassified in real scenarios, bit errors arise, and thus, original bit stream and decoded bit stream slightly differ from each other.

As previously mentioned, channel coding techiques are capable of correcting errors and restoring the original bit streams, under the assumption that the redundancy introduced in the encoding phase is enough. The main problem considered here is how much redundancy can be sufficient for 2D color barcodes. This depends on many parameters, including the algorithms used for color classification. In order to estimate the most suitable redundancy rate (RR), we implemented five different methods for color classification, each of them being representative of a general class of algorithms (minimum distance classifiers, decision trees, clustering, probabilistic classifiers and support vector machines), and measured their error rates. Next, we briefly describe these five algorithms.

A. Minimum Distance Classifiers (Euclidean distance)

Minimum distance classifiers assign unlabelled samples to classes which minimize the distance between unlabelled data and classes in the feature space. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity. We have used the Euclidean distance (in RGB, YUV, ...) for identifying similar colors (that is, colors with minimum distance), because it is one of the simplest and most popular distance measures. This can be taken as a basic reference method in our experiments: it is a very simple-minded method, and thus we expect all other methods to produce much lower error rates but to be much slower in their running times.

B. Decision Trees (LMT)

A decision tree is a classifier in the form of a tree structure, where each node is either a leaf node (which indicates the value of the target class) or a decision node, which specifies some test to be carried out on a single feature value, with one branch and sub-tree for each possible outcome of the test. There are a variety of algorithms for building decision trees; we have used the Logistic Model Trees (LMT), because they have been shown to be very accurate and compact classifiers. As in ordinary decision trees, a test on one of the attributes is associated with every inner node. Unlike ordinary decision trees, the leaves have an associated logistic regression function instead of just a class label.

C. Classification using Clustering (K-means)

Clustering is the task of assigning a set of objects into groups (denoted as clusters) so that the objects in the same cluster are more similar to each other than to those in other clusters. Hence, color cells are classified once clustering is completed. Clustering itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. We have used the K-means algorithm, which is an unsupervised learning algorithm that classifies a given data set through a certain number of clusters (exaclty k clusters) fixed a priori. Using the K-means algorithm, we can exploit the *a priori* knowledge about the number of colors in color palettes, so

that the algorithm generate exactly 4 or 8 clusters. The starting points (for centroid initialization) can be taken by averaging the series of color palettes.

D. Probabilistic Classifiers (Naive Bayes)

A probabilistic classifier is a function that maps an unlabelled sample to a distribution over class labels. There are a variety of probabilistic classifiers; we have used the Naive Bayes algorithm because it only requires a small amount of training data to estimate the parameters. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. In simple terms, a Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, assume that a color cell (in YUV space) is represented by luma (Y value) and by chroma (U and V values). Even if these luma and chroma values depend on each other, a Naive Bayes classifier considers all of these properties to *independently* contribute to the probability that this color cell is of a given color. The Naive Bayes classifier can be trained very efficiently in a supervised learning setting, using in our case a trained set of known samples taken from the color palette patterns.

E. Support Vector Machines (SVM)

Support vector machines (SVM) are supervised learning models, that is, machine learning tasks of inferring a function from labeled training data (in our cases, labelled color cells belonging to color palettes). A support vector machine constructs a hyperplane or set of hyperplanes in a high dimensional space (in our case, a three-dimensional color space such as RGB or YUV), which can be used for classification. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (denoted as functional margin), since in general the larger the margin the lower the error of the classifier. We have used the Sequential Minimal Optimization (SMO) algorithm for training support vector machines. This is because SMO efficiently solves the optimization problem which arises during the training, avoiding the use of time-consuming numerical optimizations.

V. Experimentation

We developed a prototype for generating and acquiring HCC2D codes. Even if we restricted our experiments to HCC2D codes only, most of the results (in relative terms) can be generalized to the color classification of any other color barcode. This is due the fact that our experiments focused only on the color classification task.

Barcode reading requires a detection and a decoding phase, where color classification is only one part of the decoding phase. The risk is that errors arising in steps other than color classification may affect the experimental results in an unpredictable way. To prevent this, we proceeded as follows. Even if many factors other than the classification algorithm affect the experiment (e.g., the specific hardware involved, routines for detecting or for sampling color cells which are specific to HCC2D codes), their impact has been kept constant through the use of a common set of barcode scans as input for each classifier, along with the use of common routines for every processing step other than color classification (such as image processing, barcode detection or grid sampling routines). For this reason, even if results of our experiments (in absolute terms) depend on the hardware involved and on the HCC2D code, the relative performance of the algorithms considered seems to be of more general extent.

A metric that we adopt is the byte error rate (ByER). By definition the byte error rate is the ratio of the number of incorrectly received bytes compared to the total number of bytes transmitted. It depends on characteristics of the channel such as the signal-to-noise ratio (SNR) at the reveiver and on the accuracy of the "sensors". In our case, the unreliable channel is a printing and scanning channel, while the "sensors" are represented by the classification algorithms and their accuracy is our parameter of interest. We performed an experiment for computing performance statistics for byte error rates (ByER) and computational time of classifier algorithms. By computing these error rate statistics, we are able to identify the most effective color classifier and the most suitable redundancy rate (RR) for it. This allows us to optimize the data rate (DR), that is, the actual data density of color barcodes.

A. Experimental Set-up

We collected 100 barcode scans which make up the sample upon which performance statistics (such as the mean error rate) are computed for each of the 5 methods. Our experimental set-up is as follows:

- We collected 100 barcode scans from real-life applications. Each barcode underwent a real printing and scanning process (i.e., no artificially distorted barcodes).
- Each scan was decoded by each one of the 5 classifiers.
- Color barcodes were printed and scanned at 600 dpi.
- The print out size of each code was 1 square inch.
- The size of each color cell was 4×4 printer dots.
- Each code was made up of 149×149 cells (out of which we had 512 known color cells to use for training classifiers and 19,720 color cells to classify).
- Each code stored 4,930 bytes in 1 square inch, considering data and redundancy from error correction.
- Each code used 4 colors for encoding 2 $bit/module$.
- Color features were expressed in the YUV space, because the explicit separation of luminance and chrominance components makes this colorspace more attractive for color analysis.
- Codes stored different input data so that results were not dependent on the specific barcode instance.

All our experiments were run on a low-end machine equipped with OS Linux Debian 6.0 running on a 1.73 GHz Intel dual core with 2 GB RAM. Documents were printed and scanned on low cost color laser multifunction printers.

TABLE I
SUMMARY STATISTICS FOR BYER CORRESPONDING TO EACH OF THE STUDIED ALGORITHMS.

	Mean	99% Confidence Interval for Mean	Standard Deviation	Standard Skewness	Standard Kurtosis
Euclidean	0.0956	0.0956 ± 0.0160	0.0610	5.8994	4.8087
LMT	0.0851	0.0851 ± 0.0177	0.0675	7.4592	6.9833
K-means	0.0454	0.0454 ± 0.0097	0.0369	5.7278	4.0926
Naive Bayes	0.0621	0.0621 ± 0.0124	0.0473	8.0778	8.9108
SVM	0.0809	0.0809 ± 0.0131	0.0500	3.8256	0.6889

B. Experimental Results

We now turn to the experimental results, by starting with the analysis of the byte error rates (ByER). We remark that in the application at hand, the byte error rate (ByER) is more meaningful than the bit error rate (BER). This is due to the fact that 2D barcodes use block error correcting codes, such as Reed Solomon codes (rather than codes protecting against single bit errors), since they have to withstand accidental damages such as ink spots, affecting a contiguous portion of the barcode.

A Box-and-Whisker plot for ByER data is depicted in Figure 4. Box-and-Whisker plots are convenient way of graphically depicting groups of numerical data through their five-number summaries: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (sample maximum). Quartiles are the three points that divide the data set into four equal groups, each representing a fourth of the population being sampled. The red cross indicates the mean. Outliers, which are observations that are numerically distant from the rest of the data, are depicted too. They are often indicative either of measurement error or that the population has a heavy-tailed distribution.

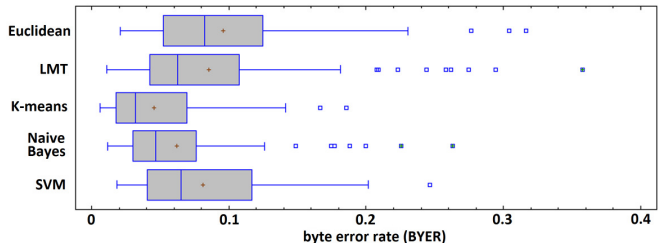


Fig. 4. Box-and-Whisker plot indicating the smallest observation, lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation. (Viewed better in color).

The shape of each distribution is asymmetric and with several strong outliers. Table I show summary statistics for ByER corresponding to each of the studied algorithms. It includes measures of central tendency, measures of variability and measures of shape. It turns out that the most effective algorithm (with the smallest mean and standard deviation) is the K-means clustering algorithm. The ByER of K-means is 4.54% on average; this sample mean, which has been computed on a basis of 100 input images, is close but different from the true mean of the distribution. We have computed the 99.0% confidence intervals for the mean of each ByER distribution, which are reported at the corresponding column in Table I. The classical interpretation of these intervals is that, in repeated sampling, these intervals will contain the true mean of the population from which the data come 99.0% of the time. In practical terms, we can state with 99.0% confidence that the true K-means ByER is somewhere between 0.0357 and 0.0551. Even if these intervals assume that the population from which the sample comes can be represented by a normal distribution (which is not the case here), the confidence interval for the mean is quite robust and not very sensitive to violations of this assumption. Confidence interval for the standard deviation would be quite sensitive, and thus, they are not computed. Of particular interest here are the standardized skewness and standardized kurtosis (reported at the last columns of Table I), which can be used to determine whether the sample comes from a normal distribution. Values of these statistics are outside the range of -2 to +2, indicating significant departures from normality. Figure 5 illustrates percentiles of ByER distributions. Percentiles are values below which specific percentages of the data are found.

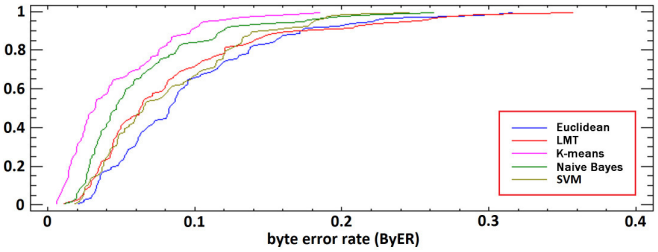


Fig. 5. Percentiles for ByER data of each algorithm. (Viewed better in color).

We can interpret the percentile plot as follows. Consider the 90-th percentile (the value below which 90% of the cases fall) for K-means ByER, its value being 0.0971. This means that 90 barcodes out of 100 would be decoded if the symbols were robust to ByER up to the 90-th percentile ($\approx 9.71\%$). If the K-means algorithm is used for color classification, we may state that $Prob(ByER < 10\%) \approx 90\%$, even if this is just an estimation of the probability on the basis of ByER data collected. Consider that if locations of errors are not known in advance, then a Reed Solomon code can correct half as many errors as there are redundant symbols. For this reason, in order to achieve a success rate of 90%, the redundancy rate (RR) should be around twice the 90-th percentile ($RR \approx 19.42\%$). Data Rate (DR) is therefore reduced to the 80.58% of the overall capacity. Because HCC2D codes are capable of storing 4,930 $bytes/inch^2$ (data plus redundancy), this would results in an effective data density of 3,972 $bytes/inch^2$ with a success rate of 90%. Table II shows this trade-off between data density and reliability (in terms of success rate) for each method. Error rates to tolerate for achieving the target success rate are illustated for three levels (80%, 90% and 95%), along with the corresponding data rate (DR), which is expressed as ratio of data bytes to overall bytes (data plus redundancy bytes). Redundancy rate (RR) is omitted being exactly twice

TABLE II
PERFORMANCE AS FUNCTION OF SUCCESS RATE FOR BARCODE READING.

	Error Rate to Tolerate ($ByER$)	Effective Data Rate (DR)	Effective Data Density $bytes/inch^2$
	85% Success Rate		
Euclidean	0.1547	0.6906	3,404.65
LMT	0.1399	0.7202	3,550.58
K-means	0.0837	0.8326	4,104.71
Naive Bayes	0.1081	0.7838	3,864.13
SVM	0.1320	0.7360	3,628.48
	90% Success Rate		
Euclidean	0.1723	0.6554	3,231.12
LMT	0.1717	0.6566	3,237.03
K-means	0.0971	0.8058	3,972.59
Naive Bayes	0.1190	0.7620	3,756.66
SVM	0.1463	0.7074	3,487.48
	95% Success Rate		
Euclidean	0.2223	0.5554	2,738.12
LMT	0.2511	0.4978	2,454.15
K-means	0.1126	0.7748	3,819.76
Naive Bayes	0.1755	0.6490	3,199.57
SVM	0.1818	0.6364	3,137.45

the ByER to tolerate (because of the Reed Solomon code). Finally, barcode data density is expressed in terms of data bytes per square inch. In order to state that K-means outperforms the other 4 algorithms, we did not just rely on data shown in Table II; we have addressed the statistical significance of our experimental results. Statistical hypothesis testing is used to determine whether an experiment conducted provides enough evidence to reject a null hypothesis. We are interested to reject the null hypothesis for which there is no (statistically significant) difference among the ByER distributions related to the 5 classifiers. We can consider ByER distributions two-by-two as paired samples. "Paired"samples means there are two measurements on each sample unit, e.g., measurements on the same subject before and after an intervention. This is the case here, because there are measurements on the same barcode scan before and after the "intervention" (i.e., substitution of the color classifier). We have run paired tests on these samples such as the sign test and the Wilcoxon signed-rank test, for testing the null hypothesis that there is "no difference in medians" between the distributions.

The result of each test is denoted as P-value, which is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. If the P-Value is under 0.01, the medians of the samples are significantly different at the 99.0% confidence level. For instance, running the Wilcoxon test on the Naive Bayes ByER distribution and on the K-means ByER distribution results in a P-Value of $1.03 \cdot (10^{-9}) \ll 0.01$. Running paired tests on ByER distributions taken two-by-two as paired samples, we have rejected all null hypotheses but one; we can say nothing about the performance difference between logistic model trees and support vector machines. Values computed on 100 barcode scans suggest that SVM has smaller average ByER than LMT, but the difference was found not to be significant (P-value resulting from the Wilcoxon signed-rank test is $0.7413 \gg 0.01$ and from the sign test is $0.6170 \gg 0.01$). We cannnot use more powerful statistical tests (parametric tests such as the t-test) because the normality assumption would be violated. In summary, we can state that, at the 99.0% confidence level, K-means outperforms Naive Bayes, which in turn outperforms support vector machines and logistic model trees, which in turn outperform Euclidean classifiers.

Finally, we address the computational overhead introduced by the color classifiers considered. We stress that the overall running time is important, since in many applications a color barcode can be decoded on low-end devices, such as mobile phones or tablets. Figure 6 illustrates the Box-and-Whisker plot for computational time distributions (the sample size is 100 elements for each method). As expected, the Euclidean classifier is the simplest and fastest algorithm among the studied methods; it is capable of classifying 19,720 color cells in a few milliseconds. In our experiments, the K-means algorithm was also fast (order of milliseconds), while all other classifiers had a much higher computational overhead, as they required up to several seconds for the classification phase.

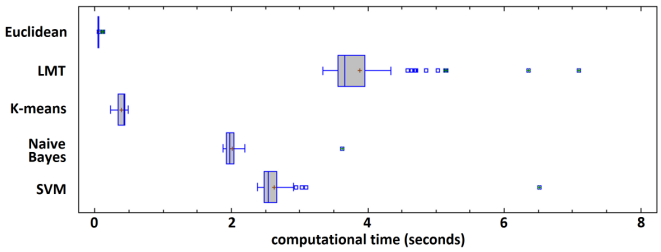


Fig. 6. Box-and-Whisker plot for computational time. (Viewed better in color).

C. Experiments with Mobile Phones

In this section we extend the experiments carried out previously on desktop scanners to other devices, such as mobile phones. The experimental setup is the same as in our previous experiment except that we use mobile phones (such as Samsung Galaxy and Google Nexus 4) for reading color barcodes and that we increase the print out size to 1.5 inch per side (with a cell size of 6×6 printer dots), in order to allow the camera focus to work properly.

TABLE III
SUMMARY STATISTICS FOR BYER DATA RELATED TO BARCODE READING BY MOBILE PHONES.

	Mean	99% Confidence Interval for Mean	Standard Deviation	Standard Skewness	Standard Kurtosis
Euclidean	0.1065	0.1065 ± 0.0382	0.1457	7.4383	5.4823
LMT	0.0851	0.0851 ± 0.0294	0.1122	9.3484	11.380
K-means	0.0832	0.0832 ± 0.0454	0.1729	15.406	31.354
Naive Bayes	0.1293	0.1293 ± 0.0329	0.1255	6.8177	6.0854
SVM	0.1023	0.1023 ± 0.0266	0.1016	8.1937	10.373

Table III reports, with exactly the same format used in Table I, the results of our experiments with mobile phones. It can be seen that the average error rate and the standard deviation are larger (in absolute value) for mobile phones than for desktop

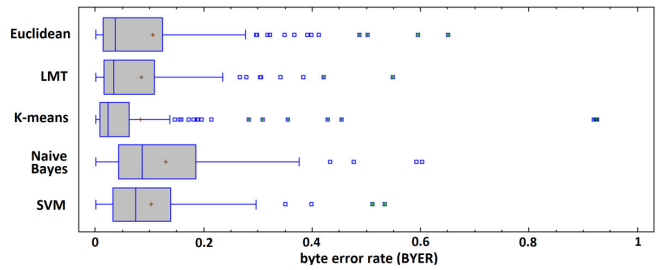


Fig. 7. Box-and-Whisker plot for ByER data related to barcode reading by mobile phones. (Viewed better in color).

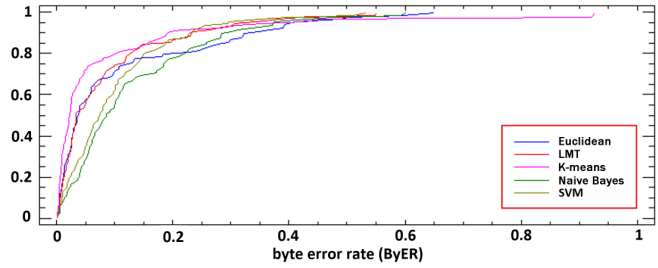


Fig. 8. Percentiles for ByER data related to barcode reading by mobile phones. (Viewed better in color).

scanners. This is due to the fact that desktop scanners have controlled light intensity, while pictures taken from phone cameras present a much larger variation in light conditions.

Beside the absolute values of error rates, which are dependent on the specific devices used for printing and scanning, it can be seen that, in any case, the choice of the classifier has a non-negligible impact upon the error rate distribution. Figure 7 illustrates, similarly to Figure 4, the Box-and-Whisker plot for byte error rate in mobile environment. The K-means algorithm is still the most effective, even if not as effective as in the previous experiment, because there are few cases in which the initial centers do not allow K-means to converge to the optimal solution. This situation occurs especially in case of strongly non-uniform illumination of the barcodes. Analogously to Figure 5, Figure 8 illustrates percentiles of ByER distributions, showing that the K-means curve tends to perform quite well in comparisons with the other curves, except for the fact that there are few observations with high error rates (the far right of the curve) in which the algorithm does not converge to the solution.

VI. Conclusions and Future Work

Our work addressed the trade-off between reliability and data density in 2D color barcodes, performing an experimental study in both desktop and mobile environments. The experimentation showed that the impact of the choice of the color classifier on the error rate is significant and that more complex classifiers do not necessarily achieve better accuracy in classifying color barcode cells. By means of this study, we identified the most suitable ways to convert analog color cells to digital bit streams. We have so far used the same algorithms in both desktop and mobile environments; those methods were found to be more suitable for desktop settings than for mobile scenarios.

For future work, we see a number of interesting directions where our study maybe extended, in particular for mobile devices. An important extension of this work will be to design specific approaches to decode color barcodes acquired by mobile phones. Optimizations for the mobile scenario will attempt to minimize the error rate by taking into account models of light variation for addressing the problem of strong non-uniform illumination, which is significant in mobile scenarios. Furthermore, this study could be extended to barcodes that use 8 instead of 4 color classes.

Acknowledgments: This work has been partially supported by the EU under Contract no. FP7-SME-2010-1-262448 - Project SIGNED (Secure Imprint GeNerated for papEr Documents).

References

[1] D. Hecht, "Printed embedded data graphical user interfaces," *Computer*, vol. 34, no. 3.

[2] Microsoft Research, "High Capacity Color Barcodes," http://research.microsoft.com/projects/hccb/, 2012, [Online; accessed 21-September-2012].

[3] D. Parikh and G. Jancke, "Localization and segmentation of a 2D high capacity color barcode," in *Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision*. IEEE Computer Society, 2008.

[4] O. Bulan, V. Monga, and G. Sharma, "High capacity color barcodes using dot orientation and color separability," in *Proceedings of Media Forensics and Security*, vol. 7254. SPIE, January 2009.

[5] M. Querini, A. Grillo, A. Lentini, and G. Italiano, "2D color barcodes for mobile phones," *International Journal of Computer Science and Applications (IJCSA)*, vol. 8, no. 1, pp. 136–155, 2011.

[6] M. Querini and G. Italiano, "Facial biometrics for 2D barcodes," in *Federated Conference on Computer Science and Information Systems (FedCSIS 2012), Wroclaw, Poland*, 2012.

[7] H. Tack-don, C. Cheol-ho, L. Nam-kyu, S. Eun-dong *et al.*, "Machine readable code image and method of encoding and decoding the same," 2006.

[8] O. Bulan and G. Sharma, "High capacity color barcodes: Per channel data encoding via orientation modulation in elliptical dot arrays," *Image Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1337–1350, 2011.

[9] ISO 18004:2006, *QR Code 2005 bar code symbology specification.*

[10] H. Bagherinia and R. Manduchi, "A theory of color barcodes," in *ICCV Workshops*, 2011.

[11] S. Buluswar and B. Draper, "Color recognition in outdoor images," in *Sixth International Conference on Computer Vision*. IEEE, 1998, pp. 171–177.

[12] D. Guo, L. Chen, Z. Lu, and L. Han, "Vehicle plate location techniques based on plate grounding-color recognition," *Computer Engineering and Design*, vol. 5, p. 023, 2003.

[13] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recogn.*, vol. 40, no. 3, pp. 1106–1122, Mar. 2007.

[14] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *TENCON 2000*, vol. 2, 2000, pp. 421–424.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 619–626

A Novel Portable Surface Plasmon Resonance Based Imaging Instrument for On-Site Multi-Analyte Detection

Sara Rampazzi, Francesco Leporati, Giovanni Danese, Nelson Nazzicari
Dip. di Ing. Industriale e dell'Informazione
Università degli Studi di Pavia
via Ferrata 1, Pavia, Italy
Email: sara.rampazzi01@universitadipavia.it, {francesco.leporati, gianni.danese}@unipv.it, nelson.nazzicari@gmail.com

Lucia Fornasari, Franco Marabelli
Dip. di Fisica "A. Volta"
Università degli Studi di Pavia
via Bassi 6, Pavia, Italy
Email: {franco.marabelli, lucia.fornasari}@unipv.it

Andrea Valsesia
Plasmore s.r.l.
via G. Deledda, Ronco (Varese), Italy
Email: andrea.valsesia@plasmore.com

***Abstract*—In the last decade the need for portable Surface Plasmon Resonance (SPR) biosensors capable of on-site simultaneous multiple assays increased steadily. Several devices are available affected, however, by limitations in terms of costs, size, complexity and portability.**

A compact low-cost SPRi biosensor based on a novel method for multi-analyte detection is presented.

The prototype consists of a nanohole array biochip integrated with a compact optics and an elaboration system. A CMOS image sensor captures reflected light from the biochip surface irradiated by a 830 nm LED. The entire system is managed by an ARM9 processor.

The biosensor was able to detect a ~10^{-5} RIU change in the refractive index without analyte receptors at a glycerol concentration equal to 0.2%. Results are available in 14 seconds on LCD display and immediately stored to external SD memory. Preliminary experiments confirmed the strong biosensor's usability in a wide range of applications and fields.

I. Introduction

THE last 30 years have witnessed a rapid growth in the use of biosensors in both research and practical applications. Molecular biology, biotechnology together with genetic, protein and pharmaceutical engineering are the most common applications fields of these devices. This trend is due to technological innovations that increased their sensitivity, versatility and integrability within microprocessor-based electronics. However, most of these powerful instrumentations are only used in specialized laboratories because of their high cost. In particular the biosensors based on Surface Plasmon Resonance (SPR) have been developed into a very useful technology applications due to their high sensitivity but the majority of this devices are only for research use [1]. The Surface Plasmon Resonance is an optical phenomenon that offers specific advantages in bio-molecular studies. First of all, it is label-free technique that allows real-time direct detection of molecular binding, enabling the determination of their concentration during this interaction without the need for fluorescence or radioisotope labelling. Moreover, it can be applied for kinetic measurements and to perform simultaneous assays on the same biochip (SPR *imaging*-SPRi). Antibody–antigen interactions, peptide/protein–protein interactions, DNA hybridization conditions, biocompatibility studies of polymers, biomolecule–cell receptor interactions and DNA/receptor–ligand interactions have been well analyzed by means of this approach [1]. However the complex fabrication procedure to make the measures more accurate and fast, increases their cost (more than 10,000$) and their size (most of them are benchtop instruments). This limited the diffusion of SPR biosensors outside the industrialized countries, hampers significant improvement of hygiene and environmental conditions. We present a novel portable SPRi biosensor based on a nano-structured crystal biochip developed by [2]. It is a multi-parametric system for biological and molecular interaction monitoring using the Localized Surface Plasmon Resonance (LSPR) phenomenon. Our device allows to detect the presence and the amount of specific target molecules (called *analytes*) in a liquid samples without the use of an external computation device (like computer). The potential applications fields of this instrument are enormous: biochemical and chemical assays in medicine for diagnostic and monitoring of the patients, for water and soil pathogen detection, for process control in pharmaceutical chemistry and drug and food monitoring for toxins detection. This biosensor allows the acquisition of the light reflected from crystal's surface irradiated at a specific wavelength; the crystal is a biochip in which micro-spots of antibodies deposited on the surface are sensitive to different analytes to be identified in the sample. The estimate of the presence and concentration of the searched analytes is available in about 14 seconds on a LCD display and stored in a MicroSD card. This work will focus on the development of the prototype together with the implementation choices to realize a compact low-cost and low-power consumption SPR device. The Linux kernel-based modules conceived for the images acquisition and elaboration are also discussed with a novel method to identify analytes presence and amount. Finally, the detection capabilities of the device are provided

978-83-60810-53-8/$25.00

measuring bulk refractive index changes in glycerol solutions at different concentrations.

II. Integrated SPR System Available

The most common commercial SPR biosensors in the last decades are generally based on the Kretschmann's configuration where a laser source radiates a glass prism covered with a thin metal film. However this simple technique features sensitivity limitation for low molecular weight analytes [3]. The technology improvement allowed new approaches to SPR for increase the sensitivity (plasmon waveguide resonance, channel parallel array, etc...) but this made biosensors more expensive with less attention to an integrated, portable, low-cost and low-power consumption device.

Only recently the market geared toward the fabrication of compact instrumentations for the simultaneous detection of one or multiple analytes and this implied the miniaturization of each system's components and the develop of new SPR technology solutions.

The most best known example of marketed compact SPR devices is Spreeta (Texas Instruments, USA). These instruments measure approximately 1.5 cm x 0.7 cm x 3 cm and consist of a plastic prism assembled on a PCB that contained a infrared LED, a linear diode array detector and a non-volatile memory [4]. The LED light beam through a plastic sheet polarized, strikes a glass chip coated with a gold layer. The SPR waves produced are captured by the diode array detector. Spreeta cost is about 50$ [5] with a resolution of $5x10^{-6}$ RIU [4]. Spreeta, however, requires an integration with external fluidics systems and suitable elaboration units to manage acquisition and analysis. Another limitation is due to temperature which significantly affects the refractive index, requiring a control system to keep it constant.

Many research works exploit Spreeta technology to develop compact multi-analyte SPR instruments. Chinowsky's describes a semi-automatic 24-channel system for toxin identification, called “SPIRIT”. This lunchbox-size instrument weighs approximately 3 Kg and is made by eight replaceable three-channels Spreeta 2000, a 1 MHz analog-to-digital converter and a DSP microcontroller [6].

The device also contains a touch-screen LCD display and a flash memory for data storage, however needs a PC connection for long-term reaction monitoring and displaying of all the 24 channels.

Another low cost biosensor based on Spreeta is described by Hu et al. in 2009. This bioanalyzer uses a tree-channels Spreeta TSPR1K23 and three processors for temperature control, data analysis and display control [7]. Similarly to the previous one, this instrument needs a high precise temperature sensor and uses a PID algorithm control for temperature regulation. Only three analytes for sample are detectable and its considerable weight makes it unsuitable to perform on-site assays also in unfavourable environments.

In addition to Spreeta devices, different approaches can be considered to optimize the Kretschmann configuration. For example, in 2010 Cai et al., developed a portable SPR biosensor based on the image scanner chip LM9833. A wedge shape laser beam radiated a cylinder prism and the reflected light was captured by a CCD camera driven by the image scanner [8]. A single board industrial PC was used to provide the analysis results featuring about 1 Kg weight. The refractive index sensitivity was about 6×10^{-5} RIU, however this device required 30 minutes of baseline stabilization before each analysis.

A palm-size biosensor based on light source modulation was developed by Shin et al. in 2010. The beam of a diode laser was modulated by a rotating mirror and the reflected light was captured by a CMOS [9]. The refractive index resolution was quite interesting (about $2.5x10^{-6}$ RIU at a 3% glycerol concentration) but this device requires temperature control and PC for elaborating signals via LabVIEW code. It also suffers from mechanical solicitations that require an highly synchronization between the frame rate of the CMOS image sensor and the revolution speed of the mirror, to limit the noise level due to the present artefacts that compromise the portability.

Another solution has been adopted by Wichert et al. His team used the Plasmon Assisted Microscopy of Nanoobject (PANOMO) technique with FPGA technology and CCD camera. However it allows to detect only specific viruses [10].

In parallel to Kretschmann-based systems, optical fibers and waveguide structures have been used to fabricate miniaturized biosensors (Tzyy-Jiann et al., 2004). The use of waveguides is similar to Kretschmann configuration since by coating the fiber with a metal film, each reflection of light, corresponds to the reflection spectrum of a Kretschmann configuration. However this solution requires a careful choice of the fiber type and complex design and assembly that increase fabrication costs [11].

Another technique uses diffraction gratings to couple the plasmon resonance with the optical wave. In 2009 a 4-channel compact biosensor has been presented by Piliarik et al. [12] and one year later Vala and his team developed a device capable of simultaneous detection of 10 analytes [13]. The beam reflected by the grating is collimated by lens and captured by a CCD camera that evaluate the average over 50 frames by onboard electronics and transmits it via USB. Although featuring both a high resolution ($3x10^{-7}$ RIU and $6x10^{-7}$ RIU respectively when the RI index change of a NaCl water solution is 0.00312 in both cases) and compact size they need an external PC for elaborating images.

In the last years the significant progress in nanotechnology have led to remarkable results in the study of Localized Surface Plasmon Resonance (LSPR) phenomenon. The use of nano-scaled metallic structures provides a new route to overcome the limitations described before, allowing to perform SPRi (see Section 3).

We describe here the design, implementation and characterization of a new SPR imaging biosensor based on LSPR with a novel efficient analysis method overcoming temporal light variations and noise. This approach considers two control regions in the sensible area of the nanostructured

biochip and one macro area where the sample will be injected. The entire system doesn't require the use of external PC for the elaboration and does not have moving parts or additional temperature control sensor.
The goal of our research is to develop a palm-size low-cost instrumentation that allows relatively untrained user to perform rapid multiple simultaneous assays outside of specialized laboratory for a wide range of applications.

III. Principles of Local Surface Plasmon Resonance and imaging

Surface plasmons (SP) are longitudinal charge density waves along the interface between a metal and a dielectric [14]. When a P-polarized beam radiates the interface between the media at a specific resonant angle and the wave component parallel to the surface matches the wavevector evanescent component of the plasmonic mode, it can resonantly couple and excite the mode. The coupling to the plasmonic mode results in an energy loss and then in an intensity reduction of reflected light.
In a SPR biosensor, receptor molecules are chemically immobilized on the metal surface. Therefore, when the target substance diluted in a liquid sample reacts with the receptor, a change in the resonance conditions is observed. Analyte/receptor reaction produces a refractive index change close to the surface which can be related to analyte concentration [14]. Typically, SPR system maximum sensitivity is about 10^{-7} RIU [15].
The rapid development of nanotechnologies, stimulated the study of the physical properties of metallic nanostructuctured, whose optical response can be exploited to improve surface plasmon resonance technique.
Metallic nanostructures support charge density oscillations called *Localized Surface Plasmons* (LSPs). Metal conduction electrons can be excited through a light beam to a collective oscillation state with a specific resonant frequency that depends on the size, shape, composition, dielectric properties of the material and spatial distributions of nanoparticles [16, 17]. Haes and Van Duyne, in 2004, demonstrated that LSPR has a sensitivity which is approximately equivalent to traditional SPR techniques. As a consequence, SPR systems can be miniaturized without significant sensitiveness loss.
The strongly dependence on the type of metal, shape and size of nanostructures, pushed researchers to identify the optimal design improving optical sensing.
In the last 10 years, significant results have been reached: basically, two possible approaches can be exploited.
In the first, metallic nanoparticles are immobilized on a glass substrate, while in the second one a glass substrate is covered with a metal thin film in which a periodic array of nanoholes is carried out [18]. A significant difference between these structures was explored in 2009 by Parsons et al. who demonstrated that the optical response of nanoparticles is independent of interparticle separation, while the response in periodic subwavelength nanohole arrays is largely dependent on interhole separation [19]. In particular, experiments highlighted that the peaks of transmittance spectrum were correlated to the period between nanoholes and that the amount of transmitted light was greater than that predicted by the classical theory. This phenomenon is called *Extraordinary Optical Transmission* (EOT) effect [20].
The possibility of producing nanostructured plasmonic surfaces with versatile and simple techniques such as lithographic one, makes this approach more compatible with the SPR imaging method.
An imaging SPR biosensor allows to detect multiple types of molecules simultaneously on a single chip. The binding interaction between the receptor substances attached to the metal surface and the molecules to be detected is monitored by a CCD or CMOS camera. The simplicity of the entire structure with no moving parts and the use of nanoholes array structure allows to overcome the difficulty to develop a multi-channel efficient biosensors for high-throughput analysis [18].

IV. Instrument design

According to the previously described state-of-the-art, our device combined the high performance of nanohole biochip with an embedded elaboration system to realize a new portable SPRi biosensor.

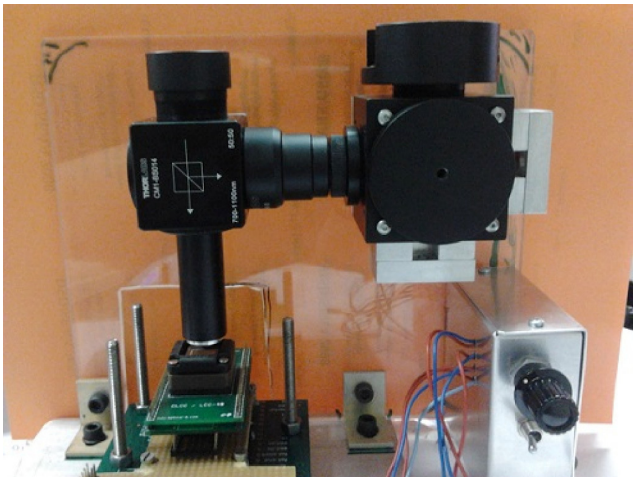


Fig 1. The biosensor prototype

The developed prototype showed in figure 1 is composed by two main parts: a bio-sensing surface with optical set up and an electronic unit for images acquisition and data elaboration.

A. Bio-sensing device

In a typical EOT biosensor, a collimated light beam is focused on the sensing region and the reflected light is captured by a spectrometer.
Many researches about EOT sensors performed by [21, 22, 23] demonstrated that their refractive resolution measured is similar to conventionally SPR system sensitivity (~10^{-6} RIU).

For our prototype we used a nanostructured biochip realized by [2] in 2011. Its sensitive surface is composed of an nanoholes array embedded in a gold film [2].
The chips were produced through a colloidal lithographic technique: a glass substrate was covered by plasma polymerized poly-acrylic acid via plasma enhanced chemical vapour deposition (PE-CVD) and a subsequent layer of polystirene beads (PS) are deposited on the top of the ppAA. In order to form a grating structure of a regularly spaced pillars, the layers were exposed to oxygen plasma *etching*. A gold layer is deposited on the nanostructure by vapour deposition to fill-in the gaps between pillars. Then, the residual PS mask is removed using lift-off by an ultrasonic bath in ultra-pure water. The obtained biochip features periodic gold cavities with shapes that widen to their bottom. The opening width is in the range of 50-250 nm, the bottom of 100 to 450 nm with a cavity periodicity of 200 to 1000 nm [2].
A recent work performed by Giudicatti et al. shows that this particular asymmetric pillars geometry increased the electric field in the cavities where the analytes receptors are located, contributing to strengthen the EOT effect and increasing the sensitivity in the order of 10^{-5} RIU [2].
Furthermore, the reflectance measured from the glass side of the biochip is sensitive to the refractive index at the opposite side. This characteristic allows to measure the optical

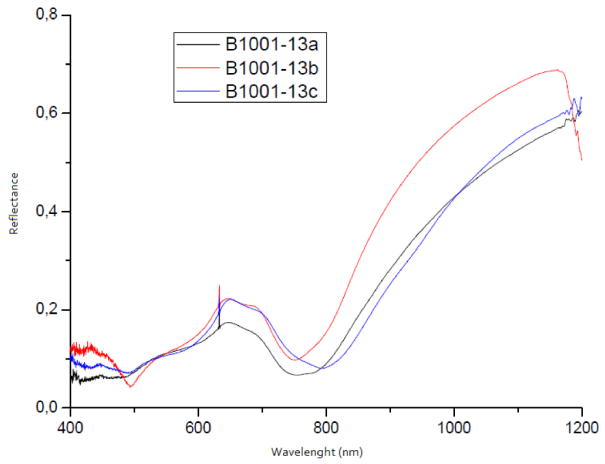


Fig 2. Reflectance spectra of different biochip used on trials. The minimum of reflectance is located in the 700-850 nm region.

response without expensive optical apparatus and collimated light source [24].
We studied the biochip reflectance spectra to identify the correct signal region where the resonance occurs. The trend showed in figure 2 highlights that the sensible region is located between 700 and 850 nm. Accordingly to this study we chose a LED light source (Vishay TSHG8200 IR LED) with 830 nm wavelength peak.
The light beam passes through one cube beamsplitter (CM1-BS015 by Thorlabs Inc.) and an achromatic lens (AC127-025-b by Thorlabs Inc.) to ensure the proper uniform irradiation of the sample minimizing aberrations. Run-time, the sensor is orientated so as the light beam illuminates the cavities from the widest bottom side while the detector captures the reflected light as shown in fig. 1.
The simple lithographic techniques together with no complex optics required allow the biochips mass production, providing an efficient solution for low-cost SPR biosensors fabrication.

B. *Electronic unit*

The electronic platform is composed by the image detector and the elaboration system. For capturing the reflected light of the biochip surface, we selected an APS CMOS. Since 1995, the increasing interest in CMOS image sensors was related to their design flexibility, reduced costs and low power consumption. The Active Pixel Sensor (APS) CMOS differs from the traditional passive CMOS for its robustness to noise due to the pixel amplification. Studies demonstrated that APS CMOS performance is comparable to their CCD equivalent [25]. We choose the MT9M001C12STM (Aptina Corporation), a monochromatic 1024x1280 pixel APS CMOS with a quantum efficiency (QE) suitable for our purposes. A 10 bit onboard analog-to-digital converter (ADC) codifies each pixel directly to the elaboration unit responsible for image processing and controlling the sensor via I^2C protocol.
In particular this unit acquires each image by the sensor, elaborates them, shows the analysis results on a LCD display and stores all the information on a MicroSD. This kind of approach requires a low power consumption and low costs flexible architecture for the portability and robustness of the instrument. For these reasons we selected the ARM9 family (AT91SAM9260 Atmel Corporation). The processor is mounted on the SAM9-L9260 development board developed by Olimex Ltd. The board features 64 MB SDRAM and 512 MB Nand Flash, an Ethernet 100 Mbit controller and a SD/MMC card connector directly linked to the processor.

V. Software platform

The primary tasks of the realized software platform are to ensure the communication between all the devices components, run the analyte detection algorithm and interact with the end user for the trials configurations and results visualization. These services must be managed by a specific software designed to optimize the performance but easily modifiable to add or change functionalities.
Since many years Linux has become an efficient solution to perform these tasks. It is a Unix-like, modular and multitasking operating system that supports a wide range of devices and configurations [26]. Furthermore, the open-source code availability allows software designers to modify the code according to their own needs improving costs and efficiency.
A typical Linux operative system scheme is a monolithic kernel where user applications run in the user space and operate on a virtual address to protect the internal memory.

The users program cannot directly access the system hardware but can request services to kernel by primitive system calls.
A specific *bootloader* at the starting time, initializes the hardware, defines the memory space map, enables the MMU (Memory Manage Unit) if present, and configures the processor for loading the operating system's kernel.
In our case, the onboard bootloader U-boot (http://www.denx.de/wiki/U-Boot/SourceCode with the Olimex board patches available on https://www.olimex.com/Products/ARM/Atmel/_resources/u-boot-olimex-patches-20090717.tgz) has been used to initialize the ARM9 processor.

A. Linux-based Operating System

The basis of our prototype is a minimal operating system called SSW realized by A. Rubini in 2010 and release under GNU General Public License (GPL). It is a small Linux-based operating system derived from THOS (www.gnudd.com/wd/thos.pdf). It is conceived for several family processors that we modified and extended to include the biosensor configuration and management.
The SSW operating system uses ARM9_v5 architecture composed by the MMU initialization, exceptions handlers, a simple I/O model, the module interface, and the tasks scheduler.
The Memory Management Unit controls the memory access and translates virtual to physical addresses. In our operating system the 4 GB virtual memory is divided in 1 MB sections with no cache [27].
The ARM exceptions are implemented through a vector table with branch instructions to specific code that saves registers and performs the right operations. The seven types of exceptions supported drive the processor in specific privileged modes accessing to different registers, stack and handler.
The kernel modules are initialized through a suitable initcall routine. At the system boot, modules set up is performed according to a specific hierarchy. At first, the modules with no dependencies like the serial port and the interrupt controller. Then, the architecture modules like the memory controller and timers, the peripheral devices and lastly the tasks modules.
The interface to modules is defined by two suitable macros *request()* and *provide()*.
The application programs that use modules can be activated periodically or not. A simple Round Robin scheme was implemented for the tasks scheduling. The tasks are kept in two doubly linked lists: one for the running ones and one for the idle ones. The initialization function prepares the first set up parameters (activation time, periodic execution or not, other parameters) and put the task in the idle list, from which it will be selected according to the Round Robin scheduling. This approach is a simple timeline-based scheduling where the system design chooses the activation time and period for each task. The scheduler temporization is implemented using timer interrupts. Specifically, a processor timer is programmed to generate an interrupt signal every 10 ms.

B. Biosensor device drivers and application

We extended the previously described operative system with an hierarchical collection of modules and drivers that use the simple GPIO pins to peripherals management and an application program to allow the biosensor utilization. The AT91SAM9260 configuration drivers, the acquisition interface for the image sensor, the LCD display and SD/SDHC modules are implemented using the minimal hardware and configuration required. This approach assures the software reusability with different hardware or other Linux-based operating systems with the possibility to easily add or change functionalities.
The biosensor application program realized configure the initial parameters of the biosensor, sets the trials periodicity and the number of images acquired, invoke the modules needed and provide to the end users the analysis results.
The main part of the described software platform is the communication management between the board, the image sensor, the display and the external SD memory.
The image sensor device uses I^2C protocol to read and write the internal sensor registers. These registers control various features like black level calibration, integration time for the pixels and readout modes of the sensor. The data output is controlled by three different signals: the pixel clock, the line valid and the frame valid.
The image data is read out sequentially. Each 10 bit pixel is ready on the falling edge of the pixel clock signal when both the line valid and frame valid are activated.
We developed specific modules for the images acquisition using the GPIO processor interface via a simple PWM signal. Each image is elaborated in about 4 seconds by the application software that makes available on LCD display the trial results and stores on SD memory in 14 seconds. A FAT16 partition is created on external memory during the application software boot to neatly memorize them on specific files.

VI. Analytes detection method

The goal of the biosensor is to provide a robust detection of the presence and amount of the target analyte on a liquid sample. This required a fast and efficient identification algorithm. The conducted research about the raw pixels of the images has highlighted temporal fluctuations of the light intensity acquired from the sensor that significantly influences the analysis results. In particular, the main effect is an upward trend of the pixels grey level average observable in figure 3.
Experiments conducted with different black level calibration, integration times and LED sources, demonstrated that not homogeneous light diffusion and photons accumulation on the sensor during long time exposition are the major causes of this phenomenon.
The study also revealed a saturation level not time-predictable, but affecting each image region.

To overcome the instability we used the pixel average ratio between specific image areas.
The approach is to normalize the grey level pixel average of the area where the antibodies are deposited, with the average of an external area of the same surface designed as control region where there is no active molecules.
With this method when the antibody-analyte reaction takes

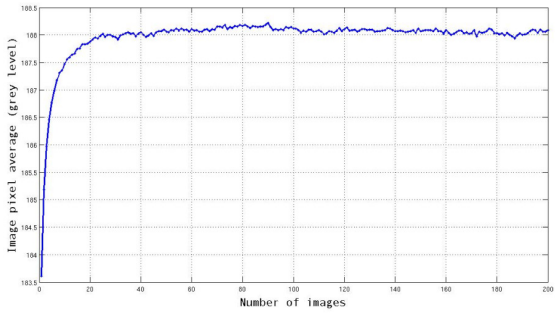


Fig 3. Gray levels pixel average measured during 200 images acquisition without antibodies on the biochip surface. After 50 images the light intensity reaches a saturation level.

place, the different light intensity measured is only influenced by the LSPR phenomenon without light aberrations.
Figure 4 a) shows the pixel average ratio of three regions of the biochip surface without immobilized antibodies. Two areas (2 and 3) are used as control regions and in the third distilled water was injected after 10 images acquired. The normalization applied eliminates the previously described

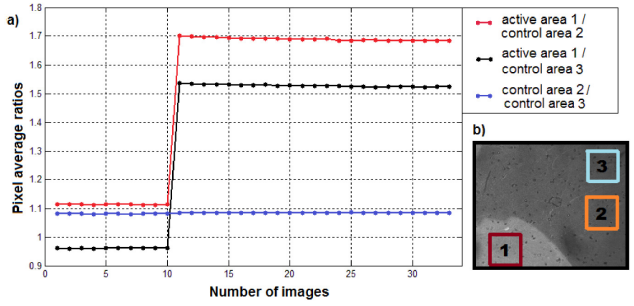


Fig 4. (a) Trend of pixel average ratio between three biochip regions without antibodies. After the tenth image distilled water was injected on the area 1; (b) 128x128 pixels areas used for the experiment. Areas 2 and 3 are control region.

instability and the ratios values are perfectly constant before and after the water injection.
It can be also seen the different average ratios measured are independent of each other since relative to biochip regions with different illumination. All the relevant information is included on the deviation before and after the water injection that corresponds to refractive index change due to LSPR phenomenon.

On this basis, we used the time derivative of average ratio as a suitable measurement of the refractive index variation on the biochip surface. Fig. 5 illustrates the average ratios and time derivatives of three areas, two control regions (area 1 and area 2) and a sample area (area 3). Distilled water was used as reference baseline, then (after 14 acquired images) a 5% glycerol solution was injected on the sample area.
The figure shows the time derivative change due to the refractive index variation on the sample area. Peak position and ratio increase of course correspond perfectly.
Such as in previous case (injection on area 1), we observed a change in the measured ratio according to refractive index variation of the biochip surface in contact with the sample.
The developed algorithm measures the amplitude of the time derivative peak that is compared with a specific refractive index calibration curve in the internal memory of the processor. The obtained value is used to define the target analyte concentration in a sample.

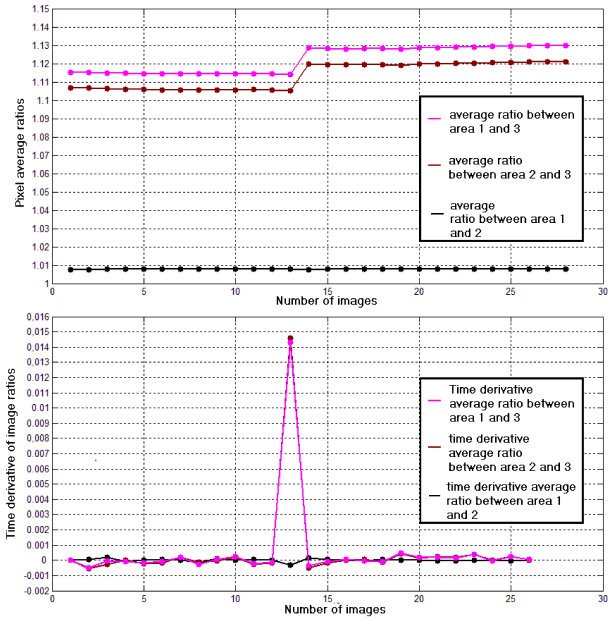


Fig 5. Pixel average ratio and time derivative when the sample area 3 is filled with distilled water (refraction index = 1.333) and then with a glycerol solution with concentration 5% (refraction index = 1.339).

VII. Results and Discussion

A. Sensitiveness of the biosensor

In order to evaluate the biosensor sensitiveness to changes in refractive index of bulk solutions and define a preliminary calibration curve for the algorithm developed, we prepared glycerol solutions with different concentrations (0.2% - 5%).
The biochip surface is subdivided in three rectangular areas (150x800 pixels), two taken as control regions and one selected region where the solutions were injected.
Distilled water is flowed on selected region as stable baseline reference for a few number of images (5-10) and

then the glycerol solution was introduced. The acquisition session relative to each concentration took 20 minutes.
Fig. 6 shows the linear correlation between the peak amplitude of the time derivative of average gray level pixel ratio at the various glycerol concentrations.

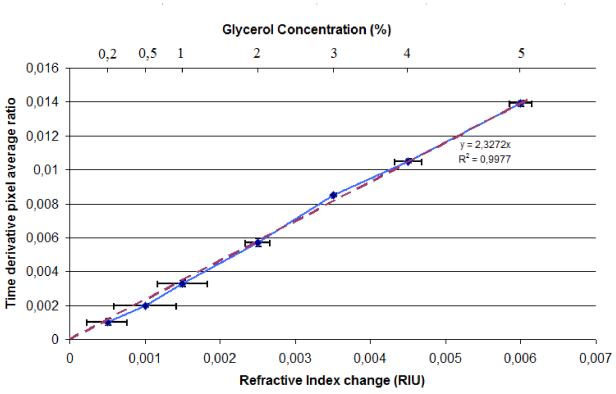


Fig 6. Trend of the time derivative pixel average ratios vs. refractive index change. Distilled water (n = 1,333) was assumed as reference. Biochip model B1001-13b was used.

Each experiment was repeated three times to assure measurements repeatability and the pixel average values are filtered with a moving average filter to eliminate outliers during the acquisitions.
The refractive index change was measured as the difference between the refractive index of distilled water and the solutions ones measured with an Abbe refractometer.
The correlation coefficient (R^2) of standard y-intercept and the measured data was 0.9977.
The variation of the response with respect to small changes in refractive index is caused by various factors as the detector noise and the LED fluctuations. The CMOS image sensor parameters as black level calibration and properties as the quantum efficiency at the wavelength used, may cause statistical fluctuations in the quantity of light captured. In the same way the electrical properties of the LED may affect the measurement.
Fig. 6 shows also a resolution limit for bulk refractive index respect to the baseline noise (measuring the standard deviation), equal to ~10^{-5} RIU for a glycerol concentration of 0.2%, which corresponds to the theoretical detection limit of the biochip [24]. This value increases as the concentration to 10^{-6} RIU for a glycerol concentration of 4% that corresponds to 10^{-3} Refractive Index change. This is sufficient in many fields and biodetection applications.
Furthermore this novel method allows to exploit the entire biochip surface for achieving multiple analytes detection, by simply distributing different antibodies spots and control areas along the surface.

B. *Future work*

Future work will involve tests of biological samples with antibodies immobilized on the biochip surface. The goal is to monitoring the adsorption and evaluate the real sensitivity limit for small molecule analyte detection.
Another important future step concerns the integration of the device with a microfluidic cell for the injection of the sample and the connection with a touch-screen monitor for display the signal response for the reaction kinetics observation.

VIII. Conclusions

The multiparametric SPRi biosensor described here provides efficient solution to the limit of the available SPR instrumentation, in terms of portability, power consumption, low-cost (about 1000$) and simplicity design. This device offers a real-time analytes detection useful for a large number of applications as medical diagnostic, monitoring of food allergens, toxins and pathogens detection in water and soils without the aid of specialized research laboratories.
A simple embedded system is used to elaborate images of a nanostructured biochip surface irradiated by a IR LED. Specific modules and device drives for Linux-based operating system have been developed to manage the biosensors components.
In particular, the pixels captured by a CMOS sensor are sent to an ARM9 processor via its GPIO interface and elaborated with a novel detection method. This new approach uses the time derivative pixels average ratios to identify change of refractive index on the biochip surface. The analysis results are available on LCD display in about 14 seconds and all the information are stored in an external SD memory.
The sensitivity for bulk refractive index changes is also measured to test the biosensor potential. The results for glycerol solutions at different concentrations indicate a best resolution of 10^{-5} RIU, suitable in several applications and fields. This detected biosensor response corresponds to the biochip theoretical response estimated by [24].
The device peculiar characteristics and the results achieved so far highlight a promising direction for a massive use for on-site analysis in thirdworld countries.

Acknowledgments

The authors are very thankful to prof. Alessandro Rubini for his help during the development of the ARM9 modules.

References

[1] Ramanavièius A., Herberg F. W., Hutschenreiter S., Zimmermann B., Lapënaitë I., Kaušaitë A., Finkelšteinas A., Ramanavièienë A., “Biomedical application of surface plasmon resonance biosensors (review) ”, *Acta Medica Lituanica*, 2005, Vol. 12, No. 3, pp. 1-9.

[2] Valsesia et al., 2013, “SPR sensor device with nanostructure”, European Patent Application 11174058.5, 16/01/2013.

[3] Hoa X. D., Kirk A. G., Tabrizian M., “Towards integrated and sensitive surface plasmon resonance biosensor: A review of recent progress”, *Biosensor and Bioelectronics* , 2007, No. 23, pp. 151-160.

[4] Chinowsky T. M., Quinn J.G., Bartholomew D.U., Kaiser R., Elkind J.L., “Performance of the Spreeta 2000 integrated surface plasmon resonance affinity sensor”, *Sensor and Actuators B*, 2003, No. 91, pp. 266-274.

[5] Koel M., Kaljurand M., “Green Analytical Chemistry”, 2010, RSC Publishing, ISBN: 978-1-84755-872-5.

[6] Chinowsky T. M., Soelberg S. D., Baker P., Swanson N. R., Kauffman P., Mactutis A., Grow M. S., Atmar R., Yee S. S., Furlong C. E., "Portable 24-analyte surface plasmon resonance instruments for rapid, versatile biodetection", *Biosensor and Biolectronics,* 2007, Vol. 22, pp. 2268-2275.

[7] Hu J., Hu J., Luo F., Li W., Jiang G., Li Z., Zhang R., "Design and validation of a low cost surface plasmon resonance bioanalyzer using microprocessor and a touch-screen monitor", *Biosensor and Bioelectronics*, 2009, Vol. 24, pp. 1974-1978.

[8] Cai H., Li H., Zhang L., Chen X., Cui D., "Portable Surface Plasmon Resonance Instrument based on a Monolithic Scanner Chip", *3rd International Conference on Biomedical Engineering and Informatics*, 2010, pp. 1153-1155.

[9] Shin Y.-B., Min Kim H., Jung Y., Chung B.H., "A new palm-sized surface plasmon resonance (SPR) biosensor based on modulation of light source by rotating mirror", *Sensor and Actuators B: Chemical*, 2010, No. 150, pp.1-6.

[10] Weichert F., Gaspar M., Timm C., Zybin A., Gurevich E.L., Engel M., Müller H., Markwedel P., "Signal Analysis and Classification for Surface Plasmon Assisted Microscopy of Nanoobjects", *Sensor and Actuators B: Chemical*, 2010, No. 151 pp. 281-290.

[11] Roh S., Chung T., Lee B., "Overview of the Characteristics of Micro and Nano Structured Surface Plasmon Resonance Sensors", *Sensors*, 2011, No. 11, pp. 1565-1588.

[12] Piliarik M., Vala M., Tichý I., Homola J., "Compact and low-cost biosensor based on novel approach to spectroscopy of surface plasmons", *Biosensor and Biolectronics*, 2009, No. 24, pp. 3430-3435.

[13] Vala M., Chadt K., Piliarik M., Homola J., "High-performance compact SPR sensor for multi-analyte sensing", *Sensor and Actuators B: Chemical*, 2010, No. 148, pp. 544-549.

[14] Green J. R., Frazier A. R., Shakesheff M. K., davies C. M., Roberts J. C., Tendler J.B. S., "Surface plasmon resonance analysis of dynamical biological interactions with biomaterials", *Biomaterials*, 2000, Vol. 21, pp. 1823-1835.

[15] Homola J., "Surface Plasmon Resonance Sensor for Detection of Chemical and Biological Species", *Chemical Reviews*, 2008, Vol. 108, pp. 462-493.

[16] Hutter E., Fendler H. J., "Exploitation of Localized Surface Plasmon Resonance", *Andvanced Material*, 2004, Vol. 16, No. 19, pp. 1685-1706.

[17] Zhao J., Zhang X., Yonzon C. R., Haes J. A., Van Duyne P. R., "Localized surface plasmon resonance biosensor", 2006, *Future Medicine Ltd*, No.1, pp. 219-228.

[18] Byun K.-M., "Development of nanostructured plasmonic substrates for enhanced optical biosensing", *J. Opt. Soc. Kor*., 2010, No. 2, Vol. 14, pp. 65-76.

[19] Parsons J., Hendry E., Burrows C.P., Auguié B., Sambles J. R., Barnes W., L., "Localized surface-plasmon resonance in periodic nondiffracting metallic nanoparticle and nanohole arrays", *Physical Review B*, 2009, Vol. 79, Issue 7, 073412.

[20] Lesuffleur A., Im H., Lindquist N. C., Lim S. K., Oh S., "Plasmonic Nanohole Arrays for Real-Time Multiplexing Biosensing", *Biosensing*, 2008, Vol. 7035, 703504-1.

[21] De Leebeeck A., Kumar S. L. K., De Lange V., Sinton D., Gordon R., Brolo A. G., "On-Chip SurfaceBased Detection with Nanohole Arrays", *Analytical Chemistry*, 2007, Vol. 79, No. 11, pp. 4094-4100.

[22] Eftekhari F., Ferreira J., Santos M. L. J., Escobedo C., Brolo A. G., Sinton D., Gordon R., "Development of Portable SPR Sensor Devices Based on integrated Periodic Arrays of Nanoholes", *Optical Sensors*, 2009, Vol. 7356, 73560C-1.

[23] Im H., Sutherland J. N., Maynard J. A., Oh S., "Nanohole-based SPR Instruments with Improved Spectral Resolution Quantify a Broad Range of Antibody-Ligand Binding Kinetics", *Analytical Chemistry*, 2012, Vol. 84(4), pp. 1941-1947.

[24] Giudicatti S., Valsesia A., Marabelli F., Colpo P., Rossi F., "Plasmonic resonance in nanostructured gold/polymer surfaces by colloidal lithography", *Physica Status Solidi A*, 2010, Vol. 207, No. 4, pp. 935-942.

[25] Carlson, B.S., "Comparison of modern CCD and CMOS image sensor technologies and systems for low resolution imaging", *Sensors*, 2002, Proceedings of IEEE, Vol.1, pp. 171- 176.

[26] Hu J., Zhang G., "Research transplanting method of embedded linux kernel based on ARM platform", *International Conference of Information Science and Management Engineering (ISME)*, Vol. 2, pp. 35-38, 7-8 Aug. 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 627–633

A SCORE-BASED PACKET RETRANSMISSION APPROACH FOR PUSH-PULL P2P STREAMING SYSTEMS*

Muge Sayit, Erdem Karayer, Kemal Deniz Teket, Yagiz Kaymak, Cihat Cetinkaya, Sercan Demirci, and Geylani Kardas
Ege University, International Computer Institute, 35100, Izmir, Turkey
{muge.fesci, yagiz.kaymak, cihat.cetinkaya, sercan.demirci, geylani.kardas}@ege.edu.tr, {erdemkarayer, denizkemal}@gmail.com

ABSTRACT

In this paper we propose an inference based packet recovery technique which considers past scores indicating retransmission success of the peers. Past scores are calculated by considering several parameters such as requested packets availability and round trip time. The importance of packets to be retransmitted is also considered in the proposed model. In order to obtain comparable results, we also implement a different retransmission approach similar to the models proposed in the literature. The ns3 simulations show that retransmission model increases the Peak Signal to Noise Ratio (PSNR) value even under high peer churn and limited resource index. Furthermore, score-based approach provides a decrease in reset counts and the number of duplicate packets, when it is compared to different retransmission approaches.

Keywords - peer-to-peer networks, live video streaming, packet recovery

1. INTRODUCTION

Peer-to-peer (P2P) video streaming systems represent one of the applications having a huge effect on the network traffic and there are remarkable live streaming systems proposed in the literature for both wired and wireless networks. Although most of these applications fall into two main categories as push-based and pull-based, several hybrid systems bringing the advantages of these two approaches have reported their success [1, 2]. In pull-based systems the nodes in the system send and receive video data chunks from one or more nodes [1] whereas each node has one parent and one or more children node in push-based systems [3]. In hybrid push-pull-based streaming; video data are divided into substreams. Nodes in the system connect one or more parents to receive these substreams and play video by combining them. During streaming, buffer maps, indicating the received blocks of each substream periodically, are exchanged between partner nodes in order to detect congestion and determine the candidate parents in case of parent re-selection.

We implement a hybrid push-pull system having the main properties of CoolStreaming [1] as to be used as an underlying framework. Although such systems have remarkable success, the system performance may degrade if there are not enough special-aimed nodes such as Content Delivery Network (CDN) nodes or super peers in the system. In this case, stream can be supported via retransmission of important packets. One may discuss if parent selection algorithms proposed for pull-based systems can be implemented in the selection of node which will send retransmission packets. Nevertheless, retransmission or packet recovery has more limited time to receive the requested packets when compared to the required time to receive packets from parents during streaming. The reason is that the packets received from the parents are generally obtained before their playout time since they are kept in buffer for a while whereas buffering time for retransmitted packet is relatively small. Since requested packets must be received in short time period, the selection of node sending retransmission packets should be realized considering more constrained time period.

In this work, we propose a new packet recovery technique based on retransmission. The contributions of this paper can be listed as follows: (*i*) We propose a pull-based retransmission model designing as to run combined with hybrid push-pull systems. It is shown that the performance of the hybrid system is significantly improved under high peer churn with the proposed model. (*ii*) The selection of the retransmission packets is based on the frame type and the selection is done in the sender side. (*iii*) Proposed model considers past behaviors of the peers, and the algorithm has low computational complexity and easy to implement.

The rest of the paper is organized as follows: In section 2, we give the summary of the related work. In section 3, we introduce the main contribution of this paper, namely score-based retransmission approach. The simulation results and conclusion are given in section 4 and 5, respectively.

2. RELATED WORK

Packet recovery techniques for missing packets can be classified into two categories, namely Forward Error Correction (FEC) and retransmission. FEC algorithms can

*This work is funded by the Scientific and Technological Research Council of Turkey (TUBITAK) Electric, Electronic and Informatics Research Group (EEEAG) under grant 111E022.

978-83-60810-53-8/$25.00

recover lost packets if the necessary number of packets is received. However, these techniques may not be useful in case of burst packet lost [4]. Furthermore, finding optimal FEC redundancy rate is a difficult problem due to the unestimated packet loss nature of P2P video streaming applications [5]. With the usage of retransmission techniques, there is no need to use redundant recovery packets introduced by FEC algorithms. Although there are well-known retransmission algorithms such as Automatic Repeat Request (ARQ) for classical server-client model [6], there are also some approaches proposed in the literature for retransmission in P2P video streaming systems [7, 8]. In [9], lost packets are obtained by retransmission at each hop from source to destination in the application overlay. This approach may cause retransmitted packets to reach lately to the destination nodes especially in leaf positions. A model in which parent nodes deciding retransmit some packets according to received NACK from children is proposed in [10]. Nevertheless, retransmission of missing packets from the same node, i.e. from the parent as proposed in [9, 10] has some disadvantages in case of parents do not have missing packets. In this case, requesting retransmission from a node other than the parent node provides higher continuity index [5, 11]. In [11], while selection of the node to request lost packets, the availability of the requested packets is not considered since this information is not available to the nodes in hybrid push/pull system. However, in pull based P2P video streaming systems, nodes have the information about the chunk availability and the selection of nodes to request lost packets can be done by taking into account this information [5]. On the other hand, when considering the high success of pull-based P2P video streaming systems, requesting missing packets from a set of nodes may improve the performance of the system. In [12], authors proposed a model for selecting this set of nodes. This selection is done by considering two criteria, the distance in terms of Round Trip Time (RTT) between requester node and the node retransmitting packets, and the position of the node retransmitting packets in the multicast tree.

This paper presents a work based on recovery of missing packets via pull-based retransmission. We prefer to use the term "retransmission parent" for the node which sends retransmission packets. Our study differs from the literature in several dimensions. First, we select more than one retransmission parents by considering the packets that they have. Second, after retransmission parent selection, we implement packet selection and give more priority to the retransmission of packets carrying I frames. Furthermore, we also propose a new algorithm to select candidate retransmission parents by considering their past behaviors, which provides increase in Quality of Experience (QoE) in terms of reset counts and decrease in message complexity.

In order to evaluate the performance of the proposed algorithm and to give the comparable results, we also implement a different retransmission approach similar to related studies proposed in the literature. The obtained results show that the performance of the proposed method exceeds the performance of the state-of-the-art solutions in terms of continuity index and total number of reset counts, as will be discussed in the following sections of this paper.

3. SCORE-BASED RETRANSMISSION

We used substream based, i.e. hybrid push-pull-based P2P live video streaming system as an underlying framework since it is reported that users can obtain higher continuity index than that of users in pure pull systems [1]. As mentioned before, each node in the system connects one or more parents to receive substreams in hybrid push-pull-based streaming. After joining the P2P system, the nodes obtain the list of online nodes in the system by communicating with the tracker server and construct partnership table by selecting a subset of these nodes. Parents of the nodes are selected from partnership table and parent selection is done according to the buffer maps of the partners, in other words, candidate parents. Each buffer map indicates the latest received packet of the related substream, thus if the video data are partitioned into the k substreams, buffer maps are the k-dimensional vectors. Buffer maps are periodically exchanged between partner nodes in order to watch the buffer condition of the partners. We embed additional data such as playout index, i.e. the current position of the video and the buffer indicating lost and existing packets between playout index and latest received packet into these messages.

In order to recover lost packets, a node must obtain them before their playout time. After streaming started, each node tries to recover lost packets by requesting them from their partners. An example scenario showing the buffer of a peer and the packets received from the substream parents and retransmission parent is given in Fig. 1. According to the figure, the peer is subscribed four parents to obtain four substreams. The packets numbered with 55 and 60 are not received from the parents hence these can be received from the retransmission parent. Note that 52^{nd} packet is not received and will not be requested from the retransmission parent since the playout time is passed for this packet. In packet recovery process two selections are made: selection of which partner(s) to request retransmission and selection of which packet(s) to request.

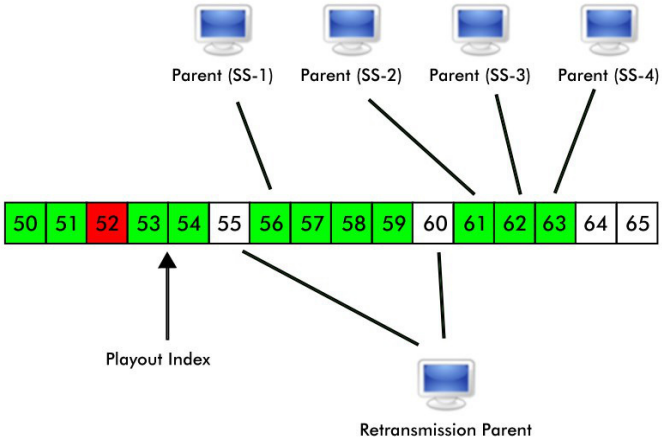


Fig. 1. An illustration of a retransmission scenario.

For retransmission parent selection, we propose a new approach based on previous scores of partners. A partner score is determined as the success ratio of received requested packets. Parent selection algorithm for determining candidate partners to request lost packets is given in Fig. 2.

$S_{requested\_packets}$: the set of packets which will be retransmitted

for each substream k
  for each node p_i in partnership table
    fullness_p_i = number of requested packets /
        number of requested packets existing in p_i's buffer;
  end for
  normalize fullness and RTT for all pi;
  for each node p_i in partnership table
    expected_score_p_i = α.fullness_p_i + (1-α).(1/RTT_p_i);
    expected_score_p_i = β.expected_score_p_i
        + (1-β).(previous_score_p_i);
  end for
  //rank all p_i according to their expected scores
  construct $\gamma(p)$; // ranked list
  while $|S_{requested\_packets}| > 0$
    p_{first} = first partner in $\gamma(p)$;
    S_p = the set of existing packets within requested packets
        in p_{first}'s buffer;
    add packets in S_p to the list representing requests from p_{first};
    $S_{requested\_packets} = S_{requested\_packets} - S_p$;
    calculate expected scores for new set of request packets and
        construct $\gamma(p)$;
  end while
end for

Fig. 2. Selection algorithm for retransmission parent.

Selection algorithm for retransmission parent starts with assigning fullness ratio of each partner in the partnership table. This value is calculated by examining the buffer map messages received from partners. In the next *for* loop, expected score for all candidate retransmission parents are calculated by considering three parameters, the fullness, the RTT value and previous expected score of the partners. Note that RTT is an important parameter to consider since requested packets need to be received before playout deadline. In the second step of the calculation of expected score, past scores of the partners is evaluated since even if a partner has high fullness and low RTT value, it may have limited available upload bandwidth or it may not be a stable node. For partners whose previous score is not calculated yet, this value is given as 0.5. After expected scores of each partner are calculated, the ranked list according to these values, $\gamma(p)$, is constructed. Requested packets which also exist at the first partner in the ranked list are put in a list representing the requested packets from the first partner in $\gamma(p)$. After the set of requested packets is reconstructed again, the expected scores are recalculated by considering new set of requested packets and $\gamma(p)$ is re-ranked. This process continues until retransmission parents are selected for all requested packets or remaining requested packets does not exist in the buffer of any partner.

Since the size of the partnership table is limited and quite small when compared with the size of the system, the space and the computational complexity of the algorithm given in Fig. 2 are negligible.

After request messages are sent, requester node waits for a period to receive the packets and to evaluate the performance of the retransmission parent for this session. This period is completed if the video playout index reaches to the playout time of the requested frame. In each buffer map exchange, retransmission parent selection is done again for the previously requested and not received packets. But, requester nodes may select the same retransmission parent with high probability since previously selected parent is not evaluated yet and still have the highest score. There are three

reasons for this choice of evaluation period, in order to evaluate the performance of the retransmission parent fairly, to decrease the number of request messages and to prevent the number of duplicate packets. With the completion of this period, the success ratio S_R is calculated by dividing the number of received packets to the number of requested packets and previous score of the partners is updated by the smoothing function given in (1).

$$previous\_score = \lambda . S_R + (1-\lambda) . previous\_score \quad (1)$$

Since the loss of I or P frames causes more distortion on display, sender node gives more priority to the packets containing I and P frames. In order to make nodes to detect which packet carries a part of an I or P frame, video server marks the packets according to its type and each node matches the received packet to its frame type.

4. SIMULATION

The simulations are implemented on ns3 [13] with the networks consisting of 50, 100, 150, 200 and 300 nodes. All the topologies used in simulations are generated randomly with BRITE topology generator [14] and Barabasi model [15]. Simulations are repeated several times in order to obtain averaged performance results.

There are one video server and one tracker server in the system. The video server has an upload capacity that provides the server to handle 20 percent of all peers in the system. Besides, we do not employ any CDN or super peer to support the video data dissemination. However 10 percent of the peers are selected as robust nodes which have higher bandwidth and tend to stay longer in the system. The tracker server is assigned to serve as the entry point of the streaming system. All newly joined peers connect to the tracker server first to request a random list of online peers to establish their initial connections.

In our simulations we use Foreman video sequence having QCIF resolution. The video is looped several times as to be used in 30 minutes length of simulations and encoded at 300 Kbps. We use frame copy as an error concealment method for missing frames. The original stream is divided into 4 substreams in each simulation in order to increase the potential suppliers of video. An exponential distribution which has an expected value of 1000 seconds for online and 400 seconds for offline periods is used to generate on/off intervals. Furthermore, we employ 10 percent of the nodes as free riders. The cumulative upload bandwidth distribution of all peers is given in Table I. We choose to send all messages including control and video over TCP in order to be able to connect all users even located behind a firewall. Retransmission requests are embedded to buffer map exchange messages, hence requests messages do not cause an extra load.

TABLE I. THE CUMULATIVE UPLOAD BANDWIDTH DISTRIBUTION

Percentage	Upload Bandwidth
10%	<50 Kbps
50%	<300 Kbps
90%	<1000 Kbps
100%	<3000 Kbps

In order to compare the performance of the proposed approach with the performance of other studies proposed in the literature, a different approach is also implemented for retransmission of missing packets. As proposed in [5], the selection of the retransmission parent is done randomly among the partners having requested packets and among the partners located close to the server, i.e. the position in the tree [12]. In this approach, if the node cannot receive the requested packets, it concludes that the current retransmission parent has not sufficient upload capacity and requests these packets from another node at the next buffer map exchange period. For this reason, we prefer to use the term “greedy-based” for this approach. In the simulations, buffer map messages are exchanged in every 2 seconds, hence if the requester node cannot receive retransmission packets within 2 seconds; it repeats the requests whereas in score-based retransmission, the nodes do not re-send request messages with high probability. Both greedy and score-based approaches give priority to the packets carrying I and P frames for fair comparison.

We measure PSNR, continuity index and total reset count as the video quality metrics. If a node consumes all data in its buffer, stops to receive video packets and cannot find suitable parent to connect, it resets itself, in other words, leaves the system and then joins immediately. The reset procedure causes the duration of the video playout for a while. As we observe, the time of this duration can change from 15 seconds to 20 seconds. The change on the average PSNR and the total reset count metrics are presented in Fig. 3 and Fig. 4 with respect to the change in network size. The PSNR of packet recovery techniques is higher than sole hybrid push-pull as expected. Although the greedy pull technique is slightly better than the score-based pull technique in terms of PSNR values, this difference is not easily observable at the end-user. In Fig. 4, an example frame received in hybrid push-pull, greedy-based and score-based approaches is given. As it can be seen from the figure, the distortion in the frame is similar in greedy and score-based approaches even the difference of the PSNR values of both approaches bigger than the difference of the averaged PSNR values given in Fig. 3.

The greedy pull technique has a higher control overhead and higher reset counts which can be seen in Fig. 5. The total reset counts are calculated by summing the number of resets of each node in the system. As it can be seen from the figure, the total reset count of proposed score-based pull packet recovery technique are always lower than sole hybrid

push-pull and greedy pull packet recovery techniques except for 50 peers. Furthermore, resets of the system implementing score-based retransmission stay almost identical in network size of more than 100 nodes. The greedy pull packet recovery technique causes receiving the most part of the packets via retransmission. Since stream parents have not sufficient residual upload bandwidth due to the over-requested packets, hence this technique has the highest total reset count.

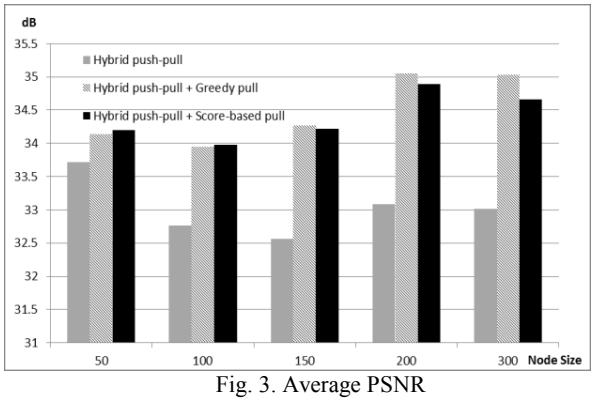
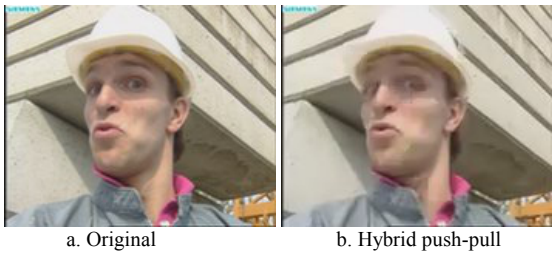


Fig. 3. Average PSNR



a. Original

b. Hybrid push-pull (PSNR: 31.21 dB)

c. Greedy (PSNR: 35.30 dB)

d. Score-based (PSNR: 34.66 dB)

Fig. 4. Example frames from the videos obtained in different P2P systems.

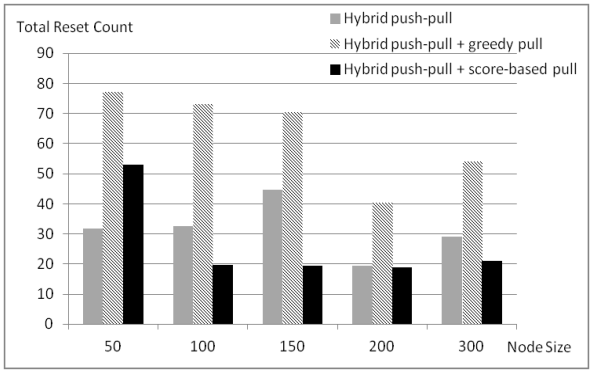


Fig. 5. Total Reset Count

In order to observe the inference performance of the score-based retransmission with respect to time, we measure two parameters, the ratio of the requested retransmission packets to the received retransmission packets and the change of continuity index, in Fig. 6 and Fig. 7, respectively. Since the greedy pull technique sends more request messages than the score-based pull, the score-based pull has higher reception ratio than the greedy pull technique on requested packets.

The graph given in Fig. 6 shows that the selection of retransmission parents is more successful in the score-based approach when compared to the greedy-based approach. This achievement provides the increase in continuity index by time as it can be seen in Fig. 7. In this figure, the graph shows that the inference process of large size of networks may be longer than that of small size of networks. Thus, it is expected that an increase in PSNR values for especially large size of networks such as network containing 300 nodes by time. Note that the most difference in PSNR values is observed between the greedy pull and the score-based pull for the network size of 300 nodes. This difference can be closed if the simulation lasts longer or if the nodes stay longer in the system. In Fig. 7, the observed change in continuity index for greedy-based retransmission shows a random pattern since it does not use any inference mechanism.

Finally, we give the comparative duplicate packet ratio values for greedy and score-based retransmission technique in order to indicate message complexity overhead in Fig. 8. The ratio is calculated by dividing the number of received packets to the number of requested packets. The graph shows that 50% of retransmitted packets are duplicate packets in the greedy-based approach.

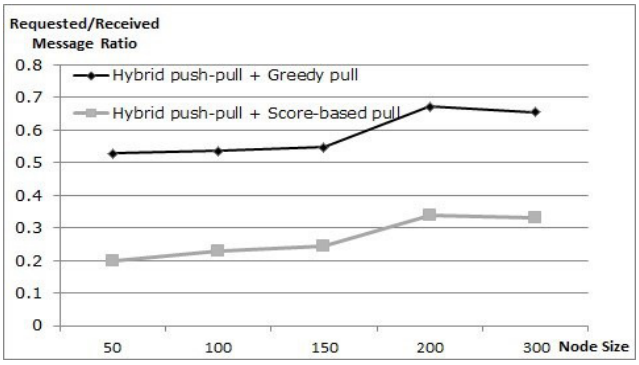


Fig. 6. Requested/Received Message Ratio

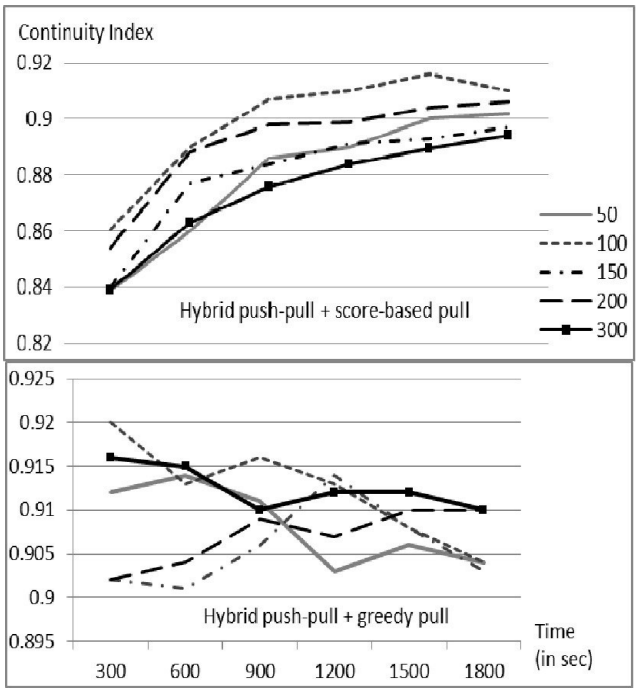


Fig. 7. Continuity Index

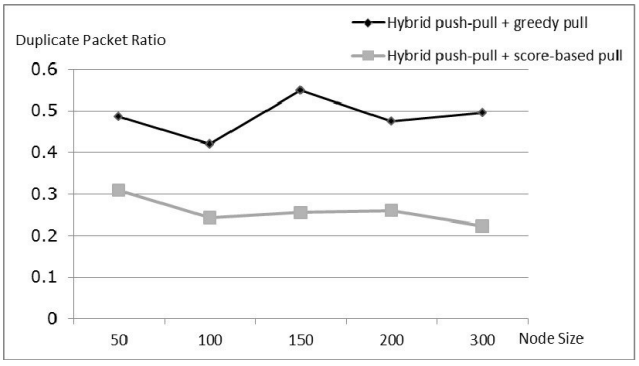


Fig. 8. Duplicate Packet Ratio

5. CONCLUSION

In this paper, we have introduced a new packet recovery technique based on retransmission for hybrid push-pull P2P live video streaming system. For selection of retransmission parent, an inference based approach has been discussed. The proposed retransmission strategies provides an increase in QoE parameters in terms of PSNR, total reset count and continuity index. Furthermore, the score-based retransmission approach achieves PSNR values similar to the greedy-based retransmission approach while the number of duplicate packets is relatively small. It is shown that the performance of the score-based approach increases by time. With the help of node clustering, it is also possible to improve the effectiveness of the proposed inference-based approach in P2P systems consisting high number of nodes.

We plan to implement the proposed score-based retransmission approach for the live P2P video streaming systems using different type of video codecs as our future direction. In this case, for example, some additional parameters such as packet priority or layer information may be considered during the selection process of the retransmission parent for scalable video codec.

6. REFERENCES

[1] S. Xie, B. Li, G.Y. Keung, and X. Zhang, “CoolStreaming: Design, Theory, and Practice,” IEEE Transactions on Multimedia 9, pp. 1661-1671, 2007.

[2] Wang F., Xiong Y., and Liu J, “mTreebone: A Collaborative Tree-Mesh Overlay Network for Multicast Video Streaming,” IEEE Transactions on Parallel and Distributed Systems, vol. 21, pp. 379-392, 2010.

[3] M. F. Sayıt, T. Tunalı, and A.M. Tekalp, “Resilient peer-to-peer streaming of scalable video over hierarchical multicast trees with backup parent pools,” Elsevier Signal Processing: Image Comm., vol. 27, pp. 113-125, 2012.

[4] M.-F. Tsai, C.-K. Shieh, T.-C. Huang, and D.-J. Deng, “Forward-Looking Forward Error Correction Mechanism for Video Streaming over Wireless Networks,” IEEE Systems Journal, vol. 5, no. 4, December 2011.

[5] H. Wehbe, G. Babonneau, and B. Cousin, “Fast Packet Recovery for PULL-Based P2P Live Streaming Systems,” The Second International Conference on Advances in P2P Systems (AP2PS'10), October 25 - 30, 2010, pp. 7-13, Florence, Italy, 2010.

[6] L. Rizzo, “Effective Erasure Codes for Reliable Computer Communication Protocols,” ACM SIGCOMM Computer Communication Review, vol. 27, pp. 24–36, 1997.

[7] E. Setton, P. Baccichet, and B. Girod, “Peer-to-Peer Live Multicast: A Video Perspective,” Proceedings of the IEEE, vol. 96, no. 1, pp. 25–38, 2008.

[8] O. Abboud, K. Pussep, A. Kovacevic, K. Mohr, S. Kaune, and R. Steinmetz, “Enabling resilient P2P video streaming: survey and analysis,” Multimedia Systems, pp. 1-21, 2011.

[9] B. Akabri.; H. Rabiee, and M. Ghanbari, "Packet Loss Recovery Schemes for Peer-to-Peer Video Streaming," Third

International Conference on Networking and Services (ICNS), IEEE Computer Society, pp. 94, 2007.
[10] P. K. Hoong and H. Matsuo, “Push-pull incentive-based P2P live media streaming system,” WTOC, vol. 7, no. 2, pp. 33-42, February 2008.
[11] C. Liu, K. Wang, and Y. Hsieh, "Efficient push-pull-based P2P multi-streaming using application level multicast," Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium, pp. 2586-2590, 26-30 Sept. 2010.
[12] X. Jin, H.-S. Tang, S.-H. G. Chan, and K.-L Cheng, "Deployment Issues in Scalable Island Multicast for Peer-to-Peer Streaming," IEEE Multimedia, vol. 16, no. 1, pp. 72-80, January-March 2009.
[13] Network Simulator 3, http://www.nsnam.org.
[14] “BRITE Topology Generator”, http://www.cs.bu.edu/brite.
[15] A.L. Barabasi and R. Albert, “Emergence of scaling in random networks,” Science 286, pp. 509–512, October 1999.

Education, Curricula & Research Methods

ECRM is a FedCSIS conference area aiming at interchange of information, ideas, new viewpoints and research undertakings related to university education and curricula as well as recommended methods of doing research in all computing disciplines, i.e. computer science, computer engineering, software engineering, information technology, and information systems. This area spans typical FedCSIS events (conferences, workshops, etc.) with rigorous paper submissions and review processes as well as panels, PhD and research consortia, summer schools, etc. Events that constitute ECRM are:

- DS-RAIT'13 - Doctoral Symposium on Recent Advances in Information Technology
- ISEC'2012 - Information Systems Education & Curricula Workshop

Doctoral Symposium on Recent Advances in Information Technology

THE first international Doctoral Symposium on Recent Advances in Information Technology (DS-RAIT 2013) will be held in Cracow (Poland) on September 8-11, 2013 as a satellite event of the Federated Conference on Computer Science and Information Systems (FedCSIS 2013) and Education, Curricula & Research Methods (ECRM 2013) conference.

The aim of this meeting is to provide a platform for exchange of ideas between early-stage researchers, in Computer Science, PhD students in particular. Furthermore, the symposium will provide all participants an opportunity to get feedback on their studies from experienced members of the IT research community invited to chair all DS-RAIT thematic sessions. Therefore, submission of research proposals with limited preliminary results is strongly encouraged.

Besides receiving specific advice for their contributions all participants will be invited to attend plenary lectures on conducting high-quality research studies, excellence in scientific writing and issues related to intellectual property in IT research. Authors of the two most outstanding submissions will have a possibility to present their papers in a form of short plenary lecture.

Topics

DS-RAIT 2013 invites the submission of papers on all aspects of Information Technology including, but not limited to:

- Automatic Control and Robotics
- Bioinformatics
- Cloud, GPU and Parallel Computing
- Cognitive Science
- Computer Networks
- Computational Intelligence
- Cryptography
- Data Mining and Data Visualization
- Database Management Systems
- Expert Systems
- Image Processing and Computer Animation
- Information Theory
- Machine Learning
- Natural Language Processing
- Numerical Analysis
- Operating Systems
- Pattern Recognition
- Scientific Computing
- Software Engineering

Honorary Conference Chair

Kacprzyk, Janusz, Systems Research Institute of the Polish Academy of Sciences, Poland

Event Chairs

Golunska, Dominika, Cracow University of Technology, Poland

Kowalski, Piotr Andrzej, Systems Research Institute of the Polish Academy of Sciences, Poland

Lukasik, Szymon, Systems Research Institute, Polish Academy of Sciences, Poland

Program Committee

Arabas, Jaroslaw, Warsaw University of Technology, Poland

Atanassov, Krassimir T., Bulgarian Academy of Sciences, Bulgaria

Balazs, Krisztian, Budapest University of Technology and Economics, Hungary

Castrillon-Santana, Modesto, University of Las Palmas de Gran Canaria, Spain

Charytanowicz, Malgorzata, Catholic University of Lublin, Poland

Corpetti, Thomas, University of Rennes, France

Courty, Nicolas, University of Bretagne Sud, France

Fournier-Viger, Philippe, University of Moncton, Canada

Gil, David, University of Alicante, Spain

Hu, Bao-Gang, Chinese Academy of Sciences, China

Koczy, Laszlo, Szechenyi Istvan University, Hungary

Kulczycki, Piotr, Systems Research Institute, Polish Academy of Sciences, Poland

Lilik, Ferenc, Szechenyi Istvan University, Hungary

Mesiar, Radko, Slovak University of Technology, Slovakia

Noguera i Clofent, Carles, Academy of Sciences of the Czech Republic, Czech Republic

Petrik, Milan, Masaryk University, Czech Republic

Tormasi, Alex, Szechenyi Istvan University, Hungary

Yang, Yujiu, Tsinghua University, China

Zadrozny, Slawomir, Polish Academy of Sciences, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 639–642

Inexact Newton method as a tool for solving differential-algebraic systems

Paweł Drąg
Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
Janiszewskiego 11/17, 50-372, Wrocław, Poland
Email: pawel.drag@pwr.wroc.pl

Krystyn Styczeń
Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
Janiszewskiego 11/17, 50-372, Wrocław, Poland
Email: krystyn.styczen@pwr.wroc.pl

***Abstract*—The inexact Newton method is commonly known from its ability to solve large-scale systems of nonlinear equations. In the paper the classical inexact Newton method is presented as a tool for solving differential-algebraic equations (dae) in fully-implicit form $F(\dot{y}, y, t) = 0$. The appropriate statement of dae using the backward Euler method makes the possiblity to see the differential-algebraic system as a large-scale system of nonlinear equations. Because a choice of the forcing terms in the inexact Newton method significantly affects the convergence of the algorithm, in the paper new variants of the inexact Newton method were presented and tested. The simulations were executed in Matlab environment using Wroclaw Centre for Networking and Supercomputing.**

***Index Terms*—differential-algebraic equations, systems of nonlinear equations, inexact Newton method.**

I. Introduction

DIFFERENTIAL-algebraic equations (dae) play a key role in control science and engineering [16], [17]. Describing the system with equations that incorporate dynamics and conservation laws, creates new opportunities for the development of the numerical methods and has a direct application in the industry [2], [3]. Design and control of chemical reactors and motor vehicle requires precise knowledge of the links between the system and the signals flowing from the environment, as well as between the internal elements of the system. Needs arising from the control of the large complex installations always outweight the modern computing capabilities, and are becoming a cause for the progress of both the hardware as well as the algorithms and the numerical methods.

The question raised in the article refers to the situation when the considered system is described by differential-algebraic equations in a general way possible. This approach has a chance to wide and common use in industry. The presented method is part of a widely used approach, which reduces infinite dimensional task to the large-scale finite-dimensional problem.

The paper is constructed as follows. In the next section the backward differential formula (bdf) is presented as the tool for solving dae systems. New aspects of the inexact Newton method were presented in $3rd$ and $4th$ sections. The presented algorithms were tested on the kinetic batch reactor model. The results were presented in $5th$ section.

II. The backward Euler method

The codes for solving dae in the $fully-implicit$ form are based on a technique which was introduced by Gear [12]. The backward differential formula is the first general technique for the numerical solution of dae and have emerged as the most popular. The idea of this technique is that the derivative $\frac{dy(t)}{dt}$ can be approximated by a linear combination of the solution $y(t)$ at the current mesh point and at several previous mesh points ([14]).

Bdf was initially defined for the systems of differential equations coupled to algebraic equations. This method was soon extended to apply to any fully-implicit system of differential-algebraic equations

$$G\Big(\frac{dy(t)}{dt}, y(t), z(t), t\Big) = 0. \tag{1}$$

The simplest method for solving differential-algebraic systems is the first order bdf, or the backward Euler method, which consists of replacing the derivative in (1) by a backward difference

$$F\Big(\frac{y_n - y_{n-1}}{h}, y_n, z_n, t_n\Big) = 0. \tag{2}$$

where $h = t_n - t_{n-1}$.

The resulting system of nonlinear equations for y_n at each step is then usually solved by the Newton method [4]. In this way, the solution is advanced from time t_n to time t_{n+1}. It is assumed, that $y(t_0)$ is known. Assume too, that t (time) is the independent variable. In practical applications in chemical enginnering, as the independent variable is used usually the lenght of the reactor. If the time interval, in which the system has to be considered, is known, it can be scaled to the interval $[0, 1]$.

III. The inexact Newton method

The methodology presented in the previous paragraph leads to the following equation

$$F(x) = 0. \tag{3}$$

This equation is very general and is often found in scientific and engineering computing areas. We assume that the function F is considered, where $F : \mathcal{R}^n \to \mathcal{R}^n$ is a nonlinear mapping with the following properties

978-83-60810-53-8/$25.00

(1) There exists an $x^* \in \mathcal{R}$ with $F(x^*) = 0$.
(2) F is continously differentiable in a neighborhood of x^*.
(3) $F'(x^*)$ is nonsingular.

There are a lot of methods for solving this nonlinear equation. One of the most popular and important is the Newton method. The Newton's method is attractive because it converges rapidly (quadratically) from any sufficiently good initial point. Its computational cost can be expensive, particularly, when the size of the problem is very large, because in each iteration step the Newton equations

$$F(x_k) + F'(x_k)s_k = 0 \tag{4}$$

should be solved. Here x_k denotes the current iterate, and $F'(x_k)$ is the Jacobian matrix of $F(x)$ at point x_k. The solution s_k^N of the Newton equation is the Newton step. Once the Newton step is obtained, the next iterate is given by

$$x_{k+1} = x_k + s_k^N. \tag{5}$$

In 1982 Dembo, Eisenstat and Steihaug proposed the inexact Newton method, which is a generalization of the Newton method [8]. The inexact Newton method is any method which, given an initial guess x_0, generates a sequence x_k of approximations to x^* as in Algorithm 1.

ALGORITHM 1. The inexact Newton method
1. Given $x_0 \in \mathcal{R}^n$
2. For $k = 0, 1, 2, \cdots$ until x_k convergence
  2.1 Choose some $\eta_k \in [0, 1)$
  2.2 Inexactly solve the Newton equations and obtain a step s_k, such that $\|F(x_k) + F'(x_k)s_k\| \le \eta_k \|F(x_k)\|$. $(\star)$
  2.3 Let $x_{k+1} = x_k + s_k$.

In the Algorithm 1, η_k is the forcing term in the k-th iteration, s_k is the inexact Newton step and $(\star)$ is the inexact Newton condition.

In each iteration step of the inexact Neton method a real number $\eta_k \in [0, 1)$ should be chosen. Then the inexact Newton step s_k is obtained by solving the Newton equation approximately with an iteration solver for systems of nonlinear equation. Since $F(x_k) + F'(x_k)s_k$ is both residual of the Newton equations and the local linear model of $F(x)$ at x_k, the inexact Newton condition $(\star)$ reflects both the reduction in the norm of the local linear model and certain accuarcy in solving the Newton equations. Thus the role of forcing terms is control the degree of accuracy of solving the Newton equations. In particular, if $\eta_k = 0$ for all k, then the inexact Newton method is reduced into the Newton method.

The inexact Newton method, like the Newton method, is locally convergent.

Theorem 1 ([8]): Assume that $F : \mathcal{R}^n \to \mathcal{R}^n$ is continously differentiable, $x^* \in \mathcal{R}^n$ such that $F'(x^*)$ is nonsingular. Let $0 < \eta_{max} < \beta < 1$ be the given constants. If the forcing terms η_k in the inexact Newton method satisfy $\eta_k \le \eta_{max} < \beta < 1$ for all k, then there exists $\varepsilon > 0$, such that for any $x_0 \in N_\varepsilon(x^*) \equiv \{x : \|x - x^*\| < \varepsilon\}$, the sequence $\{x_k\}$ generated by the inexact Newton method converges to x^*, and

$$\|x_{k+1} - x^*\|_* \le \beta \|x_k - x^*\|_*, \tag{6}$$

where $\|y\|_* = \|F'(x^*)y\|$.

If the forcing terms $\{\eta_k\}$ in the inexact Newton method are uniformly strict less than 1, then by Theorem 1, the method is locally convergent. The following result states the convergence rate of the inexact Newton method.

Theorem 2 ([8]): Assume that $F : \mathcal{R}^n \to \mathcal{R}^n$ is continously differentiable, $x^* \in \mathcal{R}^n$ such that $F'(x^*)$ is nonsingular. If the sequence $\{x_k\}$ generated by the inexact Newton method converges to x^*, then

(1) x_k converges to x^* superlinearly when $\eta_k \to 0$;
(2) x_k converges to x^* quadratically if $\eta_k = \mathcal{O}(\|F(x_k)\|)$ and $F'(x)$ is Lipschitz continuous at x^*.

Theorem 2 indicates, that the convergence rate of the inexact Newton method is determined by the choice of the forcing terms.

IV. A CHOICE OF FORCING TERMS

In the literature, researchers proposed some strategies to determine a good sequence of forcing terms. Here, four representatives strategies were selected.

(1) The choice of Dembo and Steihaug [9]:

$$\eta_k = \min\left\{\frac{1}{k+2}, \|F(x_k)\|\right\}. \tag{7}$$

The two strategies given by Eisenstat and Walker are more popular [11]. Among this two strategies, choice (2a) reflects the agreement between $F(x)$ and its local linear model at the previous step. Choice (2b) reflects the reduction rate of $\|F(x)\|$ from x_{k-1} to x_k.

For computational purposes of preventing the forcing terms from becoming quickly too small, some safeguards were added. The following strategies were obtained.

(2a) Given $\eta_0 \in [0, 1)$, choose

$$\eta_k = \begin{cases} \xi_k, & \eta_{k-1}^{(1+\sqrt{5})/2} \le 0.1, \\ \max\{\xi_k, \eta_{k-1}^{(1+\sqrt{5})/2}\}, & \eta_{k-1}^{(1+\sqrt{5})/2} > 0.1, \end{cases} \tag{8}$$

where

$$\xi_k = \frac{\|F(x_k) - F(x_{k-1}) - F'(x_{k-1})s_{k-1}\|}{\|F(x_{k-1})\|}, \tag{9}$$

$k = 1, 2, \ldots,$ or

$$\xi_k = \frac{|\,\|F(x_k)\| - \|F(x_{k-1}) + F'(x_{k-1})s_{k-1}\|\,|}{\|F(x_{k-1})\|}, \tag{10}$$

$k = 1, 2, \ldots.$

(2b) Given $\gamma \in (0, 1]$, $\omega \in (1, 2]$, $\eta_0 \in [0, 1)$, choose

$$\eta_k = \begin{cases} \xi_k, & \gamma(\eta_{k-1})^\omega \le 0.1, \\ \max\{\xi_k, \gamma(\eta_{k-1})^\omega\}, & \gamma(\eta_{k-1})^\omega > 0.1, \end{cases} \tag{11}$$

where

$$\xi_k = \gamma\left(\frac{\|F(x_k)\|}{\|F(x_{k-1})\|}\right)^\omega, \tag{12}$$

$k = 1, 2, \ldots$.

(3) Choice of H.-B. An et al. [1]. Assume, that x_k is the current iterate and s_k is the step from x_k. The actual reduction $Ared_k(s_k)$ and predicted reduction $Pred_k(s_k)$ of $F(x)$ at x_k with step s_k are definned as follows

$$Ared_k(s_k) = \|F(x_k)\| - \|F(x_k + s_k)\|, \tag{13}$$

$$Pred_k(s_k) = \|F(x_k)\| - \|F(x_k) + F'(x_k)s_k\|. \tag{14}$$

Furthermore, let

$$r_k = \frac{Ared_k(s_k)}{Pred_k(s_k)}. \tag{15}$$

In this approach, r_k is used to adjust the forcing term η_k. Considering the value of r_k, one can distinguish four situation, which can have a place in computations.

(a) If $r_k \approx 1$, the the local linear model and nonlinear model will agree well on their scale and $\|F(x)\|$ usually will be reduced.

(b) If r_k nears 0, but $r_k > 0$, then the local linear model and nonlinear model disagree and $\|F(x)\|$ can be reduced very little.

(c) If $r_k < 0$, then the local linear model and nonlinear model disagree and $\|F(x)\|$ will be enlarged.

(d) If $r_k \gg 1$, then the local linear model and nonlinear model also disagree, but $\|F(x)\|$ will be reduced greatly.

The acceptable situations are, when $r_k \approx 1$ or $r_k \gg 1$, because in this cases the local linear model and nonlinear model agree well or at least leads to a great reduction point.

According to the property of r_k, one can choose forcing terms as follows.

$$\eta_k = \begin{cases} 1 - 2p_1, & r_{k-1} < p_1, \\ \eta_{k-1}, & p_1 \le r_{k-1} < p_2, \\ 0.8\eta_{k-1}, & p_2 \le r_{k-1} < p_3, \\ 0.5\eta_{k-1} & r_{k-1} \ge p_3, \end{cases} \tag{16}$$

where $0 < p_1 < p_2 < p_3 < 1$ are prescribed at first and $p_1 \in (0, \frac{1}{2})$. Assume, that η_0 is given.

The choice of forcing terms proposed in [1] is to determine η_k by the magnitude of r_{k-1}.

It is worth to note, that the current forcing term η_k is determined by the previous value r_{k-1} and η_k determines the value r_k through solving the Newton equations approximately.

V. Numerical results

As an example the kinetic batch reactor was choosed. This example is known from the literature [2], [6], [7] The concentrations are modeled by the system of differential and algebraic equations. The desired product AB is formed in the reaction

$$HA + 2BM \rightarrow AM + MBMH. \tag{17}$$

For the formulation given here the differential and algebraic variables are denoted by y_j and z_j respectively (Table 1).

The kinetic model is stated in terms of six differential mass balance equations

$$\dot{y}_1 = -k_2 y_2(t) z_8(t), \tag{18}$$

TABLE I
Batch reactor dynamic variables.

y_1	Differential State	$[HA] + [A^-]$
y_2	Differential State	$[BM]$
y_3	Differential State	$[HABM] + [ABM^-]$
y_4	Differential State	$[AB]$
y_5	Differential State	$[MBMH] + [MBM^-]$
y_6	Differential State	$[M^-]$
z_7	Algebraic State	$[H^+]$
z_8	Algebraic State	$[A^-]$
z_9	Algebraic State	$[ABM^-]$
z_{10}	Algebraic State	$[MBM^-]$

$$\dot{y}_2 = -k_1 y_2(t) y_6(t) + k_{-1} z_{10}(t) - k_2 y_2(t) z_8(t), \tag{19}$$

$$\dot{y}_3 = k_2 y_2(t) z_8(t) + k_3 y_4(t) y_6(t) - k_{-3} z_9(t), \tag{20}$$

$$\dot{y}_4 = -k_3 y_4(t) y_6(t) + k_{-3} z_9(t), \tag{21}$$

$$\dot{y}_5 = k_1 y_2(t) y_6(t) + k_{-1} z_{10}(t), \tag{22}$$

$$\dot{y}_6 = -k_1 y_2(t) y_6(t) - k_3 y_4(t) y_6(t) + k_{-1} z_{10}(t) + k_{-3} z_9(t), \tag{23}$$

an electroneutrality condition

$$z_7(t) = -0.0131 + y_6(t) + z_8(t) + z_9(t) + z_{10}(t) \tag{24}$$

and three equlibirum conditions

$$z_8(t) = \frac{K_2 y_1(t)}{K_2 + z_7(t)}, \tag{25}$$

$$z_9(t) = \frac{K_3 y_3(t)}{K_3 + z_7(t)}, \tag{26}$$

$$z_{10}(t) = \frac{K_1 y_5(t)}{K_1 + z_7(t)}. \tag{27}$$

with initial conditions $y_1(0) = 1.5776$, $y_2(0) = 8.32$, $y_j(0) = 0.0, j = 3, 4, 5$, $y_6(0) = 0.0131$, $z_7(0) = 0.5(-K_2 + \sqrt{K_2^2 + 4K_2 y_1(0)})$, $z_8(0) = z_7(0)$, $z_j(0) = 0.0, j = 9, 10$.

The following values of rate and equlibirum constants were used $k_1 = 21.893(hr^{-1} \cdot Kg \cdot gmole^{-1}), k_{-1} = 2.14E09(hr^{-1}), k_2 = 32.318(hr^{-1} \cdot Kg \cdot gmole^{-1}), k_3 = 21.893(hr^{-1} \cdot Kg \cdot gmole^{-1}), k_{-3} = 1.07E09(hr^{-1}), K_1 = 7.65E-18(gmole \cdot Kg^{-1}), K_2 = 4.03E-11(gmole \cdot Kg^{-1}), K_3 = 5.32E-18(gmole \cdot Kg^{-1})$.

The equations were considered in the time domain $t \in [0, 2.5]$. Then the equations were discretized into equidistant points with distnace 0.025. It resulted in 600 differential and 400 algebriac state variables. Then, 1000 equality constraints from the backward Euler method were imposed. The Jacobian matrix was obtained analitically and stored as the 1000×1000 sparse matrix.

This large-scale system of the linear equations was solved using GMRES algorithm [15]. The inexact Newton backtracking method [10] was used with four presented approaches for adjusting the forcing terms.

The results in Table 2 indicate, that the considered problem is difficult to solve. Iterations quickly converge to the locally optimal solution. The parameter r_k gives an answer, what

TABLE II
RESULTS FOR CHOICE OF H.-B. AN ET $al.$ [1].

iter	η_3	$\|F_3(x_k)\|$	r_k
1	0.4375	5.8635e3	1.4670
2	0.4297	4.0018e3	1.7745
3	0.4287	2.7828e3	2.1213
4	0.4286	2.0803e3	3.4620
5	0.4286	1.4856e3	2.2609
6	0.4286	1.2529e3	1.5051
7	0.5000	1.1420e3	5.3385e-4
8	0.5000	1.1420e3	NaN
9	0.5000	1.1420e3	NaN
10	0.5000	1.1420e3	NaN

TABLE III
RESULTS FOR OTHER SEQUENCES OF FORCING TERMS.

iter	η_1	$\|F_1(x_k)\|$	η_{2a}	η_{2b}	$\|F_{2a,2b}(x_k)\|$
1	0.3333	5.8635e3	0.5000	0.5000	5.8635e3
2	0.2500	4.0018e3	0.8057	0.4677	4.0018e3
3	0.2000	2.7828e3	0.9226	0.4655	2.7828e3
4	0.1667	2.0803e3	0.9689	0.4654	2.0803e3
5	0.1429	1.4856e3	0.9874	0.4654	1.4856e3
6	0.1250	1.2529e3	0.9949	0.4654	1.2529e3
7	0.1111	1.2143e3	0.9979	0.4773	1.1420e3
8	0.1000	1.1986e3	1.0000	0.5000	1.1420e3
9	0.0909	1.1972e3	1.0000	0.5000	1.1420e3
10	0.0833	1.1966e3	1.0000	0.5000	1.1420e3

is the relation between linear model and the whole system. The linear model agrees with the nonlinear model only at the beginning of the solution process. It is worth to note, that only the approach presented in [1] idicates, that after 7 iterations some difficulties can occur.

The simulations were executed with the parameters: $\gamma = 0.5$, $\omega = 1.5$ for proposition $2b$ and $p_1 = 0.25$, $p_2 = 0.6$, $p_3 = 0.8$ for the choice proposed in [1].

There are results for forcing terms adjusted in other manners in Table 3. The forcing terms adjusted as presented in [9] were decreased monotonically, but there is no information about agreement between $F(x)$ and its local linear model.

The forcing terms, adjusted as presented in [11], did not decrease monotonically to 0. Its main drawback is, that either the agreement between $F(x)$ and its local linear model at the previous step or the reduction rate of $\|F(x)\|$ are reflected in adapatation of forcing terms.

The simulations were executed in Matlab environment using Wroclaw Centre for Networking and Supercomputing.

As one can see, the results presented in the Table 2 and 3 are not the optimal solutions. If the initial guess for the inexact Newton method is close enough to the desired solution, then the convergence is very fast provided that the forcing terms are sufficiently small. But a good initial guess is generally very diffcult to obtain, especially for nonlinear equations that have unbalanced nonlinearities. Then the step length is often determined by the components with the strongest nonlinearities [5]. The nonlinearities are "unbalanded" when the step length is determined by a subset of the overall degrees of freedom.

VI. CONCLUSION

In the paper the new apsects of the inexact Newton method for solving differential-algebraic equations were presented, then the dae systems in the fully implicit form were considered. The methods for the choice of forcing terms for the inexact Newton method were presented and tested on the difficult and highly nonlinear kinetic batch reactor.

The authors would like to indicate, that the choice of forcing terms, which reflects both the agreement between $F(x)$ and its local linear model and the reduction rate of $\|F(x)\|$ are especially usefull for solving the large scale differential-algebraic equations. As the next step, the new preconditioned Jacobian-free optimization algorithm, which could solve the large-scale optimization tasks, will be studied and adjusted for new challenges in solving the optimal control problems [13].

REFERENCES

[1] H.-B. An, Z.-Y. Mo, X.-P. Liu, "A choice of forcing terms in inexact Newton method", *Journal of Computational and Applied Mathematics*, vol. 200, 2007, pp. 47-60.

[2] J.T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming. Second edition*, SIAM, Philadelphia 2010.

[3] L.T. Biegler, *Nonlinear Programming. Concepts, Algorithms, and Applications to Chemical Processes*, SIAM, Philadelphia 2010.

[4] K.E. Brenan, S.L. Campbell, L.R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM, Philadelphia, 1996.

[5] X.-C. Cai, D.E. Keyes, "Nonlinearly Preconditioned Inexact Newton Algorithms", *SIAM Journal on Scientific Computing*, vol. 24, 2002, pp. 183-200.

[6] M. Caracotsis, W. E. Stewart, "Sensitivity analysis of Initial Value Problems with mixed ODEs and algebraic equations", *Computers and Chemical Engineering*, vol. 9, 1985, pp. 359-365.

[7] P. Drąg, K. Styczeń, "A Two-Step Approach for Optimal Control of Kinetic Batch Reactor with electroneutrality condition", *Przegląd Elektrotechniczny (Electrical Review)*, vol. 6, 2012, pp. 176-180.

[8] R.S. Dembo, S.C. Eisenstat, T. Steihaug, "Inexact Newton Methods", *SIAM Journal on Numerical Analysis*, vol. 19, 1982, pp. 400-408.

[9] R.S. Dembo, T. Steihaug, "Truncated-Newton algorithm for large-scale unconstrained optimization", *Mathematical Programming*, vol. 26, 1983, pp. 190-212.

[10] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, *SIAM Journal on Optimization*, vol. 4, 1994, pp. 393-422.

[11] S.C. Eisenstat, H.F. Walker, "Choosing the forcing terms in an inexact Newton method", *SIAM Journal on Scientific Computing*, vol. 17, 1996, pp. 16-32.

[12] C.W. Gear, "The simultaneous numerical solution of differential-algebraic equations", *IEEE Transactions on Circuit Theory*, vol. 18, 1971, pp. 89-95.

[13] D.A. Knoll, D.E. Keyes, "Jacobian-free Newton-Krylov methods: a survey of approaches and applications", *Journal of Computational Physics*, vol. 193, 2004, pp. 357-397.

[14] L. Petzold, "Differential/Algebraic Equations are not ODEs", *SIAM Journal on Scientific Computing*, vol. 3, 1982, pp. 367-384.

[15] Y. Saad, M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems", *SIAM J. Sci. Stat. Comput.*, vol. 7, 1986, pp. 856-869.

[16] V.S. Vassiliadis, R.W.H. Sargent, C.C. Pantelides, "Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems without Path Constraints", *Ind. Eng. Chem. Res.*, vol. 33, 1994, pp. 2111-2122.

[17] V.S. Vassiliadis, R.W.H. Sargent, C.C. Pantelides, "Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems with Path Constraints", *Ind. Eng. Chem. Res.*, vol. 33, 1994, pp. 2123-2133.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 643–646

On some quality criteria of bipolar linguistic summaries

Mateusz Dziedzic
Department of Automatic Control
and Information Technology
Cracow University of Technology
ul. Warszawska 24, 31–155 Kraków, Poland
also
PhD Studies, Systems Research Institute
Polish Academy of Sciences
Email: Mateusz.Dziedzic@ibspan.waw.pl

Janusz Kacprzyk, IEEE Fellow
and Sławomir Zadrożny
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01–447 Warszawa, Poland
Email: {Slawomir.Zadrozny,Janusz.Kacprzyk}@ibspan.waw.pl

***Abstract*—The quality measures for bipolar linguistic summaries of data, as proposed in our previous work [1], are further developed. The summaries introduced in [2] are assumed to be an extension of the "classical" linguistic summarization (cf. [3], [4]), a human-consistent data mining technique revealing complex patterns present in data. This extension consists in using the "and possibly" to build a summary and introducing the notion of context to determine the validity of the summary. We present a more detailed description of summaries quality measures/criteria and reports results of more extensive computational experiments.**

I. Introduction

THE AIM of data mining is to discover patterns in data in a form interesting and clear to the end user. A promising way to achieve this is to use (quasi) natural language. This has been a motivation for the *linguistic data summaries* introduced by Yager [3] and further developed by him [5] and other contributors, notably Kacprzyk and Zadrożny [6], [7].

Recently, an important role of bipolarity of user preferences, in particular in fuzzy linguistic querying [8], is noticed. Its essence is in considering both positive and negative evaluations of objects in question which are not necessarily complements of each other. An important and most interesting line of research focuses on the treatment of negative evaluations as obligatory while the positive evaluations as somehow secondary. This results in the introduction and study of the "and possibly" logical connective [9]. Moreover, the concept of bipolar queries involving such a connective has been proposed [10] to better model user preferences as exemplified by the query "Find a house, cheap *and possibly* located close to a station".

In our previous papers [1], [2] we began to study if relation between fuzzy linguistic queries and linguistic data summaries may be adopted for bipolar queries. The results were positive and led us to the concept of bipolar linguistic summaries of data. In this paper we focus on two quality criteria of such new type of linguistic summaries, introduced in [1] and referring to the notion of the context of a summary.

The structure of the paper is as follows. In Section II we briefly remind the basics of the fuzzy linguistic queries and "classical" linguistic summaries, and introduce the notation to be used in the rest of the paper. In Section III we discuss the concepts of bipolar queries and bipolar linguistic summaries. Section IV reports on the computational experiments focused on comparing different summary contexts and discusses the results obtained.

II. Fuzzy linguistic queries and linguistic data summaries

A. Fuzzy linguistic queries

In classical query languages, such as SQL, preferences of users must be expressed precisely. However, due to the fact that their original form is a natural language expression, they are very often imprecise. For example, one may be concerned primarily with the cost while looking for an apartment to rent and express his or her preference as:

Find *cheap* apartments for rent in Kraków. (1)

In an approach, referred here to as fuzzy linguistic queries, such imprecise terms (e.g. *cheap*) are represented by fuzzy sets defined in the domains of respective attributes.

Usually, a dictionary of linguistic terms is assumed as a part of an implementation which contains predefined linguistic terms and corresponding fuzzy sets as well as terms defined by the users. Linguistic terms collected in a dictionary are a starting point to derive meaningful *linguistic summaries* of a database.

B. Linguistic summaries of data

As linguistic summaries we understand a (quasi) natural language sentences that grasp some characteristic features of data collected in a database. We use Zadeh's calculus of linguistically quantified propositions as the underlying formalism. The statement representing a linguistic summary points out some properties shared by a number of data items and the proportion of these data items is expressed using a *linguistic quantifier*. Yager [3], [5] first proposed the use of linguistically quantified propositions to summarize data in a user consistent

978-83-60810-53-8/$25.00

way. That idea has been further developed, cf., e.g., Kacprzyk and Yager [11], and Kacprzyk, Yager and Zadrożny [4], [6].

Assuming $R = \{t_1, \ldots, t_n\}$ is a set of tuples (a relation) in a database, representing, e.g., a set of employees; $A = \{A_1, \ldots, A_m\}$ is a set of attributes defining schema of the relation R, e.g., salary, age, education_level, etc. in a database of employees ($A_j(t_i)$ denotes a value of attribute A_j for a tuple t_i), the linguistic summary of a set R is a linguistically quantified proposition which is an instantiation of one of the following abstract *protoforms* [12] of type I and type II, respectively:

$$Q_{t \in R}\ S(t) \tag{2}$$

$$Q_{t \in R}\ (U(t), S(t)) \tag{3}$$

then a linguistic summary is composed of the following elements: a *summarizer* S which is a fuzzy predicate representing, e.g., an expression "an employee is well-educated", formed using attributes of the set A; a *qualifier* U (optional) which is another fuzzy predicate representing, e.g., a set of "young employees"; a *linguistic quantifier* Q, e.g., "most" expressing the proportion of tuples satisfying the summarizer (optionally, among those satisfying a qualifier); *truth (validity)* T of the summary, i.e. a number from $[0, 1]$ expressing the truth of a respective linguistically quantified proposition.

In Yager's original approach [3] the linguistic quantifiers are represented using Zadeh's definition [13]. A *proportional*, *non-decreasing* linguistic quantifier Q is represented by a fuzzy set in $[0, 1]$ and $\mu_Q(x)$ states the degree to which the proportion of $100{\times}x$ % of elements of the universe match the proportion expressed by the quantifier Q. Thus, the truth degree of the linguistic summaries of type I (here we use only type I summaries, thus type II is omitted) is:

$$T(Q_{t \in R} S(t)) = Z_Q(S) = \mu_Q[\frac{1}{n} \sum_{i=1}^{n} \mu_S(t_i)] \tag{4}$$

III. Bipolar queries and bipolar linguistic summaries of data

A. Bipolar queries

In classical approaches to preferences modelling, notably in database querying, it is usually assumed that an alternative (tuple) is either accepted or rejected. However, the results of many studies, cf. [10], seem to suggest that the decision maker often comes up with somehow independent evaluations of positive and negative features of alternatives in question. This leads to a general concept of *bipolar query* against database, evaluation of which results in two degrees corresponding to the satisfaction of the positive and negative condition.

Most of the research on bipolar queries are focused on a special case where the positive and negative conditions are interpreted in an asymmetric way [10]. Namely, the latter is treated as a *constraint*, denoted C, which has to be satisfied, while the former plays the role of a mere *preference*, denoted P.

We follow the approach of Lacroix and Lavency [14], Yager [15], [16] and Bordogna and Pasi [9], adapted for database querying by Zadrożny and Kacprzyk [17], which combine both conditions using the "and possibly" operator which aggregates their satisfaction degrees depending on the possibility of a simultaneous matching of both conditions.

Thus, the bipolar query's condition may be formally written as:

$$C \text{ and possibly } P, \tag{5}$$

and may be illustrated with query: Find employees that are *young* and possibly earn a *high* salary. Such a bipolar query would be denoted (C, P) and interpreted as follows. If there is a tuple which satisfies both conditions, then and only then it is actually *possible* to satisfy both of them and each tuple of data has to do so, and, on the other hand, if there is no such a tuple, then condition P can be ignored. The matching degree of the (C, P) query against a tuple t may be formalized as [14]:

$$T\left(C(t)\ and\ possibly\ P(t)\right) = C(t) \wedge \left(\exists s\left(C(s) \wedge P(s)\right) \Rightarrow P(t)\right) \tag{6}$$

B. Bipolar linguistic summaries

The main idea behind the bipolar linguistic summaries is to relate the "and possibly" to *a part* of the database instead of the whole database. Let us consider the following example:

> *Most* employees have a *short* seniority *and, if possible with respect to similarly educated colleagues*, earn a *high* salary.

An employee matches such a summary if:

1) he or she has a short seniority (to a high degree) and earns a high salary (to a high degree), or
2) he or she has a short seniority (to a high degree) and there is no other *similarly educated* employee who earns a high salary.

A characteristic feature of such a summary is the use of a summarizer employing an extended version of the "and possibly" operator, which we will refer to as the "contextual and possibly" operator. This operator may be expressed as:

$$C \text{ and possibly } P \text{ with respect to } W. \tag{7}$$

For the purposes of bipolar queries (and, thus, bipolar linguistic summaries) the predicates C and P should be interpreted as the required and desired conditions, respectively, while the predicate W denotes the *context* in which the possibility of satisfying both C and P will be assessed, separately for each tuple. Then, the formula (7) is interpreted as:

$$T\left(C(t)\ and\ possibly\ P(t)\ with\ respect\ to\ W\right) = C(t) \wedge \left(\exists s\left(W(t,s) \wedge C(s) \wedge P(s)\right) \Rightarrow P(t)\right) \tag{8}$$

Our preliminary computational experiments show that usage of the standard De Morgan triples[1], both with the S- and

[1]As standard De Morgan triples we understand $(\wedge_{\min}, \vee_{\max}, \neg)$, $(\wedge_{\Pi}, \vee_{\Pi}, \neg)$ and $(\wedge_{\mathrm{L}}, \vee_{\mathrm{L}}, \neg)$ with t- and s-norms: Minimum $(\min(x, y))$ and Maximum $(\max(x, y))$; Product $(x \cdot y)$ and Probabilistic sum $(x + y - x \cdot y)$; and Łukasiewicz's $(\max(0, x + y - 1))$ and $(\min(1, x + y))$, respectively.

R-implication, in (8) may lead to somehow counter-intuitive results in terms of bipolar queries evaluation.

Thus we use the MinMax triple and Goguen $R-$implication which turns (8) into:

$$T\left(C(t)\ and\ possibly\ P(t)\ with\ respect\ to\ W\right) = \begin{cases} \min(C(t),1) & \text{for } \exists WCP(t)=0 \\ \min\left(C(t), \min\left(1, \frac{P(t)}{\exists WCP(t)}\right)\right) & \text{otherwise} \end{cases}, \tag{9}$$

where $\exists WCP(t)$ denotes $\max_{s\in R}\min(W(t,s),C(s),P(s))$.

C. Summary context quality criteria

In [1] we stated that the quality of the summary context $W(t,s)$ itself and the whole implication premise in (8) have to be considered when measuring the quality of the bipolar linguistic summaries.

Namely, if P and/or W are such that the premise of the implication in (8) is true to a very low or a very high degree *for most* of t's, then the summarizer (7) does not make much sense even if the truth value of a summary is high. This is due to the behaviour of the bipolar query "C and possibly P" which turns into "C" and "C and P", respectively, when the truth degree of $\exists_{s\in R}C(s)\wedge P(s)$ is close to 0 and close to 1.

The introduction of the context W partially alleviates this problem but W has to be chosen carefully. If for most t's there does not exist s such that $W(t,s)$, then the premise of the implication is most often false and the summary is true to a high degree for any P. We propose a solution to this problem in a form of quality measures expressed using the following linguistically quantified propositions:

$$Q_{t\in R}\exists_{s\in R\setminus\{t\}}W(t,s) \tag{10}$$

$$Q_{t\in R}\exists_{s\in R\setminus\{t\}}C(s)\wedge P(s)\wedge W(t,s). \tag{11}$$

Namely, if the truth of (10) for a summary is too small (lower than some threshold value), then such a summary should be discarded. Also, if the truth of (11) is too high (too close to 1; larger than the second threshold value) or too small (too close to 0; lower than the third threshold value), then the summary also shouldn't be taken into account. Obviously, if the first threshold is violated, then also the third one is. On the other hand, even if the first threshold is satisfied, the summary may still fail to satisfy thresholds two or three and should be discarded.

Tuple t is excluded from the range of the existential quantifiers in (10)–(11) as if the only tuple related via W with t is only t itself, then, naturally, the resulting summary is of no interest.

IV. Computational experiments and discussion

Data on the rates of return (RORs) of selected investment funds[2] (IFs) (Tab. I), are used to present examples of bipolar linguistic summaries and their semantics in scope of summary context quality.

[2]URL: http://www.analizy.pl/fundusze/ as of May 24, 2013.

Table I
Selected investment funds (IF)

No.	IF rating[a]	1-month ROR[b]	12-month ROR
1	5	-3.7	6.9
2	3	-0.8	20.8
3	2	-3.5	2.2
4	5	-3.5	2.5
5	5	-2.3	9.7
6	4	-2.2	9.5

[a] Rating from http://www.analizy.pl/fundusze/, as of May 24, 2013.
[b] Rate Of Return.

Table II
Investment funds (IF) - local neighbourhoods coefficients of tuples

	$W_1(t,s)$						$W_2(t,s)$						$W_3(t,s)$					
No.	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
1	1.0			1.0	1.0		1.0			1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0
2		1.0						1.0	1.0			1.0	1.0	1.0	1.0	1.0	1.0	1.0
3			1.0					1.0	1.0					1.0	1.0			1.0
4	1.0			1.0	1.0		1.0			1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0
5	1.0			1.0	1.0		1.0			1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0
6						1.0	1.0	1.0		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Empty cell indicates no *similarity* ($W(t,s)=0.0$).

Fuzzy predicates are represented by trapezoidal membership functions and instantiated as (see Fig. 1):

- C: has "high"/"average"/"low" 12-month ROR,
- P: has "high"/"average"/"low" 1-month ROR,
- $W_{1/2/3}$: of "the same"/"very similar"/"quite similar" Rating, the former true iff IF rating(t) = IF rating(s) and the latter two defined over |IF rating(t) − IF rating(s)|.

In Tab. III we present summaries with truth value (evaluated using Zadeh's approach) $T>0$ obtained for data in Tab. I and compare the results in scope of proposed quality criteria (10) and (11).

In order to focus the interpretation on summaries contexts we consider only one linguistic quantifier Q with the membership function indicating the proportion of tuples satisfying the summarizer below 30% and above 80% as, respectively, totally

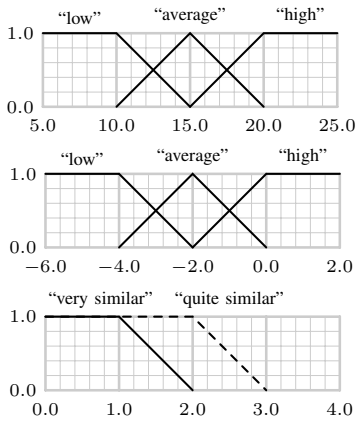


Figure 1. Membership functions of 12-month ROR (condition C – upper plot), 1-month ROR (condition P – center plot) and "similar" Rating (condition $W_{\text{Rating II/III}}$ – lower plot) predicates.

Table III
OBTAINED BIPOLAR LINGUISTIC SUMMARIES (USING ZADEH'S (Z_Q) APPROACH)

No.	Linguistic summary (Q, C, P)	W_1 [a]	W_2 [a]	W_3 [a]
1	*"Most" of IFs have "low" 12-month and possibly "high" 1-month ROR with respect to...*	*1.00; 0.50; 0.00*	*1.00; 1.00; 0.00*	*1.00; 1.00; 0.00*
2	*"Most" of IFs have "low" 12-month and possibly "average" 1-month ROR with respect to...*	0.55; 0.50; 0.23	0.52; 1.00; 0.89	0.30; 1.00; 0.74
3	*"Most" of IFs have "low" 12-month and possibly "low" 1-month ROR with respect to...*	0.75; 0.50; 0.41	0.46; 1.00; 0.71	0.46; 1.00; 0.68

[a] Truth degrees of the linguistic summary (Z_Q) and values of quality criteria (10) and (11) computed for the unitary quantifier Q for corresponding W predicates.

incompatible and compatible with the meaning of "most", while all intermediate proportions are treated as compatible to a degree in [0,1].

We focused here on showing the benefits of using contextual and possibly operator in the scope of linguistic data summarization, presenting both a theoretical and semantic justification of this concept and intuitively appealing examples.

Nine linguistic summaries (based on three different triples Q, C, P) reported in Tab. III clearly argue in favour of introduced additional quality criteria (measures) (10)–(11).

First, criterion (10) values 0.5 and 1.0 indicates that all selected contexts are meaningful (see Tab. II).

On the other hand, summaries with highest truth values (all three variants of No. 1 summary) clearly should be discarded — there are no IFs with "high" one-month ROR, which, as we stated at the beginning of section III-C, turns (7) in those summaries into a simple summarizer C (i.e. whole summary into *"Most" of IFs have "low" 12-month ROR.*).

Last three columns of Tab. III confirm that the use of (10) and (11) helps to distinguish interesting summaries (No. 2 and 3 with different W instantiations) from among all with high truth values (rejected summaries are italicized). Additional studies are needed in order to clearly determine the best summaries, yet already the results are promising.

V. CONCLUDING REMARKS

Preliminary computational results of the extension of linguistic data summaries, i.e. bipolar linguistic summaries proposed in [2], demonstrated the need for new quality criteria to determine usefulness of the summary. In [1] we introduced two of them, which have been studied deeper here. The results presented (Tab. III) show that proposed criteria fulfill their role and help select bipolar linguistic summaries valuable and interesting for an end user. Future works in this subject will mainly cover combining introduced criteria with other known quality measures, in order to determine a single value of quality of linguistic summary on one hand, and for evaluating and selecting linguistic summaries by means of heuristic methods, on the other hand.

ACKNOWLEDGMENT

Mateusz Dziedzic contribution is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme (2007-2013) and European Regional Development Fund.

REFERENCES

[1] M. Dziedzic, S. Zadrożny, and J. Kacprzyk, "Bipolar linguistic summaries: a novel fuzzy querying driven approach." in *2013 IFSA-NAFIPS Joint Congress*. Edmonton (Canada): IEEE, 2013, pp. 1279–1284.

[2] ——, "Towards bipolar linguistic summaries: a novel fuzzy bipolar querying based approach." in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*. Brisbane (Australia): IEEE, 2012, pp. 1–8.

[3] R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, pp. 69–86, 1982.

[4] J. Kacprzyk, R. R. Yager, and S. Zadrożny, "A fuzzy logic based approach to linguistic summaries of databases," *International Journal of Applied Mathematics and Computer Science*, no. 10, pp. 813–834, 2000.

[5] R. Yager, "On linguistic summaries of data," in *Knowledge Discovery in Databases*, Frawley W. and Piatetsky-Shapiro G., Eds. AAAI/MIT Press, 1991, pp. 347–363.

[6] J. Kacprzyk and S. Zadrożny, "On a fuzzy querying and data mining interface," *Kybernetika*, no. 36, pp. 657–670, 2000.

[7] ——, "Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools," *Inf. Sci.*, vol. 173, no. 4, pp. 281–304, 2005.

[8] S. Zadrożny and J. Kacprzyk, "Bipolar queries: An approach and its various interpretations," in *IFSA/EUSFLAT'09 Conf.*, Lisbon (Portugal), 2009, pp. 1288–1293.

[9] G. Bordogna and G. Pasi, "Linguistic aggregation operators of selection criteria in fuzzy information retrieval," *International Journal of Intelligent Systems*, vol. 10, no. 2, pp. 233–248, 1995.

[10] D. Dubois and H. Prade, "Bipolarity in flexible querying," in *FQAS 2002*, ser. LNAI, T. Andreasen, A. Motro, H. Christiansen, and H. L. Larsen, Eds. Berlin, Heidelberg: Springer-Verlag, 2002, vol. 2522, pp. 174–182.

[11] J. Kacprzyk and R. R. Yager, "Linguistic summaries of data using fuzzy logic," *International Journal of General Systems*, no. 30, pp. 33–154, 2001.

[12] L. Zadeh, "From search engines to question answering systems – the problems of world knowledge relevance deduction and precisiation." in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed. Elsevier, 2006, pp. 163–210.

[13] ——, "A computational approach to fuzzy quantifiers in natural languages," *Computers and Mathematics with Applications*, vol. 9, pp. 149–184, 1983.

[14] M. Lacroix and P. Lavency, "Preferences: Putting more knowledge into queries." in *Proceedings of the 13 International Conference on Very Large Databases*, Brighton (UK), 1987, pp. 217–225.

[15] R. Yager, "Higher structures in multi-criteria decision making," *International Journal of Man-Machine Studies*, vol. 36, pp. 553–570, 1992.

[16] ——, "Fuzzy logic in the formulation of decision functions from linguistic specifications," *Kybernetes*, vol. 25, no. 4, pp. 119–130, 1996.

[17] S. Zadrożny and J. Kacprzyk, "Bipolar queries: An aggregation operator focused perspective," *Fuzzy Sets and Systems*, vol. 196, pp. 69–81, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 647–650

A computational support for the group consensus reaching process in the fuzzy environment

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: kacprzyk@ibspan.waw.pl

Dominika Gołuńska
PhD Studies, Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Department of Automatic Control and Information Technology
Cracow University of Technology
ul. Warszawska 24,
31-155 Cracow, Poland
Email: dominika.golunska@pk.edu.pl

Andrzej Gorgoń
B.Sc. Studies, Department of Automatic Control and Information Technology
Cracow University of Technology
ul. Warszawska 24,
31-155 Cracow, Poland
Email: andrzej.gorgon.ag@gmail.com

***Abstract*—In this paper we present an intelligent consensus reaching support system within the group of individuals under fuzzy preferences and fuzzy majority. Our solution is based on the idea of soft degree of consensus proposed by Fedrizzi, Kacprzyk, Nurmi and Zadrożny, which is meant as the statement: "most of the individuals agree with the most of the options". Our new comprehensive model provides an effective support for the discussion guidance in the form of quantitative indices, i.e. sensitivity of individuals, option consensus degree and the cost of preference's changes. This additional measures support and simplify consensus reaching process and improve the degree of total agreement among decision makers.**

I. Introduction

CURRENTLY, any activity that a human being does involves solving problems, making choices, thus in general, involves some decisions. The essence of decision making is unified and short: there are some options to choose between and only one has to be chosen [2].

We accept the statement that the goal-directed decisions are difficult to make alone. Thus, we assume a session with a group of individuals and make the *group decision making process* the groundwork of our further consideration [3]. What matters here is respecting the preferences of all decision makers and arriving at a joint solution meant as an agreement of individuals as to the final decision. This interactive and iterative process is meant in the literature as a *consensus reaching process* and it requires: time, active participating of all individuals, creative thinking and being open-minded, active listening, etc [1]. The model of consensus reaching process is manageable only if individuals are able to negotiate and change their preferences.

Consensus reaching support system is commonly known as an intelligent, computer-based system that helps a team of decision makers solve problems and make choices [10]. The main role of this computer-based system plays *moderator*. His most important task is to support the discussion, i.e. he stimulates the exchange of knowledge, encourages appropriate individuals to change their opinions, focus the discussion on the relevant issues, etc. This is repeated until the group gets close to acceptable consensus or until some time limit is reached [6].

All of these features of consensus reaching process developed a need for a modern computer-based support with sophisticated tools which simplify this dynamic process and allow to achieve consensus in a more efficient way. There are many different methods that facilitate multi-stage consensus reaching process, but in this paper we show the implementation of the one of computer-based support systems. We consider either the improvement of consensus achieved or the cost of entire decision making process meant as a cost of total changes of individuals preferences.

II. Framework of the Consensus Reaching Model

A. Fuzzy Preference Relations

The core of our system is a human consistent representation of preferences. Preference relation is a very useful tool that gives relevant information about the comparison of options in decision making process [11].

Formally, there is a finite set of $n \geq 2$ options, $S = \{s_1, s_2, ..., s_n\}$, and a finite set of $m \geq 2$ individuals, $E = \{e_1, e_2, ..., e_m\}$. Each individual $k \in E$ presents his opinion as to the particular pairs of options in S. These testimonies are assumed to be individual *fuzzy preference relation* R_k defined over the set of options S (i.e. in $S \times S$) [3].

This work was partially supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

An individual fuzzy preference relation of expert k, R_k, is given by its membership function $\mu_{R_k}: S \times S \to [0,1]$. Namely, $\mu_{R_k}(s_i, s_j) > 0{,}5$ denotes that the alternative s_i is preferred to the alternative s_j, $\mu_{R_k}(s_i, s_j) < 0{,}5$ indicates that the option s_j is preferred to the option s_i and $\mu_{R_k}(s_i, s_j) = 0{,}5$ denotes that there is no difference between two considered options s_i and s_j [11].

We assume cardinality of S to be small enough to allow us to represent individual fuzzy preference relation R_k by a $n \times n$ matrix $R_k = [r_{ij}^k]$, such that $r_{ij}^k = \mu_{R_k}(s_i, s_j)$, *i,j=1,...,*n; *k=1,...,m*. R_k is also assumed to be reciprocal, i.e. $r_{ij}^k + r_{ji}^k = 1$, moreover, $r_{ii}^k = 0$, for all i, j, k [5].

B. Fuzzy Majority and Fuzzy Linguistic Quantifiers

An important part of our consensus reaching model is a fuzzy majority in the sense of fuzzy linguistic quantifiers, i.e. most, almost all etc. It is represented by the fuzzy logic-based calculus of linguistically quantified statements due to Zadeh [12].

A linguistically quantified statement is understood as "most individuals are satisfied" which can be written as

$$Qy\text{'s are } F \tag{1}$$

where Q is a linguistic quantifier (e.g., most), $Y = \{y\}$ is a set of objects (e.g., individuals) and F is a property (e.g., satisfied).

Our task is to find the degree of truth value of this linguistically quantified statement (1). First, a fuzzy linguistic quantifier is equated with a fuzzy set in [0,1]. For instance, "most" may be given as

$$\mu_{\text{"most"}}(x) = \begin{cases} 1 & \text{for} \quad x > 0.8 \\ 2x - 0.6 & \text{for } 0.3 \le x \le 0.8 \\ 0 & \text{for} \quad x < 0.3. \end{cases} \tag{2}$$

Property F is defined as a fuzzy set in the set of objects Y, and if $Y = \{y_1, \ldots, y_p\}$, then we suppose that truth value $(y_i \text{ is } F) = \mu_F(y_i), i = 1, \ldots p$. The degree of statement (1), that is, truth value (*Qy's are F*), is now calculated in two steps:

$$r = \frac{1}{p} \sum_{i=1}^{p} \mu_F(y_i) \tag{3}$$

$$\text{truth value}(\textit{Qy's are F}) = \mu_Q(r). \tag{4}$$

C. Soft Degree of Consensus

Here, we define a consensus measure which indicates the agreement between decision makers' opinions. We consider a "soft" degree of consensus as proposed by Kacprzyk and Fedrizzi [4]. In our context it is meant as the statement that: *"most of the* individuals agree in their preferences to most of the options.*"* Except of total agreement or disagreement between individuals as to the final decision, this approach allows some partial, acceptable consistency in the range [0,1].

The "soft" degree of consensus in the above sense is now obtained in three steps [5]:

1) for each pair of individuals we indicate a degree of agreement as to their preferences between all the pairs of options,

2) we aggregate these degrees to derive a degree of agreement of each pair of individuals as to their preferences between Q_1 (a linguistic quantifier as, e.g., "most") pairs of options,

3) we combine these degrees to obtain a degree of agreement of Q_2 (a linguistic quantifier similar to Q_1) pairs of individuals as to their preferences between Q_1 pairs of options and this is meant to be the degree of consensus.

We start with the degree of a sufficient agreement (at least to degree $\alpha \in [0,1]$) of individuals k_1 and k_2 as to their preferences between options s_i and s_j defined by

$$v_{ij}^{\alpha}(k_1, k_2) = \begin{cases} 1 & \text{if} \quad |r_{ij}^{k_1} - r_{ij}^{k_2}| \le 1 - \alpha \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where, $k_1 = 1, \ldots, m-1$, $k_2 = k_1 + 1, \ldots, m$, $i = 1, \ldots, n-1$, $j = i+1, \ldots, n$.

Then, the degree of agreement between individuals k_1 and k_2 as to their preferences between all the pairs of options is:

$$v(k_1, k_2) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} v_{ij}^{\alpha}(k_1, k_2). \tag{6}$$

Next, the degree of agreement between individuals k_1 and k_2 as to their preferences between Q_1 pairs of options is:

$$v_{Q_1}(k_1, k_2) = \mu_{Q_1}(v(k_1, k_2)). \tag{7}$$

The degree of agreement of all the pairs of individuals as to their preferences between Q_1 pairs of options is:

$$v_{Q_1} = \frac{2}{m(m-1)} \sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} v_{Q_1}(k_1, k_2). \tag{8}$$

Finally, according to the third step, the degree of agreement of Q_2 pairs of individuals as to their preferences between Q_1 pairs of options, called the degree of consensus, is:

$$\text{con}(Q_1, Q_2) = \mu_{Q_2}(v_{Q_1}). \tag{9}$$

III. INDICES OF CONSENSUS

A. Sensitivity of Individuals

The above definition of the "soft" consensus concerns most individuals as to most options without any distinguish between the individuals or the options. In this paper, we adopt a more flexible concept of a consensus reaching process which takes into account a sensitivity of individuals. This important component of the decision-making process is defined by the perturbation of every particular fuzzy preference relation matrix R_k.

If the fuzzy preference relation matrix is defined as $R_k = [r_{ij}^k]$, then the perturbed fuzzy preference relation matrix may be identified by $R_k^p = [r_{ij}^{kp}]$, such that: $[r_{ij}^{kp}] = [r_{ij}^k \pm a], a \in (0,1)$, for $i = 1,\ldots,n-1;\ j = i+1,\ldots,n,$ $k = 1,\ldots,m$. We also assume that $[r_{ji}^{kp}] = 1 - [r_{ij}^{kp}]$.

After matrix perturbation, we compute the degree of consensus for each individual which is now denoted as $con_k^p(Q_1, Q_2)$. Then, the measure of distance between $con(Q_1, Q_2)$ and $con_k^p(Q_1, Q_2)$ is obtained as:

$$d_k = | con(Q_1, Q_2) - con_k^p(Q_1, Q_2) | \qquad (10)$$

where $k = 1,\ldots,m$.

It is relevant for which individual small changes in fuzzy preference relation matrix cause the biggest change in the consensus degree. We obtain an ordered argument vector B where the b_i is the $i-th$ largest element (the most sensitive individual) among $\{d_1,\ldots,d_m\}$. B is called an ordered argument vector if each $b_i \in [0,1]$, and $j > i$ implies $b_i \geq b_j$, $i-1,\ldots,m$.

B. Option Pair Related Consensus Degree

Calculating the degree of "soft" consensus might derive additionally some partial indicators of consensus, like e.g. the option consensus degree which points out to the most controversial or popular options. Thus, this indicator facilitates the work of moderator by providing him with some hints as to the most promising directions of a further discussion.

The option pair related consensus degree [7] for options s_i and s_j, $OCD(s_i, s_j) \in [0,1]$, is the degree of truth value: "most pairs of individuals agree in their preferences in respect to the pairs of options s_i and sj ." It may be formally defined as:

$$OCD(s_i, s_j) = \mu_{Q_2}\left(\frac{2}{m(m-1)} \sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} v_{ij}^{\alpha}(k_1, k_2)\right). \qquad (11)$$

C. Cost of Changes

The cost of the entire consensus reaching process may be defined as the sum of absolute values of all changes in decision makers' preferences until the session ends [9], i.e.

$$cos\, t_{ij}^k(q) = \sum_{q=1}^{s} | r_{ij}^k(q) - r_{ij}^k(q-1) | \qquad (12)$$

where q denotes the iteration, $q \in [0, t]$.

IV. CONSENSUS REACHING SUPPORT SYSTEM

To clarify, initially preferences of the decision makers are far away from each other and this system aims at minimizing these distances [3]. Therefore, the moderator measures distances between individuals on each stage of the process and checks whether the consensus is reached (and process can be stopped)

$$con(Q_1, Q_2) \geq \beta \qquad (13)$$

where β indicates the acceptable degree of the consensus.

If the consensus level is not acceptable the moderator encourages appropriate individuals to update their preferences in order to improve the level of total agreement.

After calculating the consensus indicators (10) and (11), the moderator has to suggest the most sensitive decision makers to change their preferences in the most promising direction for a further discussion. Among the selected group of the most sensitive individuals the moderator finds the "typical preference relation" equited with their preference relations with respect to the pairs of options pointed out in (11). The "typical preference relation" is calculated by:

$$r_{ij}^c = \frac{\sum_{k=1}^{m} r_{ij}^k}{\sum_{k=1}^{m} r_{ij}^k + \sum_{k=1}^{m} (1 - r_{ij}^k)}. \qquad (14)$$

Then the moderator checks the relation:

$$| r_{ij}^k - r_{ij}^c | \leq \delta \qquad (15)$$

If the inequality (14) is not fulfilled then the new value of preference relation for each individual k is defined as a mean value between the typical preference relation and the former value of the preference relation of individual k, i.e., as an arithmetic average:

$$r_{ij}^k(q+1) = \frac{r_{ij}^k(q) + r_{ij}^c(q)}{2}. \qquad (16)$$

V. APPLICATION OF PROPOSED COMPUTER-BASED SYSTEM

The parameters applied to the group consensus reaching support system are:

a) $M = 10, N = 10$
b) Acceptable degree of the soft consensus (13) is: $con(Q_1, Q_2) > 0.7$

c) $\alpha = 0.8$ in $v_{ij}^{\alpha}(k_1, k_2)$ in (5)

d) $\delta = 0.1$ in (15) which denotes that almost everyone is supposed to update his opinion even with the small step.

The initial degree of consensus was equal to $con(Q_1, Q_2) = 0.38$ which was definitely below the acceptable agreement.

The sensitivity of individuals calculated in (10) was: $B = \{0.026;0.024;0.021;0.01;0.01,0.006;0.006;0.004;0.004,\ldots$ $\ldots 0.002\}$.

Hence the ordered argument vector of the most sensitive individuals was: $E = \{e_5, e_2, e_8, e_6, e_3, e_4, e_{10}, e_1, e_9, e_7\}$.

After calculating the indicator (11) for each pair of options, we discovered that there were almost no difference between the ordering options from the most preferred (promising) to the worst and inversely. This dependency is exemplified on Fig.1.

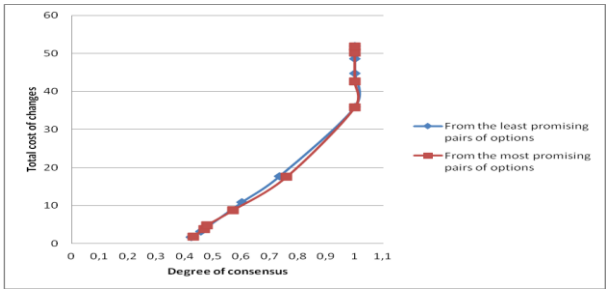


Fig. 1 Comparison of the degree of consensus and the total cost of changes for different direction of option consensus degree for 10 decision makers

However, for a smaller group of individuals (3,4 or 5) this exponential growth has differences in favor of the direction from the most promising pairs of options. TableI presents the maximum degree of the consensus obtained by a different number of sensitive individuals during the update process.

TABLE I.
MAXIMUM DEGREE OF CONSENSUS OBTAINED BY A DIFFERENT NUMBER OF SENSITIVE INDIVIDUALS COMPARED WITH THE TOTAL COST

Number of most sensitive individuals	Degree of Consensus	Total cost
3	0,61	16,86
4	0,67	22,13
5	0,77	26,73
6	0,87	32,63
7	0,99	37,23

Clearly, we can easily see an improvement in the value of the degree of consensus achieved in the group of individuals. It is also noticeable that at least the group of 5 allows us to obtain an acceptable agreement among decision makers.

VI. CONCLUSION

In this paper we proposed a new method for improving the degree of total agreement among decision makers in a consensus reaching process. We applied different procedures to find some useful indicators which allow us to run the process in the more efficient way. These procedures are to be further extended so that the improvement might take into account many aspects of this multi-criteria problem, e.g. the new optimization methods to find a best solution in the sense of aggregation either the improvement of the final degree of consensus or the total cost of changes between the decision makers during the process.

ACKNOWLEDGMENT

Dominika Gołuńska is partially supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

REFERENCES

[1] Center for Excellence in Government, Facilitator's Toolbox, 1996, http://www.employeesu.com/docs/EmployeeResources/Creating%20Consensus.pdf, [date of access: 21.05.13]
[2] Hanson S.O., Decision Theory: A Brief Introduction, Philosophy, Volume: 23, Royal Institute of Technology, Stockholm, 1994, 1-94.
[3] Kacprzyk J., Falkiewicz D., Different aspects of supporting group consensus reaching process under fuzziness, Technical Transactions, Wydawnictwo Politechniki Krakowskiej, vol. 1-AC, ss. 17-27, 2012.
[4] Kacprzyk J., Fedrizzi M., A 'soft' measure of consensus in the setting of partial (fuzzy) preferences, European Journal of Operational Research, vol. 34, 1988, pp. 315-325.
[5] Kacprzyk J., Fedrizzi M. and Nurmi H., Soft degrees of consensus under fuzzy preferences and fuzzy majorities, in J. Kacprzyk, H. Nurmi i M. Fedrizzi (Eds.): Consensus under Fuzziness, Kluwer, Boston, 1996, pp.55 – 83.
[6] Kacprzyk J., Zadrożny S., On a concept of a consensus reaching process support system based on the use of soft computing and Web techniques. In: D. Ruan, J. Montero, J. Lu, L. Martínez, P. D'hondt, E.E. Kerre (Eds.): Computational Intelligence in Decision and Control. World Scientific, 2008, pp. 859-864
[7] Kacprzyk J., Zadrożny S., On the use of fuzzy majority for supporting consensus reaching under fuzziness, Proceedings of FUZZ-IEEE'97 - Sixth IEEE International Conference on Fuzzy Systems (Barcelona, Spain), vol.3, 1997, pp. 1683 – 1988.
[8] Kacprzyk J., Zadrożny S., Soft computing and Web intelligence for supporting consensus reaching, Soft Computing, vol.14, no. 8, 2010, pp.833-846.
[9] Kacprzyk J., Zadrożny S., Wilbik A.: Linguistic summarization of some static and dynamic features of consensus reaching. In: B. Reusch (Ed.): Computational Intelligence, Theory and Applications. Springer-Verlag, Berlin Heidelberg 2006, pp. 19-28.
[10] What is a Group Decision Support System (GDSS)?, Decision Support Systems Resources, http://www.dssresources.com/, [date of access: 21.05.2013]
[11] Xia M., Xu Z., On consensus in Group Decision Making Based on Fuzzy Preference Relations, Studies in Fuzziness and Soft Computing, Springer, vol. 267, , 2011, pp. 263-287.
[12] Zadeh, L, A computational approach to fuzzy quantifiers in natural languages, Computers and Mathematics with Applications, no. 9, 1983, pp.149-184.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 651–654

Linguistic knowledge about temporal data in Bayesian linear regression model to support forecasting of time series

Katarzyna Kaczmarek
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
Email: K.Kaczmarek@ibspan.waw.pl

Olgierd Hryniewicz
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
Email: Olgierd.Hryniewicz@ibspan.waw.pl

***Abstract*—Experts are able to predict sales based on approximate reasoning and subjective beliefs related to market trends in general but also to imprecise linguistic concepts about time series evolution. Linguistic concepts are linked with demand and supply, but their dependencies are difficult to be captured via traditional methods for crisp data analysis. There are data mining techniques that provide linguistic and easily interpretable knowledge about time series datasets and there is a wealth of mathematical models for forecasting. Nonetheless, the industry is still lacking tools that enable an intelligent combination of those two methodologies for predictive purposes. Within this paper we incorporate the imprecise linguistic knowledge in the forecasting process by means of linear regression. Bayesian inference is performed to estimate its parameters and generate posterior distributions. The approach is illustrated by experiments for real-life sales time series from the pharmaceutical market.**

***Index Terms*—linguistic knowledge, time series analysis, Bayesian linear regression, posterior simulation**

I. Introduction

HUMAN-BEINGS have the unique ability to process imprecise information and solve complex problems based mostly on their intuition and expertise [14]. This ability allows us to easily interpret and describe temporal data in natural language with words and propositions. Such information, often imprecise, is called *temporal linguistic knowledge* within this paper.

As observed in a selected pharmaceutical company experts were able to predict future sales and make related decisions based on approximate reasoning about imprecise information driven from visual inspection of time series data sets. It was observed that experts recognized important dependencies between linguistic temporal knowledge even in situations when analysis of crisp time series datasets showed no significant correlations. We pose the question whether linguistic temporal knowledge may bring information about new correlations useful for the time series forecasting process.

The linguistic knowledge about temporal data is provided by the experts of selected domain or is extracted automatically thanks to the knowledge discovery and data mining techniques. Among the recent developments in the field of intelligent computing there are efficient methods that provide interpretable knowledge from huge datasets.

The problem of time series abstraction and labeling meaningful intervals by means of clustering, machine learning and function approximation methods, statistical test or multiscale methods is addressed e.g. in [2], [10], [13]. The concept of pattern recognition has been widely discussed for example in [7], [8], [9], [11], [12]. One of the goals of data mining research is to provide linguistic and human-consistent description of raw data. Within this paper we take data mining results as the input for the forecasting procedure.

We provide a predictive model to support decision making in the international pharmaceutical sales market. Linguistic knowledge about temporal data is transformed into imprecisely labeled sequences that are incorporated into the probabilistic model as explanatory variables. We adopt Bayesian linear regression model and perform posterior simulation. We operate on parameters for which linguistic concepts are transparent and could easily be interpreted by experts.

The structure of this paper is as follows. Next chapter introduces basic definitions related to temporal linguistic knowledge about time series. Chapter 3 presents the forecasting process with temporal linguistic knowledge incorporated into the regression model. The description of the experiments and results for time series from the pharmaceutical industry are gathered in chapter 4. Paper concludes with general remarks and further research opportunities.

II. Linguistic concepts about temporal data

In this section we define formal language for temporal linguistic concepts that we consider most appealing for predictive purposes.

Let $O = \{o_1, o_2..., o_q\}$ denote a finite set of objects in a considered domain. The properties of objects are measured by observables. Let $M = \{m_1, m_2..., m_r\}$ denote a finite set of observables in the considered domain.

Definition 1: ***Object's property***
A pair (o, m) *such that* $o \in O$ *and* $m \in M$ *is called object's property.*

The sequence of measurements for object's property is treated as discrete time series.

978-83-60810-53-8/$25.00

TABLE I
ILLUSTRATIVE EXAMPLES FOR IMPRECISE LABEL, OBSERVABLE AND OBJECT.

Domain	Object	Observable	Imprecise label
Pharmaceutical market	Product 1	sales	high
	Product 1	supply	high
	Product 1	sales	increasing
	Europe	inflation	increasing
Mood tracking	Patient A	anxiety	high
	Patient A	weight	constant
	Patient A	hours slept	constant
	Patient B	medication	increasing

Definition 2: ***Discrete time series***
Discrete time series $\{y_t\}_{t=1}^n \in \Psi_n$ *is a sequence of observations of given object's property* (o, m) *such that* $o \in O$ *and* $m \in M$ *measured at successive* $t \in T = \{1, ..., n\}$ *moments and at uniform time intervals. For each* $t \in T$ *the observation* y_t *is a realization of the random variable* Y_t*. Random variables* Y_t *are defined on the probability space* (Ω, A, P)*, where* Ω *is the set of all possible outcomes of the random experiment,* A *is a* σ*-field of subsets of* Ω*, and* P *is a probability measure associated with* (A, P).

As stated in [1] during visual inspection people perceive and process shapes rather than single data points. We describe the evolution of time series with adjectives like *high, medium, low, light, heavy, interesting, increasing, constant, decreasing, interesting, long, short, strong, weak, slight, etc.* Such adjectives refer to imprecise values, trends, judgments or features and are called *imprecise labels* within this paper.

Let $S = \{s_1, s_2..., s_l\}$ denote a finite set of imprecise labels referring to either qualitative or quantitative measurements for observables applicable in the considered domain. Depending on the context, values for the imprecise label are assigned subjectively by experts or are calculated based on the fuzzy numbers and membership functions. For basic definitions related to the fuzzy sets theory see e.g. [6].

In real-life situations understanding and interpretation of imprecise labels depends on context and may change in time. Within this approach we assume one interpretation that is constant in time.

As presented in Table I for application *sales, supply* and *inflation* are observables when considering application in *pharmaceutical market*. If the problem of *mood tracking to support medical diagnosis* is considered, the observables measure the *anxiety* or *weight* of a patient. However, the general idea of visual inspection and processing trends is the same regardless of the practical context.

Definition 3: ***Imprecise labeled sequence***
Let $f : \Psi_n \times S \times T \rightarrow [0, 1]$ *denote function assigning the degree of truth that label* $s \in S$ *applies at the moment* $t \in T$ *for object's property measured by time series* $y \in \Psi_n$*. Imprecise labeled sequence* $\{x_t^{s,y}\}_{t=1}^n$ *is calculated from* $x_t^{s,y} := f(y, s, t)$.

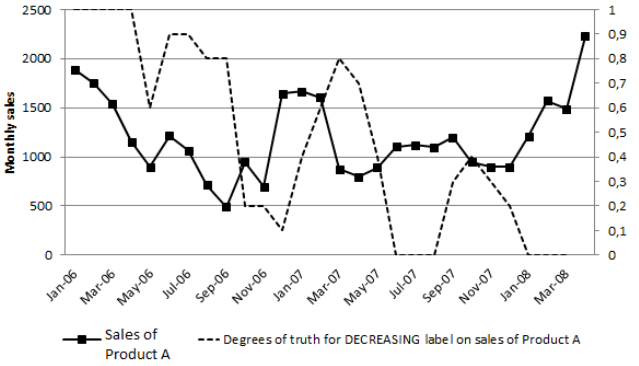


Fig. 1. Example of time series and imprecise labeled sequence.

Fig. 1 presents an illustrative example of sales time series and its imprecise labeled sequence. In this context, for each observation of time series representing sales of Product A, the expert subjectively assigned the degree of truth for *decreasing* trend. Imprecise labeled sequences are processed within the presented approach.

III. BAYESIAN REGRESSION WITH LINGUISTIC KNOWLEDGE

The forecasting procedure consists of the phase of processing temporal data and the posterior simulation. As outlined in Fig. 2 the input for the model are discrete time series and definitions of the linguistic concepts. As a result of the model, the forecast and its regressive components are provided.

A. Processing temporal data

Let $Y_n^k = \{\{y_t^1\}_{t=1}^n, ..., \{y_t^k\}_{t=1}^n\}$ denote k-vector of multivariate discrete time series. Let $\{y_t^1\}_{t=1}^n$ denote a time series of object's property to be predicted. For clarity reasons, we limit considerations to the one-step-ahead forecast and the vector of interest ω contains one element $\omega = \{y_{n+1}^1\}$. Predictions for longer horizons are iterated by repeating the procedure.

For $s \in S$ and $k-1$ time series $y \in \{\{y_t^2\}_{t=1}^n, ..., \{y_t^k\}_{t=1}^n\}$, imprecise labeled sequences $\{x_t^{s,y}\}_{t=1}^n$ are created based on data mining techniques or as a result of subjective expertise. Sequences $\{x_t^{s,y}\}_{t=1}^n$ interpreted as degree of truth that the imprecise label is valid at each moment for given object's property, represent the linguistic knowledge for the regression model.

B. Posterior simulation

Imprecise labeled sequences $\{x_t^{s,y}\}_{t=1}^n$ are included into the linear regression as explanatory variables. We adopt the multiple normal linear regression model which can be written as:

$$y = X\beta + \epsilon,\ \epsilon\ \sim N(0, \sigma^2 I_n)$$

where X is the $n \times ((k-1) \times l)$ matrix of explanatory variables, y is the $n \times 1$ vector of dependent variables and ϵ is an $n \times 1$ vector of independent identically distributed normal

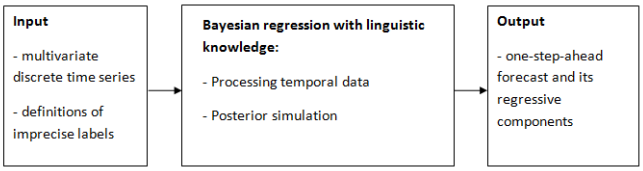


Fig. 2. Overview of the forecasting procedure based on Bayesian regression model with temporal linguistic knowledge.

random variables. We perform Bayesian inference to estimate the vector of parameters $\theta = (\beta, \sigma)$.

Following definition of Geweke [4], [5] the complete model A for Bayesian inference consists of:

1) the observables density

$$p(Y_t|\theta_A, A) = \prod_{t=1}^{T} p(y_t|Y_{t-1}, \theta_A, A)$$

in which $\theta_A \in \Theta_A$ is a $k_A \times 1$ vector of unobservables

2) the prior density $p(\theta_A|A)$

3) the vector of interest density (the posterior density)

$$p(\omega|y^o, A) = \int_{\Theta_A} p(\omega|y^o, \theta_A, A)p(\theta_A|y^o, A)\mathrm{dv}(\theta_A)$$

$$p(\theta_A|y^o, A) = \frac{p(\theta_A|A)p(y^o|\theta_A, A)}{p(y^o|A)}$$

The problem statement is to find a decision, known in Bayesian theory as an action a, which minimizes the following equation:

$$E[L(a, \omega)|y^o, A] = \int_{\Theta_A} L(a, \omega)p(\omega|y^o, A)\mathrm{d}\omega$$

Posterior predictive distributions are approached by means of Markov Chain Monte Carlo Methods (MCMC). Posterior simulation yields a pseudo-random sequence of the vector of interest to estimate its posterior moments. MCMC were initially developed in 1940s and gained popularity thanks to their great success in practical applications [5].

For simplicity, we assume that prior for β is independent from prior for σ and we apply Gibbs Sampling which leads to sampling from multivariate probability density. The sampling procedure begins with arbitrary values for β^0 and σ^0, computes mean and variance of β^0 conditional on the initial value σ^0, uses the computed mean and variance to draw a multivariate normal random vector β^1 and uses the β^1 with a random draw to determine σ^1.

Details for posterior density, drawings construction and sampling algorithm are available in [5]. Gelfand and Smith [3] proved that with Gibbs Sampler large sets of draws converge in the limit to the true joint posterior distribution of parameters.

The results for the linear regression model with linguistic knowledge are the predictive distribution for the future observations of the time series of interest and the model parameters. Parameters are easily interpreted in a natural language as they are directly linked with imprecise labels.

IV. Experimental results

The purpose of this experiment is to illustrate the performance of the forecasting method for real-life data at the example of sales time series from the pharmaceutical industry.

Train dataset consists of 6 normalized time series representing monthly sales of different products in the period from Jan'05 to Dec'09. Fig. 3 shows exemplary time series from the train dataset. The test dataset contains 6-month-long sales continuation for each product and is used for evaluation.

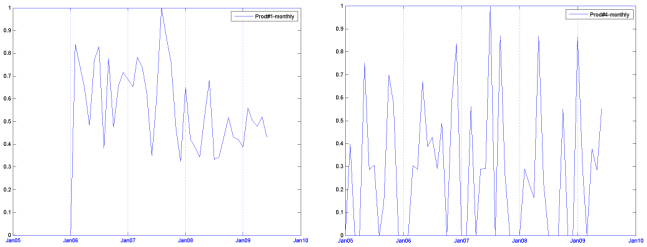


Fig. 3. Exemplary time series from the train dataset representing monthly sales of Product No.1 and Product No. 4.

We consider following 6 imprecise labels in the experiment: *low, medium, high, increasing, constant, decreasing*. Values for imprecise labeled sequences referring to *increasing, constant, decreasing* labels are defined based on experts' subjective beliefs. For labels: *low, medium, high* triangular fuzzy numbers are constructed based on the minimum, average and maximum values calculated from the time series. Then, imprecise labeled sequences are calculated from appropriate membership functions.

We first analyze correlations between the time series to be predicted and the imprecise labeled sequences to verify whether the imprecise linguistic knowledge may bring valuable information in the linear regression model. For each product we compare Pearson correlation coefficient between its sales time series and the imprecise labeled sequences derived for other products.

As demonstrated by the results in Table II correlations between sales time series and imprecise labeled sequences range from 0,10 to 0,14 and are on average by 20% higher than between different sales time series itself. Correlation coefficients are on average higher for labels of imprecise trends than values.

TABLE II
Correlation coefficients between sales time series and imprecise labeled sequences (ILS)

	Mean	Median	StdDev
Sales vs Sales	**0,10**	**0.08**	**0.05**
Sales vs Increasing	0,14	0.14	0.04
Sales vs Constant	0,12	0.13	0.06
Sales vs Decreasing	0,13	0.10	0.07
Sales vs Low	0,11	0.10	0.05
Sales vs Medium	0,12	0.12	0.05
Sales vs High	0,10	0.08	0.04
Sales vs All ILS	**0,12**	**0.11**	**0.05**

Table III provides detailed correlations per product. It is interesting to observe that for example the correlation coefficient (0,21) between sales time series of Product No. 2 and decreasing trends of other products is higher than correlation coefficient (0,17) between Product No. 2 sales time series and other sales time series itself.

TABLE III
CORRELATION COEFFICIENTS PER PRODUCT

	P1	P2	P3	P4	P5	P6
Sales vs Sales	0,16	**0,17**	0,05	0,09	0,04	0,07
Sales vs Increasing	0,17	0,18	0,10	0,11	0,16	0,11
Sales vs Constant	0,13	0,21	0,12	0,13	0,03	0,07
Sales vs Decreasing	0,22	**0,21**	0,09	0,10	0,06	0,09
Sales vs Low	0,19	0,17	0,11	0,07	0,06	0,09
Sales vs Medium	0,19	0,14	0,16	0,05	0,09	0,08
Sales vs High	0,13	0,15	0,06	0,09	0,07	0,08

The second step of the experiment is the comparative analysis of the forecasts' accuracy of the Bayesian regression model with linguistic knowledge (BRLK) and the traditional Vector Autoregression (VAR). Table IV summarizes mean absolute percentage error (MAPE) and deviation (MAPD).

TABLE IV
MEAN ABSOLUTE PERCENTAGE ERROR(MAPE) AND DEVIATION(MAPD) FOR 6- AND 1-STEP-AHEAD FORECAST OF BRLK AND VAR

	h=6 MAPE BRLK	h=6 MAPE VAR	h=6 MAPD BRLK	h=6 MAPD VAR	h=1 APE BRLK	h=1 APE VAR
P1	0,173	0,199	0,101	0,129	0,036	0,129
P2	0,425	0,509	0,193	0,165	0,202	0,317
P3	0,696	0,476	0,210	0,278	0,900	0,745
P4	0,282	0,396	0,371	0,237	0,026	0,326
P5	0,389	0,459	0,317	0,177	0,188	0,364
P6	0,444	0,371	0,281	0,222	0,703	0,559
All	**0,402**	**0,402**	**0,246**	**0,201**	**0,342**	**0,407**

As demonstrated by results in Table IV absolute percentage error for 1-step-ahead forecast is 0,342 and 0,407, respectively for BRLK and VAR. For 6-month-long forecast MAPE is the same and relatively high for both models, and amounts to 0,402. Forecasts generated by VAR are characterized by a lower standard deviation.

We conclude that Bayesian regression model with linguistic knowledge and VAR models are comparable in terms of forecasts' accuracy. The Bayesian regression model with linguistic knowledge delivers forecasts of a higher interpretability than traditional VAR as its components are naturally linked with the linguistic concepts.

V. CONCLUSION

The performed experiment confirmed that the approach with additional linguistic knowledge is adequate to support sales forecasting. Imprecise labeled sequences enable to discover new correlations in the dataset that lead to construction of the linear regression model. Produced forecasts are accurate on a similar level as forecasts provided by Vector Autoregression.

The main advantage of the proposed solution is the easy interpretation of predictions and model parameters required for forecasting process, which is of special importance for experts involved in real-life forecasting for large datasets. The proposed solution is in line with visual pattern recognition capabilities of humans and delivers additional knowledge about dependencies in multivariate time series datasets.

Next experiments for multivariate time series from other domains and on benchmark data are planned in order to analyze further benefits and limitations of the proposed technique. Another topic planned to be explored is the introduction of multiple interpretation for imprecise labels.

Within the approach simple forms of linguistic knowledge are considered. The potential to include advanced forms of linguistic knowledge like imprecise features, frequent temporal patterns, association rules and temporal linguistic summaries remains open for future research.

ACKNOWLEDGMENT

Katarzyna Kaczmarek is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from the European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

REFERENCES

[1] F. Attneave, "Some informational aspects of visual inspection," *Psychological Review*, 61(3), 1954.
[2] G. Das, K. Lin, H. Mannila, G. Renganathan, and P. Smyth, "Rule discovery from time series," *Proc. of the 4th Int. Conf. on Knowl. Discovery and Data Mining*, 1998, pp. 16-22.
[3] A. Gelfand and A. Smith, "Sampling-based approaches to calculating marginal densities," In *Journal of the American Statistical Association*, Vol. 85, 1990, pp. 398-409.
[4] J. Geweke, *Contemporary Bayesian econometrics and statistics*. Wiley series in probability and statistics. John Wiley, 2005.
[5] J. Geweke and C. Whiteman. *Handbook of Economic Forecasting*, volume 1, chapter Bayesian Forecasting, Elsevier, 2006, pp. 3-80.
[6] M. Gil and O. Hryniewicz. "Statistics with imprecise data," In *Encyclopedia of Complexity and Systems Science*, Springer, 2009, pp. 8679-8690.
[7] F. Höppner. *Knowledge Discovery from Sequential Data*. PhD thesis, 2002.
[8] A. Wilbik and J. Kacprzyk. "Temporal sequence related protoform in liguistic summarization of time series," *Proc. of WConSC*, 2011.
[9] S. Kempe, J. Hipp, C. Lanquillon, and R. Kruse, "Mining frequent temporal patterns in interval sequences," *Fuzziness and Knowledge-Based Systems* in International Journal of Uncertainty, Vol. 16, No. 5, 2008, pp. 645-661.
[10] F. Klawonn and R. Kruse, "Derivation of fuzzy classifcation rules from multidimensional data," *Advances in Intelligent Data Analysis* Windsor, Ontario., 1995, pp. 90-94.
[11] F. Mörchen, I. Batal, D. Fradkin, J. Harrison and M. Hauskrecht "Mining recent temporal patterns for event detection in multivariate time series data," *KDD*, 2012, pp. 280-288.
[12] S. Schockaert and M. De Cock "Temporal reasoning about fuzzy intervals," *Artificial Intelligence* 172, 2008, pp. 1158-1193.
[13] A. P. Witkin, "Scale space filtering," *Proc. of the 8th Int. Joint Conf. on Artifficial Intelligence*, Karlsruhe, Germany, 1983, pp. 1019-1022.
[14] L. A. Zadeh, "From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," In *Intelligent Systems and Soft Computing*, volume 1804 of *Lecture Notes in Computer Science*, Springer, 2000, pp. 3-40.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 655–658

Improving the accessibility of touchscreen-based mobile devices: Integrating Android-based devices and Braille notetakers

Daniel Kocieliński
Institute of Mathematical
Machines ul. Ludwika
Krzywickiego 34,
02-078 Warszawa, Poland
Email: d.kocielinski@imm.org.pl

Jolanta Brzostek-Pawłowska
Institute of Mathematical
Machines ul. Ludwika
Krzywickiego 34,
02-078 Warszawa, Poland
Email: j.brzostek@imm.org.pl

***Abstract* — The article presents the concept and pilot implementation of wireless (Blutetooth-based) integration of the Braille notetaker environment and the environment of touchscreen-based devices (such as smartphones) operating under the Android system. Advanced functions of Android-based devices are hardly accessible to the blind using a touchscreen; one aim of such integration is to enable accessing them with a notetaker. Another is to allow the blind who work with notetakers on a daily basis and use common touchscreen-based smartphones and tablets to write using the physical Braille keyboard of a notetaker as well as its editing functions; this would solve many problems encountered and prevent numerous errors made by the blind when using the virtual QWERTY keyboard of a touchscreen-based device. Pilot implementation of the concept included developing a communication protocol for a notetaker operated under Windows CE and an Android-based smartphone; services to be provided to notetakers by smartphones have been developed as well. The implemented services dealt with managing contacts and composing messages – operations that normally require considerable interaction with a QWERTY keyboard. Favourable results of initial research on pilot implementation conducted among the blind indicate a need for further development of this concept.**

I. Introduction

TOUCHSCREENS intended for the operation of graphical user interfaces are increasingly common in mobile devices and computers both used privately and made available to the public. Such solutions are hardly accessible to the blind: it is difficult for them to locate and select items visualised on the screen. Assistive technologies offered by mobile device manufacturers, such as Apple's VoiceOver used in iOS-based devices or Google's TalkBack and BrailleBack used in Android-based devices, do improve accessibility, but do not eliminate all the obstacles. One example of such barrier is the virtual QWERTY keyboard, where punctuation marks and other special characters are difficult to enter (it requires switching keyboard operation mode) and there is no point of reference like the bossed "J" key of a physical keyboard. Other examples include lack of haptic points on the screen (these would improve spatial orientation) and poor suitability of touchscreen gestures for the blind (they prefer gestures starting on the edge of the screen or in its close vicinity). The existing barriers create a need for research and new solutions that would improve the accessibility of touch interfaces. The concept of such solution presented in this article is based on two assumptions, tested positively for example in [3] and [4]:

- most of the tested blind smartphone users deem notetakers to be indispensable, especially for taking notes quickly;
- the use of a linear-sequential Braille interface by the blind, especially for entering text using a physical Braille keyboard, is much more efficient than the use of a virtual QWERTY or Braille keyboard.

Since, according to the research, smartphones and notetakers are indispensable in daily use for most of the blind, the main idea of the concept is to take advantage of the synergy obtained by functionally integrating devices of both types. This synergy results from combining their essential qualities: the efficient Braille interface of a notetaker and the advanced functions of smartphones and tablets. The use of a notetaker to improve touch interface accessibility gives more than just the possibility of using a physical Braille keyboard (which is accomplished by using BrailleBack app). Notetakers – computers operating under systems such as Windows CE – can provide programmable access to smartphone functions, which makes these functions available via the Braille interface. Besides, combined with a smartphone, a notetaker serves as a smart Braille keyboard with functions for efficient (quick, less error-burdened) text editing.

Advanced smartphone functions are made available through specific software that has been developed as part of concept implementation, operates under the Android system of a smartphone, and provides services for the notetaker. These services are called by specific software operated under system (Windows CE) of the notetaker and developed as part of the concept. The lowest level of integration includes communication software that runs on both devices and connects either the smartphone to the notetaker (for example for text editing) or the other way (in order to access the selected advanced smartphone function to use it as a remote service on the notetaker).

978-83-60810-53-8/$25.00

The pilot implementation of the concept involved a Samsung Galaxy S III smartphone and a Polish notetaker Kajetek SD, see [9]; it included access to contact management and message composing services as well as editing text functions (remotely serving by the smartphone, realized by a blind user on the notetaker). Results of pilot implementation tests conducted on 7 blind users with different Braille and technical experience indicate a need for further development of the concept. The testers confirmed that the initiative was heading in the right direction: for solving touch interface accessibility problems encountered by the blind. The presented considerations and research results consist the area of dissertation work of the co-author, Daniel Kocieliński.

II. Related work

The research is focused on overcoming the two greatest difficulties faced by the blind when using touchscreen-based mobile devices: enabling efficient navigation through graphical elements and their correct selection and developing a method for quick and correct text input. There are many different ways to interact with interfaces of touchscreen-based devices: a simple single tap, several simultaneous or successive taps, directional and scanning gestures as well as fixed and adaptive layouts of function fields comprising the screen.

However, software developers often neglect to adapt their methods to the needs of a blind user. Findings of Kane et al. presented in [1] point to differences in preferred gestures and ways of making them between sighted and blind users. Oliviera et al. demonstrated in [2] that the blind input text using different methods more or less efficiently depending on their personality traits and personal experience. D'Andrea found out that most of the tested blind students prefer using smartphones with Braille notetakers and value their Braille skills, especially for the possibility of taking notes, see [3]. Southern et al. in turn noticed lower erroneousness in case of entering text in Braille, using a physical or virtual keyboard, see [4] and [2]; according to [4], the use of a physical Braille keyboard is the most efficient method. Azenkot et al. confirmed these findings in [5] and [4], comparing the more efficient various Braille text input methods to entering text using a QWERTY keyboard and with the aid of Apple's VoiceOver. The virtual QWERTY keyboard has been enhanced e.g. by Findlater et al. in [6], applying extensions available through specific user gestures. In [7] Costagliola and Capua suggest implementing into the virtual keyboard an additional menu displayed upon making certain gestures around a given character. In [8] Ruamviboonsuk et al. suggest entering numbers sightlessly (to enable quicker dialling), using specific multi-touch combinations. Improvements suggested in [6], [7] and [8] have not been tested by the blind nor designed for this group of users.

In [4] the authors suggest a fixed virtual Braille keyboard that would have 6 keys: 3 on the left and 3 on the right side of the screen, near its edges. [5] in turn suggests a dynamic virtual Braille keyboard appearing where 3 fingers of one hand or 6 fingers of two hands are placed.

III. Integration concept

The main idea behind the presented research is to provide blind users with intuitive, friendly access to the advanced functions of touchscreen-based devices (such as smartphones). As a mean of improving accessibility of smartphone and tablet (hereafter referred to as smartphones) functions we propose using Braille notetakers, which are common among the blind, and Bluetooth communication mechanisms for exchanging data between touchscreen-based devices (such as smartphones) and notetakers. The adopted concept assumes that establishing a wireless connection between the notetaker and a specific smartphone will allow the user to access its advanced functionalities (given by services of Android OS and installed apps) using the physical keyboard of the notetaker and its well-known, convenient Braille interface.

As opposed to the case of graphical smartphone interfaces, where users interact with applications by making touch gestures on items displayed on the screen, operation of notetakers is based mainly on Braille keystrokes and sequential-hierarchical access to functions. Preferred interaction methods, which speed up and facilitate blind users' actions and are usually integrated in notetakers, include:

- function selection lists activated by pressing appropriate navigation keys,
- handy command menus available through hierarchical navigation or called using keystrokes,
- sets of Braille keystrokes that provide access to all the important notetaker functions at any time,
- editing functions that allow to enter text using six- and eight-dot Braille and navigation through the text quickly.

The aim of our research is to find ways of translating the graphical interface of touch-based devices (specially of smartphones) into a Braille interface that would include navigation and text input methods preferred by the blind. Our pilot work resulted in formulating basic assumptions concerning wireless operation of a smartphone with the Braille interface of a notetaker.

Such integration of the two environments – dedicated software of the notetaker and Android OS of the smartphone – allows a blind user to easily use such advanced functions of smartphones as speech recognition, cloud data (e.g. contacts) management, file sharing (e.g. Dropbox), web-based search engines, instant messaging, phonecalls or GSM and GPS navigation. Our concept makes it possible to use all the functions of a smartphone, as remote services, via a notetaker and eliminates the need for development aimed at implementing smartphone functions into notetakers. This way the blind would be able to use, conveniently, common touchscreen-based devices.

IV. Methodology

A. Implementation technology and test equipment

In order to test the concept we have developed pilot software consisting of two interacting modules: one running on an Android-based smartphone and one on a Kajetek SD

notetaker operated under Windows CE 5. Both modules support wireless Bluetooth connectivity and allow:

1) operating the smartphone with the physical keyboard of the Kajetek notetaker and
2) direct access to the advanced functions (as dedicated services) of the Android-based device.

The Android software has been implemented in Java, using Android API level 17, and tested on a Samsung Galaxy S III smartphone. During tests (navigation the contact list and input new contacts by virtual QWERTY keyboard) feedback was realised using text-to-speech, with the TalkBack screen reader.

The notetaker software is a C++ library that extends Kajetek application with support for a protocol used to communicate with the Android module. The communication protocol for the research has been developed as a sequence of XML query-response messages; the Bluetooth protocol serves as the transport layer.

B. Scope

During the research the basic concept assumptions have been tested by managing contacts using the remote interface between smartphone and notetaker. The pilot software allows the user to operate the smartphone's contact services using braille interface of Kajetek. In this way he or she can quickly navigate through contact items using notetaker keystrokes. The following contact management services have been implemented, as the remote interface, for the Kajetek notetaker:

- browsing contact lists (using specific Braille keystrokes),
- reviewing details of selected contacts,
- adding new contacts (using the editing functions integrated in Kajetek software),
- deleting contacts.

To find an appropriate input method that provides fast text entry we also measured performance of different input methods. We considered three ways of composing messages to selected recipients: using a virtual QWERTY keyboard, a virtual Braille keyboard, and the editing functions of the notetaker. Practical tests of text input speed have been conducted on a virtual QWERTY keyboard and on the physical keyboard of the Kajetek notetaker.

So, the test scenario included:

- navigating the contact list using the manufacturer's contact management app and TalkBack screen reader,
- navigating the contact list remotely, using the Kajetek Braille interface and appropriate services on the smartphone,
- adding new contacts using the manufacturer's contact management app and a virtual QWERTY keyboard,
- adding new contacts using Kajetek with its editing functions and the remote notetaker operation services on the smartphone.

C. Pilot test group

The tests have been conducted on a group of 7 users with different experience in the use of smartphones. Three of users are experienced in the use of smartphones: from a few weeks to a couple months; where one of the testers is an experienced iPhone and VoiceOver user. Next 3 of the users were able to use smartfone after training part of the tests. Seventh person said that touchscreen is too big challenge and gave up in tests on a virtual QWERTY keyboard of smartphone.

All testers know Braille well and are experienced in the practical use of Braille notetakers (5 of them are using Braille notetakers in daily work).

V. RESULTS

The results of tests proved that:

- item selection (tested on a list with 100 items) by interface of Braille notetaker is 2-3 times quicker than the item selection by the standard app of Android system (approximately from 4 to 10 seconds when using notetaker, and from 11 to 28 seconds when using standard app);
- the use of the editing functions of a Braille notetaker is much more efficient than using a virtual QWERTY keyboard (the text input speed for the slowest and the fastest tester reached, respectively, from 11 to 38 characters per minute – CPM when using a QWERTY keyboard, and from 87 to 173 CPM when using a notetaker).

Feedback from the test group confirms that the existing accessibility mechanisms of the Android system (such as TalkBack and BrailleBack), especially the ones related to navigation based on the available gestures (e.g. implemented through TalkBack) and text input using a virtual QWERTY keyboard, are subject to significant limitations. The tests have proven that managing contacts with the interface of the Kajetek notetaker and its physical Braille keyboard is more efficient than doing so using the standard smartphone application for managing contacts combined with the TalkBack screen reader, gestures and the virtual QWERTY keyboard. Selecting desired contacts from the list using Kajetek is more efficient primarily because of lesser erroneousness of browsing actions; operating a touch interface requires greater manual precision. One of the testers suggested extending the list item selection functionality of Kajetek e.g. with a quick search function looking for the first few entered letters of a contact.

The received feedback also confirms that the existing smartphone accessibility mechanisms related to text input are insufficient. Message composing tests indicated a significant advantage of the physical Braille keyboard of a notetaker over a virtual QWERTY keyboard.

VI. CONCLUSIONS

The results of preliminary research on pilot implementation into the Android system, concerning functions like managing contacts and composing messages to selected recipients, confirm that the adopted concept of using interaction mechanisms for touchscreen-based devices and Braille notetakers is a promising direction for further research and give grounds for continuing works. Practical tests of pilot application have shown that the use of a notetaker to operate the interface of a smartphone is not only

a solution that improves a blind user's work efficiency and convenience. The tests proved that quick navigation and editing functions of a Braille notetaker can be great solution for the blind users for example with a dysfunction of touch of sense. The research suggests new ways of solving the essential problem, which is the inaccessibility of new, advanced technologies to the blind. The main cause of this problem is that hardware manufacturers and software developers usually marginalise this group of users by providing (mostly behindhand) only accessibility tools (such as screen readers) that rather adapt methods of operating graphical touch interfaces than provide other, suitable ones. The proposed solution for improving the accessibility of touch interfaces is quite unusual compared to those currently provided in smartphones: it is based on using a Braille interface, smart notetaker mechanisms and a dedicated communication protocol for convenient operation of advanced mobile device functions and services. Results confirm that the our concept is a promising direction for further research and development towards a complete and efficient working environment of touch devices for the blind users.

REFERENCES

[1] Kane, S.K., Wobbrock, J.O., Ladner, R.E. Usable gestures for blind people: Understanding preference and performance. Proc. CHI'11, ACM (2011), 413-422

[2] Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., Gonçalves, D. Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. Proc. ASSETS '11, ACM (2011), 179-186

[3] d'Andrea, F.M. Preferences and Practices Among Students Who Read Braille and Use Assistive Technology. Journal of Visual Impairment & Blindness, Volume 106, Year 2012, October-November 2012, 585-596

[4] Southern, C., Clawson, J., Frey, B., Abowd, G.D., Romero, M. An Evaluation of BrailleTouch: Mobile Touchscreen Text Entry for the Visually Impaired. Proc. MobileHCI'12, ACM (2012), 317-326

[5] Azenkot, S., Wobbrock, J.O., Prasain, S., Ladner, R.E. Input finger detection for nonvisual touch screen text entry in Perkinput. Proc. Graphics Interface, Canadian Information Processing Society (2012), 121-129

[6] Findlater, L., Lee, B.Q., Wobbrock, J.O. Beyond QWERTY: Augmenting Touch-Screen Keyboards with Multi-Touch Gestures for Non-Alphanumeric Input. Proc. CHI'21, ACM (2012), 2679-2682

[7] Costagliola, G., Fuccella, V., Di Capua, M. Text Entry with KeyScretch. Proc. IUI , ACM (2011), 277-286

[8] Ruamviboonsuk, V., Azenkot, S., Ladner, R.E. Tapulator: A Non-Visual Calculator using Natural Prefix-Free Codes. Proc. ASSETS'12, ACM (2012), 221-222

[9] Information about Kajetek SD on the manufacturer's website: http://www.ece.com.pl/index.php?option=com_virtuemart&page=shop.browse&category_id=5&Itemid=2

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 659–662

A Hybrid Approach of System Security for Small and Medium Enterprises: combining different Cryptography techniques

Georgiana Mateescu
Polytechnic University of Bucharest,
Splaiul Independenței 313, Bucharest, Romania,
Email: georgiana.mateescu@gmail.com

Marius Vladescu
Polytechnic University of Bucharest,
Splaiul Independenței 313, Bucharest, Romania,
Email: vladescumariusnicolae@yahoo.com

***Abstract*—Information protection is one of the most important issues in every domain, especially when we are talking about enterprises. Information safety can be translated into three key terms: integrity, availability and data protection. There is a great number of means used in order to achieve the three objectives simultaneously. The most popular is cryptography because it offers a lot of techniques which nowadays are impossible to fail. In this paper we want to prove their efficiency by comparing the different types of crypto algorithms and by presenting their weaknesses and strengths. In order to maximize the benefits of the crypto techniques, we propose a hybrid approach that combines three crypto algorithms.**

I. Introduction

WHEN we are talking about information security we refer to it as the mean we use to protect our information from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction.

The main concepts that a security system has to respect are: confidentiality, integrity, availability and authentication. These concepts represent the information security goals and must be achieved by every security system that aims to be functional. Most security systems use cryptography because it offers various algorithms and techniques practically impossible to break because of their complexity. Cryptography, not only protects data from unauthorized access or alteration, but it can also be used for user authentication. There are three main types of cryptographic algorithms used to accomplish these goals: secret key (or symmetric) cryptography, public-key (or asymmetric) cryptography, and hash functions (Fig. 1).

In this paper, we will analyze these three ciphers: symmetric, asymmetric and hash function. After we present each of them with their strengths and weaknesses we will point out the main attacks that an efficient security system has to face in each case.

To conclude, we propose a hybrid approach of the presented cryptography techniques which combines them for taking benefits from all of their strengths and tries to reduce as much as possible the weakness of one technique with the advantages of the other, in the following manner:

- The original message's message digest is digitally signed (the digital signature uses RSA algorithm)
- Symmetrical cipher is used to code the original message (AES algorithm). The secret key is obtained using a key generator and it is periodically changed.
- The private key used for symmetric cipher is coded using also RSA algorithm, but with different keys.
- The coded private key is attached to the encrypted message together with the digital signature

These techniques will be incrementally introduced and combined into a unitary security system for small and medium enterprises. The purpose of this system is to face the vulnerabilities and threats these enterprises might encounter, by ensuring all the security components in the company's information flow.

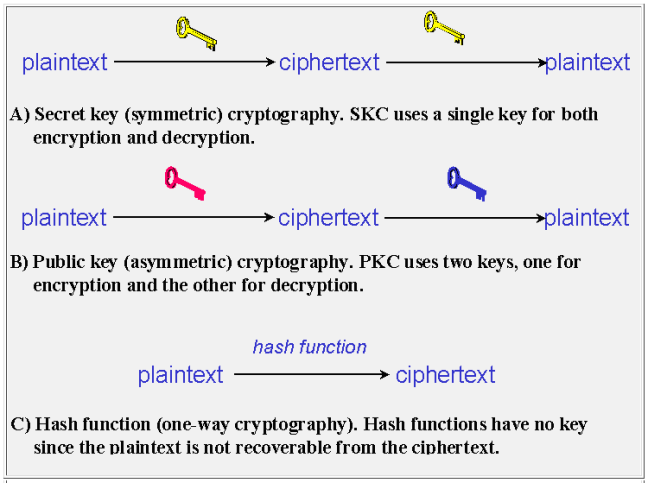


Fig. 1 The main three Cryptography Techniques [1]

II. Theoretical Background

A. Symmetric cryptography

This kind of cryptography uses a single key for both encryption and decryption, and it is also called secret key cryptography (SKC) [1]. This technique works by the following principles:

1. The plaintext is encrypted with the key and the ciphertext is sent to the receiver
2. The receiver uses the same key to decrypt the ciphertext and recover the plaintext.

978-83-60810-53-8/$25.00

The key is a set of rules, and both the sender and the receiver must know the key in order to use the technique.

The most known secret key cryptography schemes are stream ciphers and block ciphers. The stream ciphers generate a sequence of bits used as a key called a keystream, and the encryption is accomplished by combining the keystream with the plaintext. This is usually done with the bitwise XOR operation. The keystream can be independent of the plaintext and ciphertext, in which case the stream cipher is synchronous, or it can depend of the data and its encryption, in which case the stream cipher is self-synchronizing. A block cipher transforms a fixed-length block of plaintext into a block of ciphertext of the same length. The same secret key is used for the decryption by applying the reverse transformation of the ciphertext block [2].

B. Asymmetric (Public Key) Cryptography (PKC)

This technique requires two types of keys: one to encrypt the plaintext and one to decrypt the ciphertext, and it doesn't work without one or another. It is called asymmetric cryptography because it is used a pair of keys: one is the public key that can be advertised by the owner to whoever he wants, and the other one is the private key and it is known only by the owner. The most common public key algorithm is the RSA algorithm, used for key exchange, digital signatures, or encryption of small blocks of data. It uses a variable size key and a variable size encryption block. The security of the RSA algorithm is based on the factorization of very large numbers. Two prime numbers are generated by a special set of rules, and the product of these numbers is a very large number, from which it derives the key-set [3].

C. Hash Functions

A hash function offers a way of creating a fixed-size blocks of data by using entry data with variable length. It is also known as taking the digital fingerprint of the data, and the exit data are known as message digest or one-way encryption. If the data is modified after the hash function was generated, the second value of the hash function of the data will be different. Even the slightest alteration of the data like adding a comma into a text, will create huge differences between the hash values. The hash values solve the problem of the integrity of the messages. MD5 and SHA1 are algorithms for computing a fingerprint of a message or a data file. SHA-1 is described in the ANSI X9.30 standard and produces a 160-bit (20 byte) message digest. It is slower than MD5, but it has a larger digest size, which makes it stronger against brute force attacks. The advantage of MD5 is that it can be implemented faster, due to its 128 bit (16 byte) message digest [4].

III. Threats and Vulnerabilities

The main dangers an enterprise information system faces can be divided into:

- Threats –potential danger to information resources
- Vulnerabilities – weakness in application systems, network, business process or management procedures

The attacks that cryptography based security systems may suffer are dived into:

- Cyphertext-only - attempt to recover plaintext from encrypted text sent in the message.
- Known-plaintext - attempt to discover the key used when the analyst has access to the plaintext of the encrypted message.
- Chosen-plaintext same as Known-plaintext attack, but the analyst gets to choose the known plaintext.

A. Symmetric cryptography (secret key)

Although is the strongest technique that cannot be practically broken if we choose a proper complexity for the secret key, the symmetric crypto algorithms have to face a big threat in order to achieve its benefits: to safely transmit the secret key to the other part of the communication – where the decryption process is made.

Depending of the symmetric type of cipher, the usual attacks are as followed [5]:

- Block Cipher: shortcut attacks and brute force attacks. The shortcut attacks try to minimize the computational complexity required to find the correct key by exploiting the analytical and statistical characteristics of the algorithms. The most used shortcut attacks are differential cryptanalysis. The brute force attacks try one possible encryption key after another to obtain information on the correct key and/or the plaintext (For triple DES, both two-key and three-key triple DES has already been academically broken).
- Stream Cipher: its security depends on the pseudo-random number generator. If the pseudo-random numbers can be efficiently predicted from the past numbers, then the algorithm will be easily broken.

B. Asymmetric cryptography (public key)

The possible attacks on RSA are [6]:

- Searching the message space - if the message space is small, then the attacker could simply try to encrypt every possible message block, until a match is found with one of the ciphertext blocks. In practice this would be an insurmountable task because the block sizes are quite large.
- Guessing *d* - a known ciphertext attack (The attacker knows both the plaintext and ciphertext and he tries to find out the private part from the key)
- Cycle attack – it is the same as "Guessing *d*", but the coded text it is encrypted repeatedly until the original text is obtained. This number of re-cycles will decrypt any ciphertext.
- Low exponent
- Factoring the Public Key – it is considered to be the most efficient attack

C. Hash function (one way cryptography)

In the hash function case, the main vulnerability is the high probability of collisions appearance. Collisions represent the cases when two different inputs, using the same

hash function, generate the same output and therefore can be easily exploited.

In February 2005, Wang, Yin, and Yu [7] published research results which concluded that SHA-1 collisions can be found with the computational complexity equivalent to 2^{69} hash function operations. In addition, Wang, Yao, and Yao claimed that SHA-1 collisions can be found with the computational complexity equivalent to 2^{63} hash function operations.

IV. RELATED WORKS

Nowadays, small and medium enterprises use different techniques in order to achieve information protection. Some of them are based on cryptography [8, 9], others on PKI [10, 11]. All these architectures ensure a certain level of security that could be sometimes too small for the threats and the vulnerabilities that they have to face.

We propose a hybrid approach that wants to offer a complex solution with the following characteristics:

- Unified system – all the crypto techniques are combined to solve each other's threats and weaknesses
- Structured system – encapsulating different types of ciphers to maximize the efficiency

V. HYBRID ENCRYPTION SYSTEM

Data encryption is an important element of an organization's response to security threats and regulatory mandates. The enterprises are facing the fact that while encryption is not difficult to achieve, managing the associated encryption keys across their lifecycle quickly becomes a problem that creates a new set of security vulnerabilities and risks. The administration of keys must itself have built-in protection against internal maliciousness.

Encryption resources such as keys, hash algorithms, certificates, and digital signatures are dynamic and fluid. They must be changed, cycled, or renewed regularly. Furthermore, they must be archived under time-based management so that they would be available for retrieving.

By combining different cryptography techniques, this approach offers a solution for various weaknesses that must be faced in a security crypto system including:

- Key encryption management: key generator, key storage, key transmission
- Computing time
- Ensure all the security goals: integrity, availability, authentication and confidentiality

This crypto security system ensures (Fig. 2):

- Data integrity – using hash function
- Authentication and authenticity – using digital signature (DSA – asymmetric cryptography)
- Data confidentiality – using AES (Advanced Encryption Standard – symmetric cryptography algorithm)

A. Digital signature

For digital signature we can use RSA algorithm or DSA algorithm. In RSA algorithm, the message digest (the

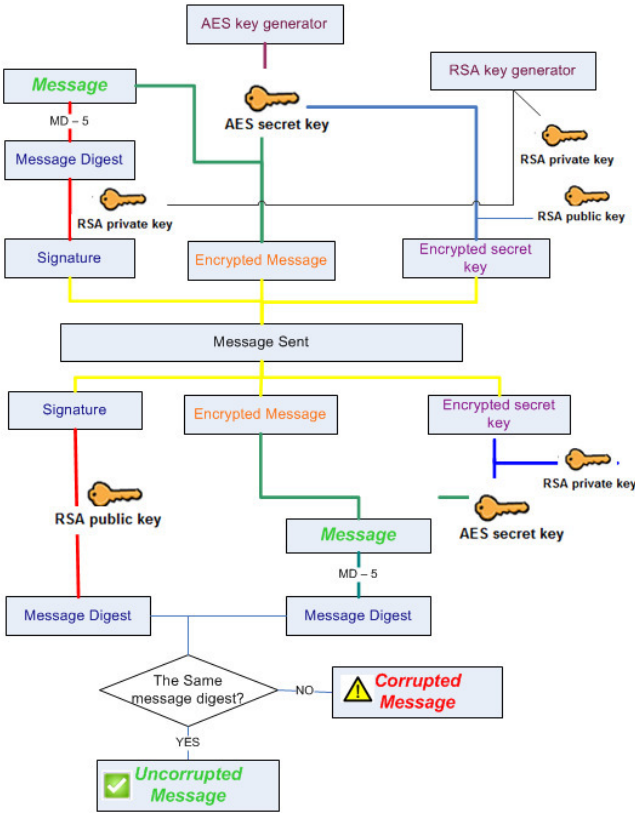


Fig. 2 Encryption schema

message hash) is encrypted with the RSA private key. This encryption represents the signature and it is attached to the message. It is obvious that this approach is impractical because:

- The ciphertext signature is the same size as the corresponding plaintext, so the messages sizes are doubled, consuming large amounts of bandwidth and storage space.
- Public key encryption is slow and places heavy computational loads on computer processors, so network and computer performance can be significantly degraded.
- Encrypting the entire contents of information produces large amounts of ciphertext, which can be used for cryptanalysis attacks, especially plaintext attacks (where certain parts of the encrypted data, such as e-mail headers, are known beforehand to the attacker).

National Security Agency developed DSS (Digital Signature Security Standard) which defines Digital Signature Algorithm. This algorithm is similar to RSA, but it does not encrypt message digests with the private key/ decrypt the message digest with the public key. Instead, DSA uses mathematical functions to generate a digital signature composed of two 160-bit numbers that are derived from the message digest and the private key. DSA uses the public key to verify the signature, but the verification process is more complex than RSA. DSA requires the use of the SHA-1 message digest function to ensure strong digital signatures and because of that and of the verification process, RSA digital signature

process generally provides better overall performance. Beside that DSA it never used to encrypt the message (for example you cannot use DSA to transmit the secret key of a symmetric cryptography algorithm).

Taking in count all these factors we will use RSA digital signature together with a RSA key generator – an algorithm used to generate secure RSA private and public keys.

B. Hash function

We use MD5 as a hash function hash function because it is faster than SHA – 1. Although it is weaker than SHA -1, we consider that the computing time is an important aspect which has to be optimised as much as possible.

C. Rijndael AES – symmetric encryption

AES algorithm has the following steps:

- Key generation
- Message encryption/decryption
- Key transmission: key encryption/decryption

Key generation is made using a cryptographically secure pseudo random number generator. We chose PRNG1 core because it is secure in wireless communications, RFID, Smart cards, electronic financial transactions. The key management system also includes the necessity of secure servers used for the key archiving and storage under time-based management so that historic data availability is ensured.

Message encryption is made using the generated key and Rijndael AES (Advanced Encryption Standard) algorithm. We chose this technique because, according to NIST, it has better security and efficiency characteristics than DES. Rijndael was designed based on the following three criteria [13]:

- Resistance against all known attacks;
- Speed and code compactness on a wide range of platforms;
- Design simplicity

Message decryption is made by the receiver using the same secret key as the sender.

Key transmission includes:

- Secret key encryption using RSA private key (generated by the key generator)
- Attach to the sent message the encrypted secret key

For the secret key transmission, we chose the most usually algorithm RSA because the practise has proved the fact that this technique successfully faces all the threats.

VI. Conclusions

Today's business environment is compliance-driven, competitive and increasingly fraught with from financially motivated hackers and frustrated employees. This creates a mounting demand for effective, practical, automated and risk-mitigating ways to manage keys throughout their lifecycle, so that only authorized users are granted access and the unauthorized user are thwarted. User and application access to these resources must be controlled, managed and audited so that authorized access is quick and reliable, all while preventing malicious attacks.

A strong crypto system together with a secure Key encryption management system can ensure all security goals. The combination of different cryptography algorithms provide a maximized efficiency, correcting or compensating each other's weaknesses.

VII. References

[1] Gary C. Kessler, *An overview of Cryptography*, 28 April 2013 http://www.garykessler.net/library/crypto.html

[2] RSA Laboratories- Chryptographic tools; section 2.1.5. unpublished; http://www.rsa.com/rsalabs/node.asp?id=2174

[3] Ing. Cristian MARINESCU, prof.dr.ing. Nicolae ȚĂPUȘ ; "An Overview of the Attack Methods Directed Against the RSA Algorithm"; Revista Informatica Economica, nr. 2(30)/2004

[4] Arash Partow –"General Purpose hash Function Algorithms"

[5] Masashi Une and Masayuki Kanda, "Year 2010 Issues on Cryptographic Algorithms", Discussion Paper No. 2006-E-8, IMES, C.P.O BOX 203 Tokyo, 100-8630 Japan

[6] Prof. Patrick McDaniel, Network and Security Research Center Department of Computer Science and Engineering Pennsylvania State University, University Park PA – "Public-Key Cryptography and Attacks on RSA", 2010

[7] X. Wang, and B. de Weger, "Colliding X.509 Certificates," Cryptology ePrint Archive, 2005 (available at http://eprint.iacr.org/2005/067).

[8] Rodrigues, J. Roberts (2007). "System security and personal help data protection"

[9] Gregory Braun –" Crypto 2000" (*For Small to Medium Businesses*)

[10] Information Technology and Organizations: "Trends, Issues, Challenges and Solutions", VOLUME 1, 2003 Information Resources Management Association, International Conference,Philadelphia, Pennsylvania, USA,May 18-21, 2003

[11] Ki Woong Park, Hyun Jin Choi, and Kyu Ho Park-"An Interoperable Authentication System using ZigBee-enabled Tiny Portable Device and PKI", Internation Conference on Next Generation PC

[12] http://technet.microsoft.com/en-us/library/cc962021.aspx

[13] Daemen, Joan; Rijmen, Vincent. "AES Proposal: Rijndael" Document version 2, 1999

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 663–666

Impact of Signalling Load on Response Times for Signalling over IMS Core

Lubos Nagy, Jiri Hosek, Pavel Vajsar and Vit Novotny
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technicka 12, 616 00 Brno, Czech Republic
Email: lubos.nagy@phd.feec.vutbr.cz

***Abstract*—This article focuses on the performance evaluation of the response time for signalling through a home Internet Protocol based Multimedia Subsystem (IMS), separately for each of IMS core nodes (Proxy-Call Session Control Function, Interrogating-CSCF, Serving-CSCF and Home Subscriber Server) and then on the investigation of the trend-line functions and their equations to describe these delays for various measured intensity of signalling generated load by high-performance tool – IxLoad. In this article, we have found out the trend-line function of response times for each measured message. Thanks to the showed results, some performance parameters like delay in selected IMS core node and their behaviour can be predicted and evaluated.**

***Keywords*—DIAMETER, IP based Multimedia Subsystem, IxLoad, Response time, SIP.**

I. Introduction

THE current trends in telecommunications lead to the network convergence and to effort the greatest number of telecommunication services through one type of transport networks and for one multifunction terminals. Therefore in the past, the operators and vendors were looking for an IP-based connectivity concept allowing the convergence network technologies and opportunities for optimization at all levels of designed communication system. Nowadays, the IMS (IP Multimedia Subsystem) in role of the IP-based service control architecture represents the standard of fixed-mobile network convergence. However to this optimization, it is necessary to know exactly the behaviour of these systems for various real conditions. One of the possible ways how to determine the behaviour of the whole IMS subsystem or only some IMS nodes is the performance analysis either based on the mathematical modelling using queueing theory or performance benchmarking (see *Section II* of this article).

The methodology of IMS/NGN (Next Generation Networks) performance benchmarking is standardized by European Telecommunications Standards Institute (ETSI) in multipart deliverable that is divided into four separate parts [1]: *Core Concepts*, *Subsystem Configurations and Benchmarks*, *Traffic Sets and Traffic Profiles*, and *Reference Load network quality parameters*. The overall concept of IMS test-beds including the IMS benchmark information model, test parameters and benchmark metrics examples is defined in the first part of this technical standard. The SUT (*System Under Test*) configuration parameters, use-cases and scenarios with metrics and design objectives are presented in the second specification. In the third part, the traffic set, traffic-time profile and test procedures are defined. The reference load network quality parameters for use-cases defined in the second part of [1] are presented in the last part of this specification.

II. Related Work

There are various research papers, documents or studies that describe the performance analysis of whole IMS. The OpenIMS Core project (in role of SUT) and IMS Bench SIPp project (in role of TS - *Test System*) are often used tools for the performance analysis of IMS subsystem. The related work concerning the performance evaluation of IMS networks can be divided into three main categories, the performance evaluation of maximum load [2]–[4], the performance evaluation of delays of SIP (Session Initiation Protocol) signalling [5] and evaluation of delays of IMS procedures like IMS registration or IMS session setup procedures [6]–[7].

In our previous works, we were mainly focused on the performance analysis using the IMS queuing network model [8] and on the performance evaluation of maximum load signalling over our laboratory IMS network [9] according to specification [1]. In [9], we investigated that the value of the maximum signalling load for these hardware and software configurations is $500cps$ for defined IHS threshold 0.025% and the HSS entity was the failure point of simulated IMS network. The same bottleneck was described in [3] for even lower values of load (during execution of registration procedures). The similar test-beds are described by others researchers in [2]–[4] and the results of maximum loads correspond to the results measured in our test-bed which is described in [9] and in this section. In [8], we presented the design of IMS mathematical model based on separated M/M/1 queuing system with feedbacks that consists of the same IMS entities, signalling and services as our laboratory IMS network. In this M/M/1 model, the new load balancing methods, that can be used for a selection of S-CSCF server during the registration procedures of subscribers, were designed and evaluated. The obtained results showed that the service latency of the whole IMS core subsystem can be optimized with the help of implemented methods into mathematical network model based on M/M/1 queuing system. However, the service times are not exponentially distributed in the real networks. Therefore, the *main motivation of this article* is targeted at the measurement of delays for each SIP and DIAMETER signalling of IMS core elements using the performance analysis of IMS core elements and standards [1]. Thanks to the obtained results, we will be able to simulate the behaviour of IMS nodes using M/G/1 queuing systems and

978-83-60810-53-8/$25.00

evaluate the designed methods for load balancing under more realistic network conditions.

III. IMS TEST-BED

The experimental topology of the test-bed (see Fig. 1) consisted of IMS core subsystem, VoD Application Server and Media Streaming Server with the same following hardware and software configurations: 4x Intel Core i5-2400S CPU @2.50 GHz with 6144k L3 cache, 8 GM RAM (DIMM 1333 MHz), 82574L Intel Gigabit Network Connection, OS GNU/Linux (Debian distribution, AMD64 architecture, kernel v3.2), the software implementations of CSCF nodes are based on the SER (SIP Express Router) and the HSS based on FHoSS (FOKUS HSS) server created by Fraunhofer FOKUS Institute.

The SIP load signalling of selected multimedia services (Video on Demand, Voice over IP and File transfer) with defined intensity (see λ in Fig. 1) of the Poisson arrival process is generated with the help of the high-performance IxLoad application (see TS in Fig. 1). Each of services consists of three phases: registration procedure with subscription , session establishment and termination procedures (only for registered subscribers), and de-registration procedure. The registration phase consists of the registrar and subscription transactions. The generated signalling flows are created with the help of the standardized document 3GPP TS 24.228.

We can define the test-bed architecture with the help of queuing theory that is often used to evaluate the performance parameters of whole networks or only some network nodes. In the Fig. 1, the IMS core nodes are shown as the M/G/1 queuing system. One of the most important performance parameters is the response time of system (see eq. (1)). The mean value of this parameter can be calculated using the Pollaczek-Khinchine formula (known as P–K mean value formula [10]):

$$\frac{T}{\bar{x}} = 1 + \rho * \frac{1 + C_b^2}{2 * (1 - \rho)} \tag{1}$$

Where the $\rho = \frac{\lambda}{\mu}$, C_b^2 is the coefficient of service time variation and the parameter $\bar{x}$ is the mean service time.

The way to calculate these values is following. In the case of P-CSCF server, the response time is always the time difference between the received and forwarded SIP messages. The response time of SIP signalling through S-CSCF server equals to the signalling through P-CSCF except the registration and de-registration procedures. In the case of these procedures, the response times of signalling through S-CSCF server are determined as the transactions between SIP and DIAMETER (DIAMETER uses TCP as its transport protocol). It means that the service times are determined as differences between the SIP request received from I-CSCF and DIAMETER request sent to HSS (sending time of DIAMETER MAR - receiving time of the first SIP REGISTER) or the DIAMETER answer received from HSS and SIP response sent to I-CSCF (sending time of SIP 401 - receiving time of DIAMETER MAA). The same way to determine the signalling response times is used for I-CSCF server. In the case of times for DIAMETER signalling through HSS database, the value is calculated as difference between the times of received DIAMETER request and sent DIAMETER answer.

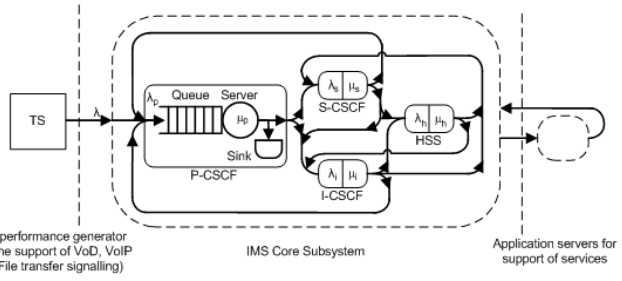


Fig. 1. The test-bed architecture as the queuing system network with feed-backs.

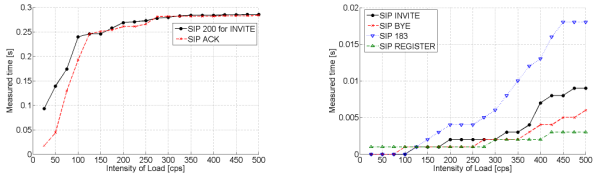
IV. RESULTS AND ANALYSIS

The traffic of three advanced telecommunication services (Video on Demand, Voice over IP and File transfer using SIP/RTSP/MSRP and RTP/RTCP signalling) over IMS experimental network is evaluated for various load intensities separately (from $25cps$ to $500cps$) for each of IMS core nodes (the P-CSCF, I-CSCF, S-CSCF and HSS) and for each of SIP or DIAMETER messages. The most important results are shown in Fig. 2 to Fig. 3(b) and Tab. I to Tab. IV. The maximum value of signalling load ($500cps$) for used hardware and software configurations was investigated in [9]. The SUT (whole IMS network) was very unstable for the load greater than $500cps$. In Tab. I to Tab. IV, the trend-lines of response times for each message through selected IMS servers are shown only for messages with the measured service time greater than $1ms$. This limitation, shown in all tables and figures, is the measurement accuracy. The measured messages are displayed in the first column of the shown tables, the formulas of trend-lines of the response time for defined range of signalling load are shown in the second column. The parameter x is the signalling load generated by IxLoad application in role of TS (see λ in Fig. 1). The delay calculation methodology has been described in the previous section of this article. The measured characteristics of the response times into arrival signalling load through SUT (see λ in Fig. 1) are shown in the Fig. 2–4.

The P-CSCF trend-lines of response times (see Tab. I and Fig. 2) are mostly defined with the exponential or logarithmic time complexity. Other measured SIP request or response times (SIP 180, SIP 200s for REGISTER, BYE, SUBSCRIBE, UPDATE and PRACK, SIP 401, PRACK and UPDATE) are set to $1ms$ (the measured times are less than $1ms$). In the graphs (see Fig. 2), the mean values of measured times within the generated load (see λ in Fig.1) for SIP messages with response times greater than $1ms$ are shown. From these graphs, it can be seen that the ACK and 200 for INVITE messages have the

TABLE I. THE FUNCTIONS OF RESPONSE TIMES FOR EACH MESSAGE THROUGH THE P-CSCF

SIP requests and responses	The trend-line function
SIP Session in Progress	$f(x) = (0.0009) * exp(0.0067 * x)$
SIP OK for INVITE	$f(x) = (-2.346) * x^{(-0.7343)} + 0.3137$
SIP REGISTER	$f(x) = (5.171e-11) * x^{(2.838)} + 0.0009$
SIP INVITE	$f(x) = (0.0003) * exp(0.0075 * x)$
SIP ACK	$f(x) = (-3.804) * x^{(-0.7923)} + 0.3139$
SIP BYE	$f(x) = (0.0004) * exp(0.0055 * x)$



(a) The messages with higher measured times. (b) The messages with lower measured times.

Fig. 2. The mean values of response times vs. signalling load through the P-CSCF.

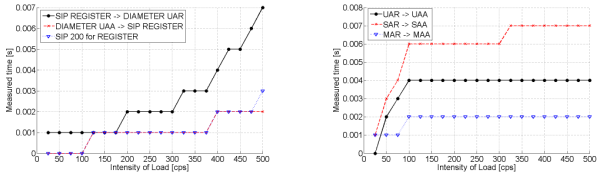
highest response times. However, the SIP Session in Progress and SIP INVITE messages have the greatest increase of the response time within the analysed interval.

The I-CSCF server is the next evaluated IMS node (see Tab. II and Fig. 3(a)). In our test-bed of a home IMS network, this server is active only during the registration or de-registration procedures. Only the SIP 401 response has different time complexity (it has the constant complexity, the measured times are less than $1ms$) than other SIP messages. The SIP REGISTER $\rightarrow$ DIAMETER UAR processing has the shortest rise of measured response times (see Fig. 3(a)).

The last node, which was evaluated within the CSCF core, is the S-CSCF server (see Tab. III). This IMS node presents the central node of the whole IMS network and therefore it can be expected that this node has the greatest response times (see Fig. 4). Actually, there are two interesting facts in this obtained results. First, the highest values of response time are associated with the SIP responses (SIP 200 for REGISTER or SIP 401) created by S-CSCF node during the registration or de-registration procedures when the DIAMETER answers (MAA or SAA) are received from the HSS node. The second interesting result is that the SIP ACK and SIP 200 for INVITE

TABLE II. THE FUNCTIONS OF SERVICE TIMES FOR MESSAGES THROUGH THE I-CSCF

SIP and DIAMETER requests and responses	The trend-line function
SIP 200 for REGISTER	$f(x) = \begin{cases} < 1ms & \text{if } x < 100cps \\ 1ms & \text{if } x \in \langle 100, 375 \rangle cps \\ 2ms & \text{if } x > 375cps \end{cases}$
SIP REGISTER $\rightarrow$ DIAMETER UAR	$f(x) = (-6.146e-11) * x^{(2.956)} + 0.0009$
DIAMETER UAA $\rightarrow$ SIP REGISTER	$f(x) = \begin{cases} < 1ms & \text{if } x < 125cps \\ 1ms & \text{if } x \in \langle 125, 375 \rangle cps \\ 2ms & \text{if } x > 375cps \end{cases}$

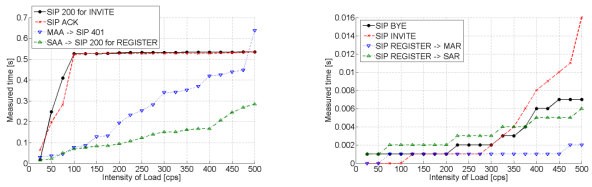


(a) The response times of messages through the I-CSCF. (b) The response times of messages through the HSS.

Fig. 3. The mean values of response times vs. signalling load through the I-CSCF (on the left) and though the HSS (on the right).

TABLE III. THE FUNCTIONS OF RESPONSE TIMES FOR MESSAGES THROUGH THE S-CSCF

SIP and DIAMETER requests and responses	The trend-line function
SIP REGISTER $\rightarrow$ DIAMETER MAR	$f(x) = \begin{cases} < 1ms & \text{if } x < 75cps \\ 1ms & \text{if } x \in \langle 75, 450 \rangle cps \\ 2ms & \text{if } x > 450cps \end{cases}$
DIAMETER MAA $\rightarrow$ SIP 401	$f(x) = (0.0016) * x^{(0.936)} - 0.0341$
SIP REGISTER $\rightarrow$ DIAMETER SAR	$f(x) = (0.0013) * exp(0.0032 * x)$
DIAMETER SAA $\rightarrow$ SIP 200 for REGISTER	$f(x) = (0.0042) * exp(0.0038 * x)$
SIP INVITE	$f(x) = (4.528e-12) * x^{(3.528)}$
SIP ACK	$f(x) = (-115.9) * x^{(-1.711)} - 0.5348$
SIP BYE	$f(x) = (-1.145e-10) * x^{(2.885)} - 0.0009$
SIP 200 for INVITE	$f(x) = (-277.9) * x^{(-1.952)} + 0.536$



(a) The messages with higher measured times. (b) The messages with lower measured times.

Fig. 4. The mean values of response times vs. signalling load through the S-CSCF.

messages have the highest response time of all SIP messages forwarded by this server. The similar result was measured also in the case of the forwarding this SIP message by P-CSCF node. The measured values of other request or response times (for SIP 180, 183, SIP 200s for BYE, UPDATE and PRACK, SIP PRACK and UPDATE) are not showed in Tab. III because the measured values are less than $1ms$.

The last measured node of IMS network is the HSS database (see Tab. IV or Fig. 3(b)). In our test-bed, the database server is active only during registration and de-registration procedures. It can be seen that the response time of all measured DIAMETER requests/answers has the logarithmic time complexity with relatively low difference between the value of minimum load and the value of maximum load.

In our case, the tested IMS core subsystem is in the role of a home IMS network. The signalling goes through each of evaluated IMS core server (see Fig.1) only the case of the de/registration procedures. The delay of IMS core, which is calculated as formula (2), consist of IMS core element delays and transport delay through network infrastructure.

$$D = \sum D_P + \sum D_I + \sum D_H + \sum D_S + \sum D_T \quad (2)$$

TABLE IV. THE FUNCTIONS OF RESPONSE TIMES FOR MESSAGES THROUGH THE HSS

DIAMETER Command-Codes	The trend-line function
300 (UA$\{R, A\}$)	$f(x) = (-1.293) * x^{(-1.792)} + 0.004$
301 (SA$\{R, A\}$)	$f(x) = (-0.046) * x^{(-0.569)} + 0.008$
303 (MA$\{R, A\}$)	$f(x) = \begin{cases} 1ms & \text{if } x < 100cps \\ 2ms & \text{if } x \geq 100cps \end{cases}$

Where $\sum D_P$, $\sum D_I$, $\sum D_S$ and $\sum D_H$ are the investigated times that the messages spent in CSCFs and HSS and $\sum D_T$ is the time the messages spend within the network infrastructure. Each of core node delays is composited from the queueing and processing delays defined in [7]. The values of $\sum D_I$ and $\sum D_H$ are greater than zero if the signalling is from the registration or de-registration procedures, else the values are equal to zero. The theorem is valid for the home IMS network simulated in this paper.

The successful registration procedure (see eq. (3)) is influenced by three delay parts, thereof two delays are influenced by time the signalling spent in SUT (the tested IMS core) and the response times of TS (IxLoad application).

$$D_{REG} = \underbrace{\sum D_{(REG1\rightarrow 401)}}_{\text{SUT}} + \underbrace{\sum D_{(401\rightarrow REG2)}}_{\text{TS}} + \\ + \underbrace{\sum D_{(REG2\rightarrow 200)}}_{\text{SUT}} \tag{3}$$

We do not tie the effect of $\sum D_{(401\rightarrow REG2)}$ and transmission delay in the following equations. The first of SUT delay is shown in eq. (4). We can define the second one based on assumptions from the first SUT delay.

$$\begin{aligned} \sum D_{(REG1\rightarrow 401)} &= \underbrace{D_{(REG1\rightarrow REG1)} + D_{(401\rightarrow 401)}}_{\text{P}} + \\ &+ \underbrace{D_{(REG1\rightarrow UAR)} + D_{(UAA\rightarrow REG1)} + D_{(401\rightarrow 401)}}_{\text{I}} + \\ &+ \underbrace{D_{(UAR\rightarrow UAA)} + D_{(MAR\rightarrow MAA)}}_{\text{H}} + \\ &+ \underbrace{D_{(REG1\rightarrow MAR)} + D_{(MAA\rightarrow 401)}}_{\text{S}} \end{aligned} \tag{4}$$

If we neglect the effects of lower signalling delays and the impact of delays outside IMS core elements (see eq. (2)–(4)) then we can define for conditions of our test-bed the delay of successful registration procedures as:

$$D_{REG} \approx \underbrace{D_{(MAA\rightarrow 401)}}_{\text{from first SUT delay}} + \underbrace{D_{(SAA\rightarrow 200)}}_{\text{from second SUT delay}} \tag{5}$$

The percentage ratio of derived D_{REG} (see eq. (5)) is 94.4% of measured D_{REG} (see Fig. 2–Fig. 4). From the characteristics and eq. (5), it can be seen that the delay of IMS procedures is mainly influenced by measured delays of S-CSCF server that is in role of IMS networks as IMS central core element.

V. Conclusion

This paper deals with the evaluation of response times for signalling through the experimental IMS core subsystem, separately for each IMS core node and message, and for various values of network load. Three advanced telecommunication services were generated by the high-performance IxLoad application. All selected IMS core nodes were situated in the servers with the same hardware configurations.

From the showed characteristics (see Fig. 2–Fig. 4), it can be seen that the central node of the whole IMS network (the S-CSCF server) has the highest values of response time and its influence on delays of signalling through whole home IMS network from eq. (5). Based on assumption from eq. (2)–(5), we can obtain the similar results for other tested IMS procedures like session establishment and that the S-CSCF server has the highest impact on delay of signalling within a home IMS network. However, the influence of S-CSCF server is not very high in the case of session termination procedure. Therefore, our future work will focus on the problem how to optimize the latency of the whole IMS network e.g. during registration procedures using load-balancing of S-CSCF servers.

Also, we have found out the trend-lines with the correlation (the lowest R-squared index was approximately 0.95, the most commonly value of R-squared was 0.98) that are described by the help of the exponential or logarithmic functions for each evaluated message and IMS core node. In the case of HSS node, only logarithmic function is used to define trend-lines of DIAMETER signalling. This functions could be used to predict the delay either in the node of IMS network or within the whole IMS network.

Acknowledgment

This research work is funded by projects SIX CZ.1.05/2.1.00/03.0072, CZ.1.07/2.3.00/30.0005, EU ECOP EE.2.3.20.0094 and CZ.1.07/2.2.00/28.0062.

References

[1] *ETSI European Telecommunications Standards Institute*, IMS Network Testing (INT); IMS/NGN Performance Benchmark. ETSI TS 186 008. November 2012.

[2] G. Din, R. Petre, I. Schieferdecker, "A Workload Model for Benchmarking IMS Core Networks," in *IEEE Global Telecommunications Conference GLOBECOM 2007*, 26.–30. Nov. 2007, pp. 2623–2627. ISBN: 978-1-4244-1043-9.

[3] R. Herpertz, J. M. E. Carlin, "A Performance Benchmark of a Multimedia Service Delivery Framework," in *IEEE Mexican International Conference on Computer Science ENC 2009*, 21.–25. Sept. 2009, pp. 137–141. ISBN: 978-1-4244-5258-3.

[4] D. Thissen, J. M. E. Carlin, R. Herpertz, "Evaluating the Performance of an IMS/NGN Deployment," in *Proceedings of the 2nd Workshop on Services, Platforms, Innovations and Research for new Infrastructures in Telecommunications. SPIRIT 2009*, Germany, Oct. 2009.

[5] S. Pandey, V. Jain, D. Das, V. Planat, R. Periannan, "Performance Study of IMS Signaling Plane," in *IEEE International Conference on IP Multimedia Subsystem Architecture and Applications - 2007*, 2007. pp. 1–5. ISBN: 978-1-4244-2671-3.

[6] Y. He, J. Veerkamp, A. Bilgic, A. Bilgic, "Analyzing the Internal Processing of IMS-based and traditional VoIP systems" in *IEEE International Conference on Telecommunications: The Infrastructure for the 21st Century (WTC), 2010*, Sept. 2010, pp. 1–6. ISBN: 978-3-8007-3303-3.

[7] A. Munir, A. Gordon-Ross, "SIP-Based IMS Signaling Analysis for WiMax-3G Interworking Architectures" in *IEEE Transactions on Mobile Computing*, May 2010. Volume 9, Issue 5. pp. 733-750. ISSN: 1536-1233.

[8] L. Nagy, J. Tombal, V. Novotny, "Proposal of a Queueing Model for Simulation of Advanced Telecommunication Services over IMS Architecture," in *IEEE International Conference on Telecommunications and Signal Processing - 2013*, 2013, ISBN: 978-1-4799-0402-0.

[9] L. Nagy, R. Krkos, V. Novotny, "Performance Analysis of IMS Network," in *the 14th International Conference on Research in Telecommunication Technologies - RTT 2012*, 2012, pp. 55.–60. ISBN: 978-80-554-0570-4.

[10] L. Kleinrock, "Queuing Systems, Vol. 1: Theory"Ed. New York: Wiley Interscience, 1975. 417 pages. ISBN: 0-471-49110-1.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 667–670

Creating a Serial Driver Chip for Commanding Robotic Arms

Roland Szabó, Aurel Gontean
Applied Electronics Department
Faculty of Electronics and
Telecommunications,
"Politehnica" Univ. of Timişoara
Timişoara, România
roland.szabo@etc.upt.ro

***Abstract*—In this paper we shall present a serial driver chip creation on FPGA. We created this serial driver chip for the RS-232 interface and we programmed it to 115200 baud rate to be able to communicate with the Lynxmotion AL5 type robotic arms. This serial driver chip was made for bidirectional communication to be able to send and receive SCPI (Standard Commands for Programmable Instruments) commands on the serial interface. If we create the layout of this chip we can create our own ASIC (Application-Specific Integrated Circuit) and this way we shall have a standalone chip which can control a robotic arm.**

I. Introduction

THIS paper presents the creation of the serial driver chip. Serial driver chip is very useful in controlling many measuring equipments, power supplies, industrial test systems and even robotic arms. No too many types of equipment have USB interface or if they have they have only USB connector, but they still emulate the COM port. This means that the serial interface is still widely used in the industry, even though it's speed it's not the highest. Basically we had to create the UART diver chip ad after it to connect whatever connector to we wish to connect DB9 or DB25, depending on what kind of robotic arm do we wish to control. In our case the Pmod (Peripheral Module) connected to our FPGA board had DB9 connector which was perfect for the Lynxmotion AL5 type robotic arm, but for the Scorbot ER-III robotic arm we needed DB25 connector which we solved with a serial converter from DB9 to DB25.

We can control robotic arms with a PC, but if we have an embedded solution is always better, because it's more portable, and consumes less power. If we find an embedded solution that is implemented in hardware, it's even better, because we shall have no issues regarding to propagation times and we don't' have problems likes bug in software or issues like the system can freeze.

II. Problem Formulation

We had a Lynxmotion AL5 type robotic arm (Fig. 1) and a Scorbot ER-III robotic arm (Fig. 2).

We had also more FPGA boards one NEXYS 2 board (Fig. 3) with Spartan-3E FPGA and one ATLYS board (Fig. 4) with Spartan-6 FPGA + and Pmod RS232 (Fig 5).

Fig. 1 Lynxmotion AL5A robotic arm

Fig. 2 Scorbot ER-III robotic arm

Fig. 3 NEXYS 2 FPGA board with Spartan-3E FPGA

Fig. 4 ATLYS FPGA board with Spartan-6 FPGA

Fig. 5 Pmod RS232 for the ATLYS board

978-83-60810-53-8/$25.00

We had all the hardware, but we needed to create the serial driver to make the communication and to control the robotic arms. For the ATLYS FPGA board we needed to use the Pmod RS232, because the board has only micro USB connector to its UART interface and that not the best solution, because it needs software driver to emulate the serial port on USB and the connector is not suitable for us too, because we need not micro USB, but DB9 or DB25. The USB UART is Ok just when the PC is the master and the FPGA board is the slave, because on the PC we can easily install the USB to serial converter driver. In our case the FPGA board is the master and the robotic arm's servo controller is the slave, this way is really hard to impellent a software driver for the serial to USB converter chip. In our case the standard serial port with DB9 connector is the best solution, this way we had to use a Pmod for the ATLYS board.

III. Problem Solution

A. Theoretical Background

The SCPI (Standard Commands for Programmable Instruments) for controlling the robotic arm will be presented next, these commands are sent on the serial driver chip and with these commands the robotic arm is moved.

```
// SSC-32 VERSION
\r\rVER\r

// INITALIZE MOTORS
QPL0\rQP0\r
QP1\r
QP2\r
//...
QP31\r

// ALL SERVOS 1500
#0P1500S0\r#1P1500S0\r#2P1500S0\r#3P150
0S0\r#4P1500S0\r#5P1500S0\r

// GRIPPER
#4P1500S1000\r

// WRIST ROTATE
#5P1500S1000\r

// WRITST
#3P1500S1000\r

// ELBOW
#2P1500S1000\r

// SHOULDER
#1P1500S1000\r

// BASE
#0P1500S1000\r
```

Somehow from these commands we managed to create some formulas too, to know the correspondence between the angles and robotic commands.

To know exactly the angles we can simply calculate with equation (1).

$$\alpha = \frac{\Delta\omega}{180° - 0°} = \frac{2500 - 500}{180° - 0°} = 11, (1)\ robotic\ values \tag{1}$$

This means the following shown in equation (2).

$$1°{\sim}11,(1)\ robotic\ values \tag{2}$$

The block diagram of the experimental setup with the ATLYS board and the Pmod RS232 is shown on Fig. 6.

The block diagram of the experimental setup with the NEXYS 2 board is shown on Fig. 7.

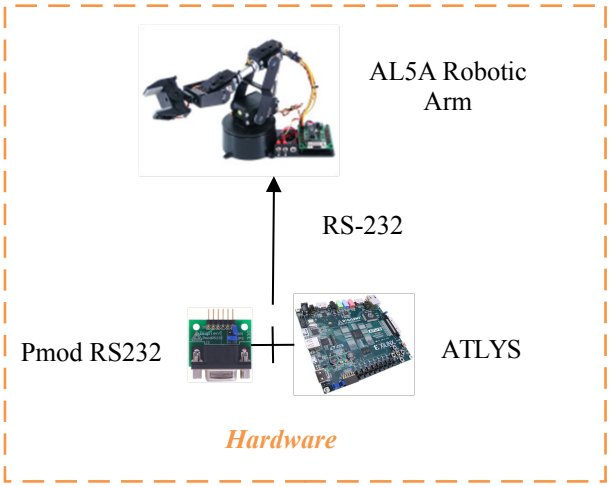


Fig. 6 The block diagram of the experimental setup with the ATLYS board and the Pmod RS232

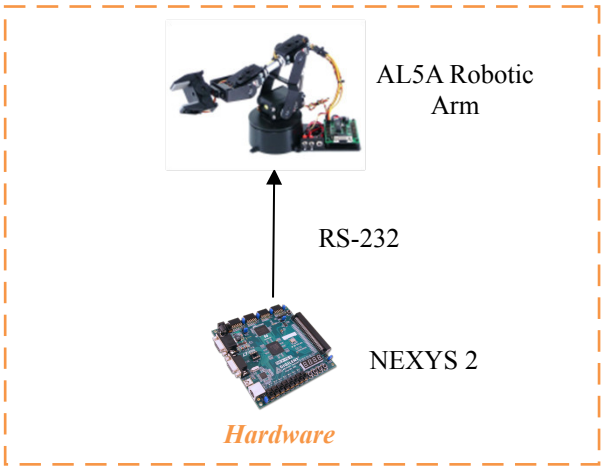


Fig. 7 The block diagram of the experimental setup with the NEXYS 2 board

B. Circuit Diagrams

These circuit diagrams were created from VHDL code in Xilinx ISE. These circuits are what we have inside the serial driver chip.

On Fig. 8 we can see the bidirectional serial driver chip, which has both the transmit (TXD) and receive (RXD) ports.

On Fig. 9 we can see the UART chip which we included in the bidirectional serial driver chip, this chip is actually the serial protocol.

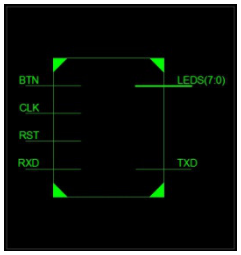


Fig. 8 The bidirectional serial driver chip structure with transmit (TXD) and receive (RXD) functions

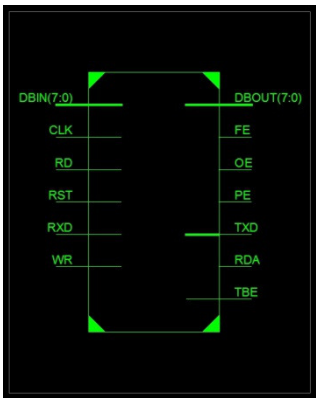


Fig. 9 The UART chip which is included in the bidirectional serial driver chip

On Fig. 10 we can see the circuit structure inside the bidirectional serial drive chip.

On Fig. 11 we can see the circuit structure inside the UART chip.

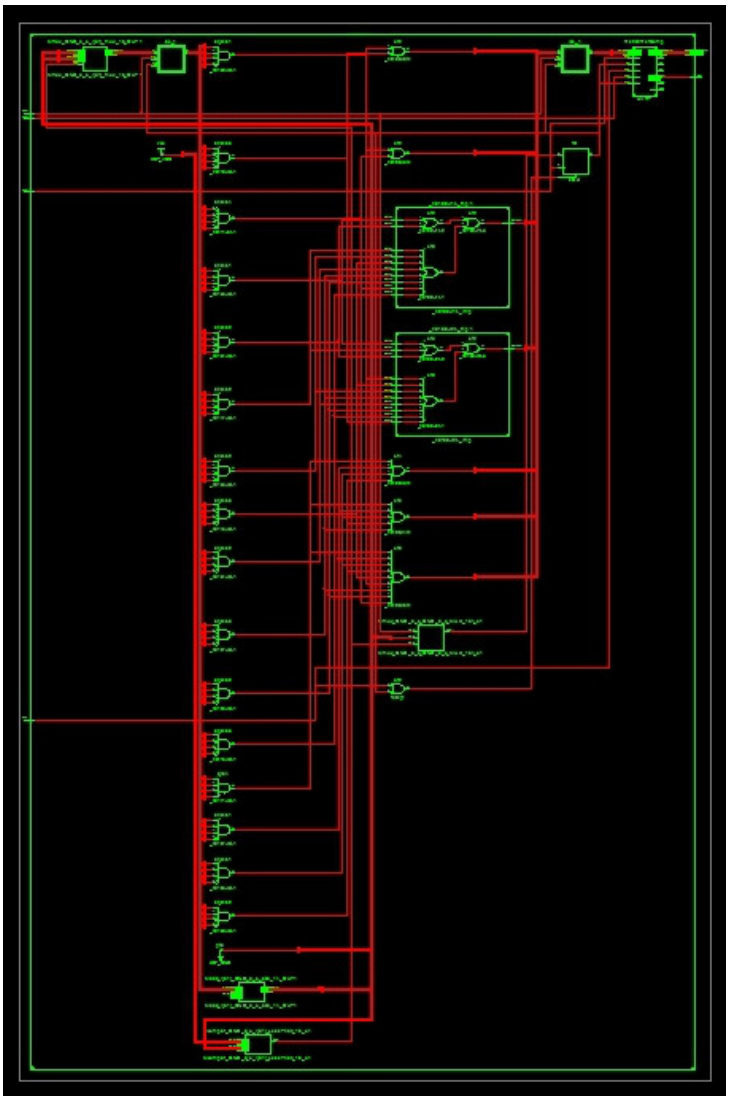


Fig. 10 The bidirectional serial driver chip's circuit (inside the bidirectional serial driver chip)

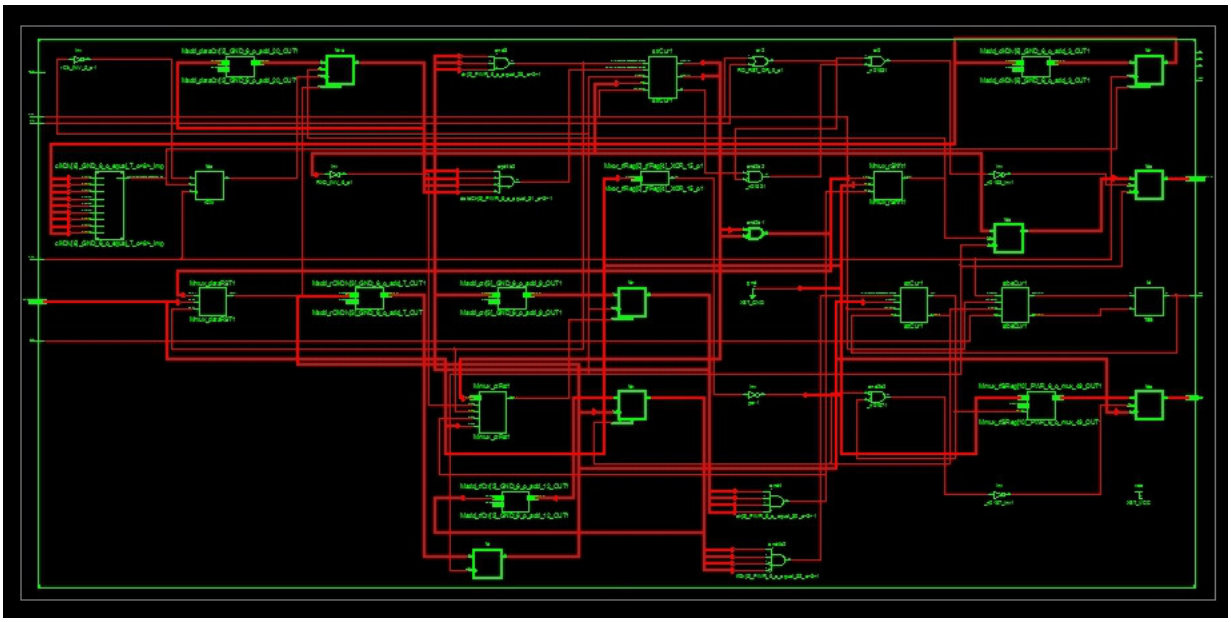


Fig. 11 The UART module's circuit (inside the UART chip)

IV. Conclusion

As we can see, we have created a serial driver chip which can control measuring equipments, power supplies, industrial equipments and even robotic arms.

With this chip we can control any equipment which has an RS-232 serial interface; the only change is that we have to load the specific SCPI commands for the specific equipment we want to control.

We created only in FPGA in two platforms on NEXYS 2 board with Spartan-3E and on ATLYS board with Spartan-6 + Pmod RS232.

After this we plan to convert the FPGA code in Verilog code and with the Mentor Graphics tools to create the chip's layout. After this we can send it to the production to create the silicon die, we shall do the packaging and with this we shall have our own serial driver ASIC.

This ASIC can be than put on a PCB (Printed Circuit Board) with a DB9 connector and some electronic components and we shall have an embedded control board for almost any equipment which has serial port or even a control board for the robotic arms. The only task is to load in a ROM memory the specific SCPI commands for each equipment which needs to be controlled.

References

[1] R. Szabó, I. Lie, "Automated Colored Object Sorting Application for Robotic Arms," *Proceedings of International Symposium on Electronics and Telecommunications ISETC 2012*, Tenth Edition, 2012, pp. 95–98.

[2] R. Szabó, A. Gontean, I. Lie, "Smart Commanding of a Robotic Arm with Mouse and a Hexapod with Microcontroller," *Proceedings of 18th International Conference on Soft Computing MENDEL 2012*, 2012, pp. 476–481.

[3] R. Szabó, A. Gontean, I. Lie, M. Băbăiță, "Comparison between Agilent and National Instruments Functional Test Systems," *8th International Symposium on Intelligent Systems and Informatics (SISY)*, 2010, pp. 87–92.

[4] A. Gontean, R. Szabó, I. Lie, "LabVIEW Powered Remote Lab," *15th International Symposium for Design and Technology of Electronics Packages (SIITME)*, 2009, pp 33–340.

[5] Fengdong Sun, Shufan Hou, Hongwei Zhao, Jianguo Liang, "The FPGA verification of USB to RS-232 bridge controller," *International Conference on Automatic Control and Artificial Intelligence (ACAI 2012)*, 2012, pp. 2123–2127.

[6] N. F. Jusoh, M.A. Haron, F. Sulaiman, "An FPGA implementation of shift converter block technique on FIFO for RS232 to universal serial bus converter," *Control and System Graduate Research Colloquium (ICSGRC)*, 2012, pp. 219–224.

[7] V. Vijaya, R. Valupadasu, B.R. Chunduri, C.K. Rekha, B. Sreedevi, "FPGA implementation of RS232 to Universal serial bus converter," *IEEE Symposium on Computers & Informatics (ISCI)*, 2011, pp. 237-242.

[8] Wong Guan Hao, Yap Yee Leck, Lim Chot Hun, "6-DOF PC-Based Robotic Arm (PC-ROBOARM) with efficient trajectory planning and speed control," *4th International Conference on Mechatronics (ICOM)*, 2011, pp. 1–7.

[9] Woosung Yang, Ji-Hun Bae, Yonghwan Oh, Nak Young Chong, Bum-Jae You, Sang-Rok Oh, "CPG based self-adapting multi-DOF robotic arm control," *International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 4236–4243.

[10] E. Oyama, T. Maeda, J.Q. Gan, E.M. Rosales, K.F. MacDorman, S. Tachi, A. Agah, "Inverse kinematics learning for robotic arms with fewer degrees of freedom by modular neural network systems," *International Conference on Intelligent Robots and Systems (IROS)*, 2005, pp. 1791–1798.

[11] N. Ahuja, U. S. Banerjee, V. A. Darbhe, T. N. Mapara, A. D. Matkar, R. K. Nirmal, S. Balagopalan, "Computer controlled robotic arm," *16th IEEE Symposium on Computer-Based Medical Systems*, 2003, pp. 361–366.

[12] M. H. Liyanage, N. Krouglicof, R. Gosine, "Design and control of a high performance SCARA type robotic arm with rotary hydraulic actuators," *Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2009, pp. 827–832.

[13] M. C. Mulder, S.R. Malladi, "A minimum effort control algorithm for a cooperating sensor driven intelligent multi-jointed robotic arm," *Proceedings of the 30th IEEE Conference on Decision and Control*, 1991, Vol. 2, pp. 1573–1578.

[14] M. H. Liyanage, N. Krouglicof, R. Gosine, "High speed electro-hydraulic actuator for a scara type robotic arm," *International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 470–476.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 671–674

Fuzzy-Based Multi-Stroke Character Recognizer

Alex Tormási and László T. Kóczy
Széchenyi István University
1 Egyetem tér, Győr, H-9026, Hungary
Email: {tormasi, koczy}@sze.hu

***Abstract*—In this paper an extension for multi-stroke character recognition of FUzzy BAsed handwritten character Recognition (FUBAR) algorithm will be presented. First the basic concept of a single-stroke version will be overviewed; in the second part of the paper the new version of the algorithm with multi-stroke symbol support will be introduced, which deploy the same algorithm overviewed in the first part and use flat and hierarchical rule bases.**

I. Introduction

LALOMIA defined the user acceptance threshold in 97% [1], however most multi-stroke character recognition methods known from the literature that are applicable for 26 symbols are well below, on the other hand with a stricted set of symbols (16 gestures) the $N recognizer reached 96.7% [2]. Despite the high accuracy these methods are not always usable for on-line (real-time) handwriting recognition as a result of their high computational complexity and processing time. It is very important to find a recognition engine, which is able to process the input strokes in a short period even on devices with limited resources such as tablets.

In this paper we present a new attempt to recognize multi-stroke letters (26 symbols) with rather good recognition rate (however definitely below LaLomia's 97% threshold). As the starting point the FUBAR algorithm that was very successful for single-strokes will be used, with modifications towards multi-stroke symbols (up to 3 strokes).

After the introduction in this paper the basic steps of the single-stroke FUBAR (Fuzzy Based Recognition) algorithm [3] will be overviewed. In Section 3 the results of the new method with the capability of recognizing multi-stroke symbols are presented. In Section 4 the average recognition rates are analyzed, for the case of the same multi-stroke recognition method with a hierarchical rule-base. In Section 5 the results of the new algorithms and other known recognition algorithms are compared.

II. The Single-Stroke Fuzzy Based Recognition Method

A. Algorithm Properties and Features

During the design of the algorithm the most important goal was to create a recognition engine which is able to process the input strokes with at least the same accuracy of other already published recognition methods, while taking less computational time. Most of the methods published in literature are using geometrical transformations, like rotation and trapezoid correction are complex and resource consuming; to reduce the complexity of the recognition method the use of such transformations was ruled out. The designed recognition method is online, which means that it uses digital ink information to represent the strokes. The alphabet used there is based on a slightly modified version of the Palm's Graffiti single-stroke symbol set [4].

B. Input Processing

The method collects the positions of the digital pen used during the writing process; the strokes are stored in a time ordered list of two-dimensional coordinates. To provide a better input for the next phases of the processing, the input stroke should be re-sampled to provide a low-level anti-aliasing (noise reduction) and provide almost equal distance between the sampled points.

Details of the input processing phase are overviewed in [3, 5].

C. Parameter Extraction

The algorithm uses two kinds of parameters to recognize symbols; the first is the width/height ratio of the stroke while the second type is formed by the average numbers of stroke points in the rows and columns of the fuzzy grid drawn around the stroke. The fuzzy grid approach was introduced in details in [3], in order to handle italics and thus replace the stroke rotation phase used by other methods. The rows and columns of the grids are represented by fuzzy sets [6], which allow a point to belong to two different rows or columns at the same time.

D. Inference

The parameters extracted from a collection of 60 single-stroke character samples were used to determine the rule base for the fuzzy system.

Each symbol in the used alphabet is described by a single rule. The antecedents of the rule are the previously collected parameters and the consequent part represents the degree of matching.

For inference a discrete Takagi-Sugeno method [7] with standard t-norms was used. The algorithm returns with the symbol assigned to the best matching rule.

This paper was supported by the National Scientific Research Fund Grant OTKA K75711 and OTKA K105529, a Széchenyi István University Grant and EU grants TÁMOP 4.2.1 B, TÁMOP 4.2.2/B-10/1-2010-0010.

The detailed description of the single-stroke algorithm and inference can be found in [3, 5].

III. Multi-Stroke Symbol Support

In this part, the results of a new Fuzzy-Based Recognition Engine (FUBAR) family member are presented. This new method is able to process multi-stroke symbols. The method handles each symbol as one non-continuous stroke with "empty" spaces between the sampled points.

The samples were collected from 10 male and 10 female Hungarian participants in the age group 18 to 40. Each subject has provided 20 samples for each 26 multi-stroke symbol. The etalon writing style for the symbols was determined by pre-collecting samples for the most widely used symbol types in the local area for the 26 letters of the English alphabet.

Symbols with obvious errors were identified and removed from the collection and the number of samples per symbols was limited to 180 for a better comparison with the single-stroke system, in which there is the same number of (different) samples was used during the tests in [8, 9, 10].

A similar method was used for data collection as overviewed in [8, 9, 10] with a modification to transform multiple strokes into a single one. This approach gives the capability to the algorithm to process multi-stroke input.

After this step the joint stroke is re-sampled to ensure almost equal distance between the neighbor points.

Each normal trapezoidal fuzzy set describing the stroke parameters in the antecedent part of the rules was constructed according to the statistical process of the parameters extracted from the first 60 samples for each symbol. Each letter is represented by one rule, where the input parameters are collected from a fuzzy grid drawn around the multi-stroke symbol. The analyzed stroke parameters are the width/height ratio and the average number of points in the rows and columns of the fuzzy grid drawn around the stroke (described in previous section); the details of the feature extraction method are the same as in [3, 5].

The same method was used for data collection as overviewed in [3, 5, 8]. The output parameters of the rules are representing the degree of matching between the parameters of the processed input stroke and the information stored in the rules for the described letter.

For the inference a discrete Takagi-Sugeno method was used with standard t-norm (Zadeh t-norm), which returns the symbol assigned to the rule with the highest matching value for the input stroke.

The average recognition rates for the symbols and for the complete alphabet were calculated for the method by the results for 120 samples from the sample set using different fuzzy grid sizes. These are the same conditions as overviewed in [8, 9, 10].

Similar single-stroke method was used for data collection as described in [3, 5], which may give a better environment to compare the results.

The best result for the multi-stroke alphabet was achieved by the algorithm using a 3x4 fuzzy grid; the letter-wise average recognition rates are listed and compared with the results of the similar single-stroke method with various modifications in Table I.

As shown in Table I, the accuracy of the system for the multi-stroke alphabet is 93.4% which is 6.03% and 5,42% less than the results of the single-stroke method (using 6x6 and 6x4 fuzzy grid).

The results have been analyzed in depth including the search for the reason of the false results. Each recognition process that has returned with false result could be traced back to the fuzzy sets describing the rule antecedents. It means there were no false-positive results caused by the overlapping sets of different parameters; and the accuracy might be increased by redefining the rule base.

IV. Hierarchical Rule-Base for Multi-Stroke Alphabet

There are many papers dealing with the use of hierarchical rule bases in fuzzy systems in different areas [11, 12, 13]. A previous work presents the results of the single-stroke method using hierarchical rule base. The details of building the hierarchical rule structure by rule input parameters for single-stroke alphabet were presented in [10].

In the modified system with multi-stroke capability the numbers of strokes were used to determine the meta-level of rules, selecting the group of rules consisting of the given number of strokes. During the tests the same data were used as in the previous section.

This method processes only the selected rules, this way the number of evaluated rules is reduced. The average recognition rates for multi-stroke and single-stroke systems are presented in Table I.

The accuracy of the system for multi-stroke alphabet is the same with flat and hierarchical rule bases. It can be explained by the results presented in the previous section in which it has been stated that there were no false-positive results and all the mistakes were caused by the limited rule base and not by an overlap of the different symbols.

V. Conclusions and Future Work

It was shown that, after the modification, the system was able to recognize multi-stroke alphabets with 93.4% average recognition rate. The results indicate that the accuracy might be further increased by a redefinition of the initial rule base. Finally in this work a similar method with multi-stroke alphabet support using hierarchical rule base was presented. The topology of the hierarchy was built based on the number of the used strokes. The modified system reached the same accuracy as the original one with flat rule base, but the computational cost of the recognition process was considerably reduced by the limited number of rules to evaluate.

The results of the previously introduced altered FUBAR methods are shown and compared to other commercial and academic methods in Fig. 1.

As you may see in Fig. 1 the recognition rates of FUBAR for single-stroke alphabet are higher than other systems.

TABLE I.
AVERAGE RECOGNITION RATES OF THE ALGORITHM FOR SINGLE-STROKE AND MULTI-STROKE ALPHABETS

Symbol	Average Recognition Rates of FUBAR (%)				
	Single-Stroke FUBAR with 6x6 fuzzy grid	***Single-Stroke FUBAR with 6x4 fuzzy grid***	***Multi-Stroke FUBAR with 3x4 fuzzy grid***	***6x4 Single-Stroke FUBAR with hierarchical rule base***	***3x4 Multi-Stroke FUBAR with hierarchical rule base***
A	100	100	96.1111	100	96.1111
B	95.5555	92.7374	89.4444	92.7374	89.4444
C	99.4444	97.7654	76.6667	97.7654	76.6667
D	98.3333	98.8827	96.6667	98.8827	96.6667
E	99.4444	97.2067	92.2222	97.2067	92.2222
F	100	100	96.1111	100	96.1111
G	100	100	97.2222	100	97.2222
H	100	100	95.5556	100	95.5556
I	97.7777	100	98.3333	100	98.3333
J	100	100	97.7778	100	97.7778
K	100	99.4413	96.6667	99.4413	96.6667
L	100	100	98.3333	100	98.3333
M	100	100	96.1111	100	96.1111
N	100	96.0894	96.6667	96.0894	96.6667
O	97.7777	93.8548	92.7778	93.8548	92.7778
P	100	100	92.7778	100	92.7778
Q	98.8888	100	97.7778	100	97.7778
R	98.3333	100	87.7778	100	87.7778
S	100	100	97.7778	100	97.7778
T	100	100	96.1111	100	96.1111
U	100	97.2067	93.3333	97.2067	93.3333
V	100	98.3240	88.3333	98.3240	88.3333
W	100	100	92.2222	100	92.2222
X	99.4444	98.3240	89.4444	98.3240	89.4444
Y	100	99.4413	91.1111	99.4413	91.1111
Z	100	100	85	100	85
Average	**99.4231**	98.8182	*93.3974*	98.8182	93.3974

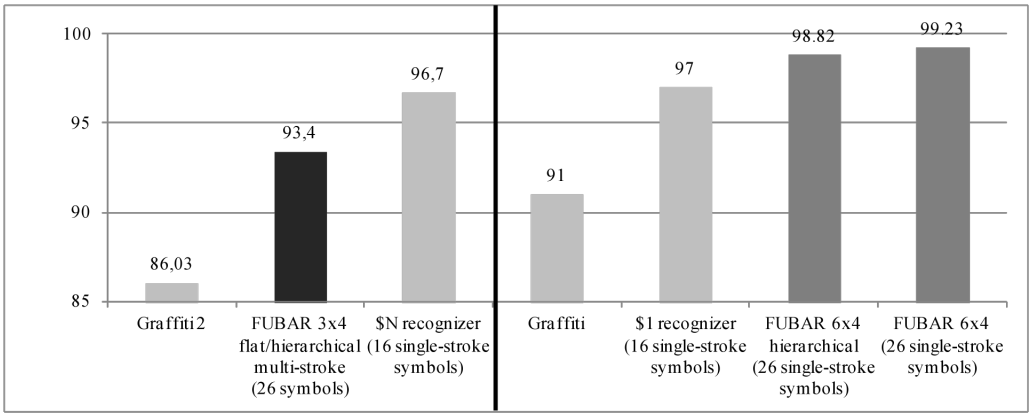


Fig. 1 Average recognition rates of various recognition engines (multi-stroke engines on the left, single-stroke engines on the right)

The $1 recognition method reached 97% average accuracy for only 16 different symbols [14], while the single-stroke version of the designed system reached over 99% recognition rate for 26 different symbols. Another advantage of the new algorithm is the computational cost of the recognition, which is linear in each phase.

Fleetwood et al. showed that users could reach only 91% average recognition rate with the original Graffiti recognition method [15], which is less than the accuracy of the presented system. The Graffiti alphabet contains 26 different English letters and other control symbols.

The average recognition rate of the modified Palm Graffiti with limited multi-stroke support (known as Graffiti 2) was studied by Költringer and Grechenig in [16]. The method reached only 86.03% accuracy. Both the single-stroke and multi-stroke versions of FUBAR performed well over the results of Graffiti 2.

The $N recognizer [2] (the multi-stroke version of the $1 method mentioned above) achieved 96.7% average recognition rate for 16 different symbols.

It is important to highlight the fact, that the computational complexity of the proposed recognition engine is linear, while most of the commercial and academic systems have a quadratic or higher computational complexity. This means that, other systems need much more time to compute the results even if the alphabet is extended by one symbol. The computational cost of FUBAR method increases however only linearly.

Currently we are working on a new method to build the initial rule base for the system, which may increase the recognition rate of the algorithm.

Another extension for the method is under development, in which the output rules are presented by discrete type-2 fuzzy sets. The preliminary results of the test are showing that the recognition rate can be increased by the mentioned modification without a significant increase of the computational cost.

In the future we intend to investigate an extension of the present algorithm, where the possibility of applying two or more rules representing a single character will be considered.

REFERENCES

[1] M.J. LaLomia, “User acceptance of handwritten recognition accuracy,” Companion Proc. CHI ’94, New York, 1994, p. 107.
[2] L. Anthony and J.O. Wobbrock, “A Lightweight Multistroke Recognizer for User Interface Prototypes,” Proc. GI 2010, Ottawa, 2010, pp. 245–252.
[3] A. Tormási, and J. Botzheim, “Single-stroke character recognition with fuzzy method,” New Concepts and Applications in Soft Computing SCI, vol. 417, V.E. Balas et al. (eds.), 2012, pp. 27–46.
[4] A. Butter and D. Pogue, Piloting Palm: The inside story of Palm, Handspring, and the birth of the billion-dollar handheld industry, John Wiley & Sons, Inc., New York, 2002
[5] A. Tormási, and L.T. Kóczy, “Comparing the efficiency of a fuzzy single-stroke character recognizer with various parameter values,” Proc. IPMU 2012, Part I. CCIS, vol. 297, S. Greco et al. (eds.), 2012, pp. 260–269.
[6] L.A. Zadeh, “Fuzzy sets,” Inf. Control, 8:338-353, 1965.
[7] T. Takagi, and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-15, pp. 116-132, 1985.
[8] A. Tormasi, and L.T. Kóczy, “Efficiency and accuracy analysis of a fuzzy single-stroke character recognizer with various rectangle fuzzy grids,” Proc. CSCS ’12, Szeged, 2012, pp. 54–55.
[9] A. Tormási and L.T. Kóczy, “Improving the Accuracy of a Fuzzy-Based Single-Stroke Character Recognizer by Antecedent Weighting,” Proc. 2nd World Conference on Soft Computing, Baku, 2012, pp. 172-178.
[10] A. Tormási and L.T. Kóczy, “Improving the Efficiency of a Fuzzy-Based Single-Stroke Character Recognizer with Hierarchical Rule-Base,” Proc. 13th IEEE International Symposium on Computational Intelligence and Informatics, Óbuda, 2012, pp. 421-426.
[11] M. Sugeno, F.M. Griffin, and A. Bastian, “Fuzzy hierarchical control of an unmanned helicopter,” Proc. IFSA ’93, Seoul, 1993, pp. 1262-1265.
[12] M. Sugeno, and K.G. Park, “An approach to linguistic instruction based learning,” Intern. Journal of Uncertainity, Fuzziness and Knowledge-Based Systems, 1(1):19–56, 1993.
[13] L.T. Kóczy, and K. Hirota, “Approximate inference in hierarchical structured rule bases,” Proc. IFSA ’93, Seoul, 1993, pp. 1262–1265
[14] J.O. Wobbrock, A.D. Wilson and Y. Li,, “Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes,” Proc. UIST '07. ACM Press, New York, 2007, pp. 159–168.
[15] M.D. Fleetwood et al., “An evaluation of text-entry in Palm OS – Graffiti and the virtual keyboard,” Proc. HFES ‘02, Santa Monica, CA, 2002, pp. 617–621.
[16] T. Költringer and T. Grechenig, “Comparing the Immediate Usability of Graffiti 2 and Virtual Keyboard,” Proc. CHI EA ’04, New York, 2004, pp. 1175–1178.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 675–678

Image Recognition System for the VANET

Štefan Toth, Ján Janech, Emil Kršák
Department of Software Technologies
Faculty of Management Science and Informatics
University of Žilina
Univerzitná 1, 01026 Žilina, Slovakia
Email: {stefan.toth, jan.janech, emil.krsak}@fri.uniza.sk

Abstract—This paper describes a system for recognition of objects in traffic scene from multiple moving vehicles. The system is based on query and image processing. It allows image recognition along with a spatial relation. A user or a machine makes a query in which description of searched objects are defined. Then the query is sent to vehicles in order to process in real-time. If any vehicle recognizes object of interest given by the query, the answer is returned to query author.

I. Introduction

AT the present time VANET (Vehicular Ad hoc NETwork) is considered as one of the most important technologies in ITS (intelligent traffic systems) [1][2]. It refers to an ad-hoc network in which vehicles communicate each other or with the infrastructure. Nowadays this is a promising research area, in which many researchers and vehicle manufacturers have proposed and developed many applications expected to improve traffic safety or increase comfort of users.

In this paper we propose one of the applications focused on image processing. It allows recognition of objects in traffic scene using a camera placed in a vehicle. It proposes a novel approach based on a query processing in which a user or a machine is able to make a complex query whereby objects are detected and recognized.

II. Similar Works

VANET networks are prospective future of the automotive industry, as it provides tremendous opportunities in terms of improving safety or comfort for users. In the field of image processing and analysis, many applications have been proposed such as a vision-based active safety system for driver assistance in intersection scenarios [3], See-Through System [4] in case of passing large vehicles and vehicles with darkened rear windscreen, which are difficult to see through and make the situation before them unpredictable. Another example is a system for locating wanted vehicles [5] based on license plate recognition and also a surveillance system proposed by Badura [6].

All applications, however, are problem specific. They solve a particular problem of exactly defined area for which they were designed. Our research has not yet identified any existing solution that would be able to search for general information – the objects contained in the image on the request.

Although many methods for real-time detection and recognition of objects have been developed using MPEG-7, QBIC / CBIR systems, we have to point out that generic object detection and recognition system suitable to be used in VANET is missing. Thus we are proposing a solution for detection of traffic objects and also general objects appeared in traffic scene visible from vehicle.

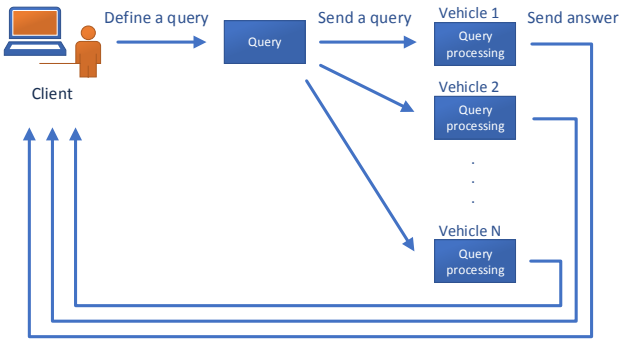


Fig. 1 Outline of the proposed query system

III. The Query System

The idea of the proposed system is very simple. It consists of following steps:

1. Query definition – a query is created manually by a person or automatically by an application. In a query an object of interest is defined, as well as required output and processing length.
2. Query sending to vehicles – a query is sent to one or more vehicles (query processing) according to defined target area in which the image is supposed to be processed (the area is defined, for example, using IP broadcast or geocast).
3. Query processing and output return – since a requested object can appear sporadically, the image captured by a camera could be processed periodically until a condition is met. Thus the length of processing can be limited either by certain time (e.g. 15 minutes), by

978-83-60810-53-8/$25.00

event (e.g. after required object was found, an engine of a vehicle was turned off) or by geographical range in space (spatial definition specified by GPS location).

The basic principle is also shown in Fig. 1.

A. Image Objects

The base of the system is image object, which constitutes either a real object or segmented part in an image. In order to recognize real objects we performed dozens of test drives in surroundings of Žilina city (Slovakia). We have identified some important objects, which could be recognized in the system:

- Sky (sun, clouds, weather condition)
- Mountains
- Roads (traffic lane, sidewalk) and intersections
- Horizontal traffic signs
- Vertical traffic signs
- Traffic lights
- Texts
- Vehicles
- People
- Animals
- Buildings and poles
- Trees, bushes
- Lakes, rivers
- Objects moving in the air (flying objects)

In order to recognize unknown objects, we define the so-called general object which is composed of a model consisting of set of templates with a description. This model will be directly incorporated into query so it will be possible to refer to it and be used in querying.

After analysis we suggest that the general object model would contain following parts:

- Unique model identifier should be included in order to refer to it in a query.
- Object characteristics – deformability, motionlessness, size etc.
- Views of model are expressed either in form of a set of source images, features or in the form of model with value parameters of a machine learning algorithm. In order to recognize an object, source images would represent the object from different viewpoints (especially if the object is not symmetric) and also capture its various visual aspects in case of deformability and especially in such positions it can frequently be seen.
- Described parts of the model which make up the model (e.g. in case of vehicle if there is a view of rear part of it, we can see a trunk, right and left light, part of right and left wheel etc.). Parts are defined for every viewpoint and will have certain position in order to refer to them in querying and e.g. to find out if they are visible or if they are in relation with other objects.

A general object can be expressed for example by using templates of still images, SIFT (Scale Invariant Feature Transform), ANN (Artificial Neural Network), or other state-of-the-art methods. Thus a general object could be represented by defining the source data such as images, descriptors, or weights as required by chosen method.

B. Querying Image Objects

We can look at querying image objects as a complex task which can be decomposed into simpler tasks and their interconnections (Fig. 2). Thanks to these interconnections the sets of image objects will pass in left-to-right direction in order to process these sets in elementary tasks.

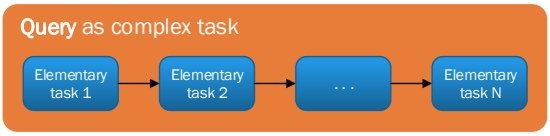


Fig. 2 Query as complex task composed of N elementary tasks

When we get to the last elementary task N, it depends on its output what happens next (Fig. 3). If the output is non-empty set, the answer will be sent in form which is defined at the end of the query. Otherwise, it depends on definition of the query whether the requirement is to be re-processed or terminated and if it terminates whether to send an answer. This behavior was chosen for a reason that it is very often necessary to search for an object in an image which is likely not to be present there. Therefore we need cyclic query processing until the answer (non-empty set) is obtained. Actually, an empty set will represent insufficient output (i.e. required object is not found).

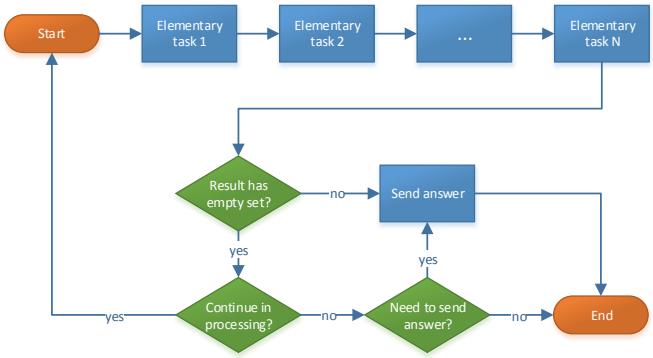


Fig. 3 The proposed system of processing a query as complex task

Processing of an image query therefore means executing of elementary tasks involved in sequentially concatenating sequence. Elementary task represents any basic operation with the image, which can be in a form of detection, localization or classification of objects, relations between them, choosing and other basic image processing algorithms built and supported in the system.

C. Inputs

The input is defined by a client. In this query, it is important to determine what to perform, when to finish the

processing and what to send as an answer. Such information is possible to write in a certain defined format, e.g. XML, JSON or others. Since we will work in VANET environment, the format should be as compact and small as possible because of the transmission. Our aim is not exactly to define exact format in which data will be transmitted via the network, but only query format in which it will be processed (i.e. what and which parts are important in a query). Therefore, we define following required and optional fields:

- `INPUT` – defines a unique query name, which is also used in the output answer.
- `MODEL` – a definition of general object recognition model. Definition of the model is optional, unless we require recognition of an unknown and undefined object in the system.
- `TERMCONDITION` – a termination query condition. It is optional in case of a query with quick response (only one image processing obtained from the camera is requested) and necessary if we want to process images cyclically (video sequence). Specified condition then determines when the processing loop finishes.
- `REQANSWER` - determines whether we always require an answer, even if the query returns an empty set.
- `QUERY` – a definition of a query. It is a mandatory field, which forms the core and allows you to define what objects with characteristics are required to obtain. The last elementary task will define the output of the query.

D. Outputs

As a result of the output can be everything what is needed by original request, e.g. original image captured by the camera or only some parts of the image where are one or more objects of interest. In addition it could contain GPS coordination, common value (number, text, logical value) containing required information (the number of objects, recognized text, licence plate number etc.) or their combination.

IV. The Query Operations and Operator

In query definition we could use many elementary tasks sequentially connected each other. These tasks constitute possible operations connected by separation operator.

A. Operations

For purpose of manipulating image objects we have introduced operations as following:

- Input and output functions
  - `GETIMAGE` – captures an image from a camera or another image source and converts it to an image object. Captured image represents traffic scene as a root image object.
  - `RESULT` – returns a set of objects defined by an expression being its argument as query result. The result should be sent back to the querying vehicle.
- Object searching operation
  - `FINDOBJECTS` – detects and recognizes requested image objects in the input set of image objects. The used algorithm of object recognition depends on the system where the query is executing. It can be used any state-of-the-art recognition algorithm.
- Manipulation operations
  - `WHERE` – selection is an operation for filtering the input set of image objects by the given condition.
  - `SELECT` – collection; it is an operation for transforming each item from the input set using the expression given as the operation argument.
- Set operations
  - `UNION` – standard set union between the input set and the set given as an operation argument.
  - `INTERSECTION` – standard set intersection between the input set and the set given as an operation argument.
  - `MINUS` – standard set difference between the input set and the set given as an operation argument.
- Temporal data storage operations
  - `SAVE` – stores its input into temporal data storage.
  - `LOAD` – loads data from temporal data storage and passes it as its output.
- Spatiotemporal operations
  - `RIGHTOF` – selects all image objects on the input that are on the right side of the objects given as operation argument.
  - `LEFTOF` – selects all image objects on the input that are on the left side of the objects given as operation argument.
  - `ABOVE` – selects all image objects on the input that are above of the objects given as operation argument.
  - `BELOW` – selects all image objects on the input that are below of the objects given as operation argument.
  - `IN` – selects all image objects on the input that are inside of the objects given as operation argument.
  - `OUT` – selects all image objects on the input that are outside of the objects given as operation argument.
  - `NN` – selects nearest object to the objects from the set on the operation input from object set passed as the operation argument.
  - `DISTANCE` – selects all objects from the set on the operation input with its distance to any object from set passed as the operation argument.

B. Separation Operator

As separator between operations, operator | was introduced. Using that operator we can express following sequence of operations:

```
operation1 | operation2 | ... | operationN
```

where the last operation `operationN` will be a result of the query.

V. Example of Usage

To demonstrate the proposed system and language we have presented some examples of using it.

A. Image of Traffic Scene

The simplest query is returning single image from a camera:

```
GETIMAGE
```

Here we do not have to use separation operator and also RESULT operation since we do not require any additional information. If we would like to return more information and from rear camera, we can use query as following:

```
GETIMAGE('rear camera') |
RESULT(image => image, GPS.GetPosition)
```

This query means to return image and current GPS position of the car.

B. Car Detecting

If we would like to find a red car, we should use sequence of following operations:

1. Get an image from a camera (`GETIMAGE`).
2. Detect car objects in the image (`FINDOBJECT`). The result of this step is a set of image objects containing a car.
3. In order to find only a red car, we can then use `WHERE` operation to select only a car with red color.

Final query will be composed as follows:

```
GETIMAGE | FINDOBJECTS('car') |
WHERE(car => car.Color = 'red')
```

It depends on a person how he defines a query. There are many possibilities how to make the same search for a red car. We could use the color segmentation at first and then apply finding objects operation:

```
GETIMAGE | FINDOBJECTS('red segment') |
FINDOBJECTS('car')
```

C. More Complex Example

To demonstrate a more complex example, we define a query in which we will detect a red car with a license plate starting with ZA and ending with AB. In addition there was a specific picture (for example skull with crossbones) on right side of the license plate number. Therefore we define a general object consisting of wanted picture in query format, so we could use it in query. Then result query by that definition could be like this:

```
GETIMAGE | FINDOBJECTS('car') |
WHERE(car => car.Color = 'red' AND
             car.View = 'rear') |
FINDOBJECTS('license plate') |
WHERE(plate => plate.Text.StartWith('ZA') AND
               plate.Text.EndWith('AB'))|
SELECT(plate => plate.ParentObject.Crop(
       plate.Location.Increase(0,100,400,100))) |
FINDOBJECTS('skull') |
RESULT(skull => skull.ParentObject, GPS.Position,
                GPS.Heading)
```

As result of that code is an image of a car (ParentObject), GPS position and heading.

Other examples of proposed system are introduced in [7][8].

VI. Conclusion

In this paper we proposed a novel application for the VANET, whereby we are able to recognize objects from images captured by a camera placed in a vehicle. We described a query system and language based on object oriented manner. Using that language we can detect any objects along with relations defined by a query.

Acknowledgment

This contribution/publication is the result of the project implementation:

Centre of excellence for systems and services of intelligent transport II., ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

"Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ"

References

[1] J. E. Naranjo, J. G. Zato, L. Redondo, M. Oliva, F. Jimenez and N. Gomez, “Evaluation of V2V and V2I mesh prototypes based on a wireless sensor network”, 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, DC, USA, 2011.

[2] J. Janech, A. Lieskovský, E. Kršák, “Comparation of Strategies for Data Replication in VANET Environment”, 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), Fukuoka, Japan, 2012.

[3] L. Hoehmann, A. Kummert, “Car2X-communication for vision-based object detection”, International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Dubrovnik, 2010.

[4] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira and M. Ferreira, “The See-Through System: A VANET-enabled assistant for overtaking maneuvers”, IEEE Intelligent Vehicles Symposium (IV), San Diego, CA, 2010.

[5] M. Ferreira, H. Conceição, R. Fernandes and R. Reis, “Locating cars through a vision enabled VANET”, IEEE Intelligent Vehicles Symposium, Xi'an, 2009.

[6] Š. Badura, A. Lieskovský, “Intelligent traffic system: cooperation of MANET and Image processing”, IEEE First International Conference on Integrated Intelligent Computing (ICIIC 2010), Bangalore, India, 2010.

[7] Š. Toth, Spracovanie obrazu s využitím dopytov v prostredí VANET (Query Based Image Processing in the VANET), PhD dissertation, Dept. of Software Technologies, University of Žilina, Žilina, Slovakia, 2013.

[8] Š. Toth, J. Janech, E. Kršák, “Query Based Image Processing in the VANET”, 5th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN2013), 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 679–682

Simulation of energy consumption in a microgrid for demand side management by scheduling

Weronika Radziszewska
Research Systems Institute
Polish Academy of Sciences
Warsaw, Poland
Email: Weronika.Radziszewska@ibspan.waw.pl

Zbigniew Nahorski
Research Systems Institute
Polish Academy of Sciences
Warsaw, Poland
Email: Zbigniew.Nahorski@ibspan.waw.pl

***Abstract*—Energy management systems (EMS) are necessary when smart grids and microgrids are considered. Simulation of energy consumption is very useful in planning and testing such systems. In this article we present the problems of simulating energy consumption and show concept of a very general load simulator. The simulator can generate time series of consumption from fixed profiles and also from the defined rules describing use of energy by the devices. The rules describe the probabilistic distribution of the device behaviour. The architecture of the implementation is also presented.**

I. Introduction

THE possibilities to test the energy management system in reality are very limited due to lack of existing microgrid infrastructures. But they may be tested using simulators. Data about wind speed, irradiance and temperature required for renewable energy sources simulation can be obtained from direct measurements or meteorological models.

Simulators of consumed energy described in the literature are usually simple as main effort is channelled towards creating management systems for the next generation of electric networks. They usually are based on general profiles collected from few devices. Each device has its own profile of energy requirement that varies in time. The amount of energy used by given equipment can be measured, but the general, statistical data of how frequently and how long people use devices are missing. Attempts have been done to measure the average amounts of power that different groups of consumers use during longer period. A report about the energy usage in Spain [1] is the most complete in that field (in [2] the short summary of the [1] in English is presented). Due to huge differences in culture, climate and wealth of the regions, the results of such research cannot be directly used in simulation of grids in different geographical locations, making the ability to simulate systems in defined localisations difficult.

The contents of the paper is as follows. The following section II will explain the idea of a microgrid and the context of the simulation. An idea of the character of the microgrid considered in this work will be also presented. In section III, the overview of the energy production consumption is described. The next section IV presents different methods of describing consumer behaviours, section V describes the concept of the simulator. The last section concludes the article.

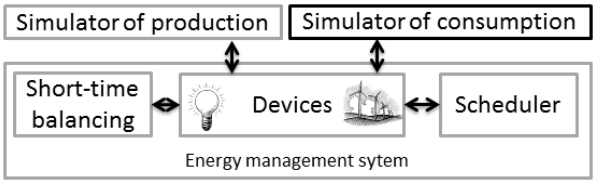


Fig. 1. A diagram of the elements of the considered system, where devices are represented by agents being parts of the Short-Time Balancing System, which uses production levels of controllable power sources from the Scheduler.

II. Smart grids and microgrids

Smart grids and microgrids seem to be the future trend in the energetic revolution that is ahead. A *smart grid* is a concept of introducing exchange of information between different elements of electrical grid (consumers, producers, storage units and prosumers). Thanks to that, controlling and coordinating of supply and demand of energy can be introduced to ensure quality of electric power in the grid, reduce the cost and promote renewable energy sources. A *microgrid* is a part of the grid, that might include producers, consumers, energy storage units and prosumers, which has the ability to connect or disconnect to/from the external power grid and balance the energy within itself.

These new technologies require an advanced control system that can use the potential of bidirectional communication. Implementing such systems requires working in real time operation mode. It is a challenge, as consumption and production is changing very dynamically, due to users activities and weather conditions.

In this article a small microgrid consisting of few buildings and connected to an external distribution network is considered. The general overview of this grid is presented in Fig. 1. The microgrid is a research and education centre with a hotel and a restaurant. Its energy producers and consumers are controlled by the complex Energy Management System (EMS), which can be divided into two main parts: the Scheduler and the Short-Time Balancing System. Detailed description of this system can be found in [3].

The Scheduler (under development by Wroclaw University of Technology) is a program that arranges the planned events and tasks in order to minimise the cost of the microgrid

978-83-60810-53-8/$25.00

operation (the cost of obtaining energy necessary to power all the tasks). The input data to the Scheduler are information on planned events, e.g. organization of a conference, a training, conduction of an experiment, hosting a person in the hotel, etc. All these events have defined time constraints, usage power profiles and locations where events can take place. The Scheduler is using a heuristic algorithm to place a task in a location at a certain interval of time.

The Short-Time Balancing System is dealing with deviations from the schedule. Its main goal is to balance the produced and consumed energy as fast as possible. It is implemented as a multi-agent system. Wooldridge in [4] defines an *agent* as a program that fulfils its goals by taking autonomous decisions based on the data received from the environment (sensors, input information). The concept of a multi-agent system as considered in [5] and [6] fits well to handle the problems of power grids. In the Short-Time Balancing System an agent is assigned to each source, energy storage unit and load node of the network. A node is an aggregation of consuming devices, e.g. one line of sockets on one floor of a building. The load nodes are divided into two groups: the ones that have to be powered (the reserved nodes) and the ones that can be switched off under power deficit (the unreserved nodes). An agent receives information about the state, the energy produced or required from its device. When the device is in an unbalanced state, the agent negotiates with other agents to contract energy for its device (both in case of its excess or deficit).

EMS is ready and working, but testing it requires running it for a certain amount of time, e.g. one year, and multiple times, to get the average time of balancing and the number of imbalances. In [7] authors assume that a test of multiagent system is statistically significant with the simulation size of at least 200.

III. Simulation of energy consumption

Simulating the amount of energy produced by renewable sources requires simulating the weather conditions. The data such as temperature, wind speed and water flow are available for us for a large number of years, but insolation was measured for much shorter time and might not be sufficiently long for an exhaustive testing.

In any case, the gathered information is not sufficient if long time simulations have to be done (e.g. to test how system copes with seasonal differences). It is necessary to generate long time series of data. We used for this a block-matched bootstrap method which samples available data and assembles them to create a time series that has statistical properties close to the original data, and is of the required length [8], [9].

Simulation of energy consumption is more complex, because there is usually a large number of heterogeneous loads considered. Consumers can be considered at different aggregation levels. In a household, usually single devices, as oven or microwave, are considered [2]. In larger networks, at levels of groups of houses, general profiles are used (like in [10]). In large networks profiles are grouped by sectors, like commercial, residential, industrial.

For some purposes the general profiles are sufficient, e.g. in [11] they are used to verify the design of the network (to identify possible overloads or violation of constraints). Only eighteen exemplary load-flow calculations are presented there, with 19 profiles for different categories of loads, but they cover all extreme situations, like e.g. extremely high consumption with no production from renewable sources. The tests confirmed that the network was well designed and there is no threat of overload. But such load profiles are not good enough (values of a profile are 1-hour averages, so there are only 24 different load values for a day) to test the dynamic behaviour of the microgrid.

Profiles for a big group of consumers can be easily derived, as any outstanding or not common behaviours tend to be compensated by each other, so they do not vary very rapidly. On the country scale they can be easily obtained from large power producers. Profiles show cycles of daily and weekly changes that reflect the human activities. Night is usually the time of lower energy usage, and its peak usage is around late afternoon. Weekends and holidays are introducing disturbances to the working day cycles. Moreover, seasonal differences are visible, caused by changes in the outer temperatures (e.g. large amount of power is used for air-conditioning), long holiday seasons and changes in labour structure [1].

On the contrary, in microgrids each consumer has a relatively bigger influence on the profile than in large grids, e.g. a 4kW induction cooking plate will not be visible in profile on the regional level, but can dominate the energy usage in a single household. Thus, profiles are not sufficient for microgrid simulation purposes.

The most comprehensive research about structure of energy usage has been done in Spain [1]. Users presented in the report are divided in 5 groups: residential, commercial, touristic, large consumers and others, whith the total contribution of power usage 20%, 6%, 0.5%, 25%, and 48.5%, respectively. These values might differ among regions and countries and depend on the method of categorisation. The authors of the report emphasise the big differences in the energy usage between user groups, as for example households, tourist facilities or companies. Other factors that influence the amount and structure of power usage are e.g. seasons of the year, days of week, times of day, months, holiday distributions, structure of labour and economic situation. It demonstrates the difficulty to obtain one reliable description of consumer structure even within small area.

The EMS considered in the present paper governs a relatively small microgrid. The maximum necessary load does not exceed 900 kW. In this situation, a room where a computer lesson takes place can use easily 4.5 kW, which is a considerable amount. Such lessons can be planned and entered to the Scheduler that would inform energy management system about an increase in power. Power usage of computers in a room, projector, air conditioning and lights are gathered and their

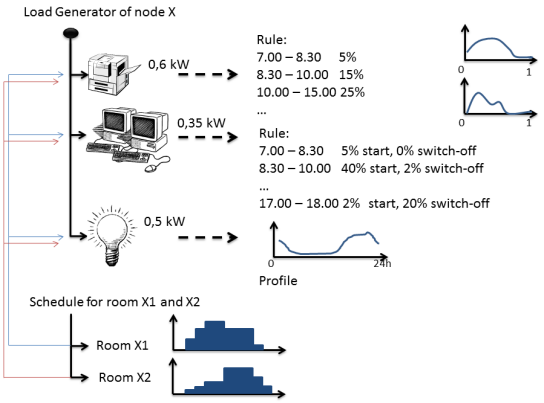


Fig. 2. A diagram of different possible descriptions of energy consumption.

average power usage is placed in the schedule for a specific time with a duration of e.g. 1.5 hour.

For the Short-Time Balancing System, the execution of the task ”computer lesson” would mean the increase of power on two nodes of network, the one that would power the computers (which is reserved, i.e. the node has priority in receiving power) and the other for lights and additional equipment. That means that two agents would ”sense” the increase of power usage and start the balancing procedure. The load simulator generates the power that is consumed in certain nodes. It also knows which nodes in what level contribute to the supplying energy for the location that Scheduler chose for the executed task. Considering both Scheduler locations and node division, it is required to perform the load simulation in smaller parts of the microgrid than belonging to one node. Simulating the power usage of each device gives much more accuracy, makes the simulation less abstract and gives possibility to base the model on existing devices, whose parameters might be measured or found in the literature. In [12] a detailed analysis of representative office environment was conducted to test the model designed. 500 electrical devices were identified, mostly usage dependent. The modelling of users behaviour regarding the use of electric equipment is the most difficult part of simulation, as people do not like to be interrogated. Unfortunately, knowledge of typical human behaviours of using devices is crucial to carry out reliable power load simulators.

IV. Description of consumer behaviour

Devices consume power because people placed them there, switched them on and use them. The load simulator, in reality, tries to mimic the patterns of human behaviour. It cannot model the whole complexity of human reasoning, but can derive general patterns and statistical distribution of certain human actions.

Usage of energy by some devices can be described as a profile, which is an approximation of a function of energy usage of the device. Device profiles are made to represent energy usage by a device during a certain time period. Such profiles come from real measurements and are applicable for the devices (or group of devices) that have stable and defined work cycles. Examples may be a coffee machine, a fridge or a freezer. Profiles are also reliable when there are many small consumers of energy, for example light bulbs. In this case a single device has little influence on the overall power consumption and multiple small deviations tend to level the usage. Profiles define the average, typical behaviour and are not suitable to describe events that happen with low frequency or of extreme power usage. For example, the profile of a coffee machine is repeatable and can be measured, but the information of how often and when users make coffees has to be derived from statistical behaviour. Simulators based on profiles encounter troubles to represent small variability in the generated data, even when random disturbances are introduced.

Simulator might increase the diversity of generated data by using multiple profiles for a single device, e.g. there might be 10 profiles for a computer. It can be switched on for 1 hour or for 24 hours, might be used for energy demanding calculations or might be in a sleep mode for most of the time. This approach would require a large number of different profiles that would represent certain cases and still would not show all possible combinations.

Power consumption of employees’ computers might be dependent on the current circumstances, the work the person is doing, the habits of different people and it is difficult to obtain a general profile. The method of describing that behaviour may be a probability distribution of switching on the device. That means describing loads by a set of rules. This type of description is introduced in [1] according to the Spanish behavioural data. The work of elements like dishwashers, ovens, etc. is described by the probability of their operating in a certain time. For example, an electric kitchen (a stove) is mainly used around 9:00, 13:00, and 21:00 hours with the respective probability around 20% at the 21 o’clock, 10% at the 13 o’clock, and 2% at the 9 o’clock [1](page 100).

We required that the simulator developed should be as general as possible, to be able to simulate operation of most of existing devices. That can be obtained by combining the ideas of rules and profiles. The example of such description for devices connected to one node is presented in Fig. 2. For devices described by a profile, like for a fridge or a freezer, the profile is used. Devices that are activated by a person and person actions control them, are described by rules. For the loads that have defined profiles for actions, but the actions are executed by users with a certain probability, the simulator uses both the profile and the rules. Rules define a probability of starting an action at certain time. When a device is active, the simulator generates consumption data according to its profile. A rule is a set of parameters that describe the operation of a device. These parameters describe the probability, the activation time and the intervals in which certain use of electric power happens. Different types of rules may define the time and the probability when the state of the device changes. For

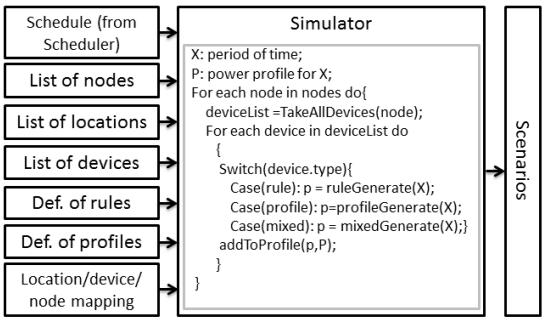


Fig. 3. Concept of the Simulator of consumption with data sources, outcome and general description of the algorithm.

example, the computer is switched on and continues this state until the user switches it off, which is also defined by a rule. Simulator use a definition of a consumption described as a set of rules with probabilities of activating an action. One rule is bound to one type of device but may be a complex one and consist of a probability of switching on, off and changing the working point of devices.

V. Concept of the simulator

The simulator is designed to generate load data for each node for a certain period of time, with a given start date and a time. Generated data are stored as test scenarios which allows to repeat the test with different configuration of sources. The schema of the system is presented in Fig. 3.

Data that have to be available for the simulator consists of the schedule made by the Scheduler, the list of nodes with information how many and what type of devices are connected to them, the mapping between devices, nodes localisations (e.g. rooms), the profiles of nodes and individual devices that are connected to a node, and the rules for devices without profiles. The outcome of the simulator are power values aggregated for each consumption nodes of the network, with the sampling frequency defined by a parameter.

The simulator processes each node separately in order of their numbering. It queries all the devices connected to the node and then generates for each device the load for the requested time period. Then it sums up all power consumptions of the loads connected to the node, at each sampling time. Each device is processed depending on the type of the device, and the load is generated from the profile or from the rule. The most important factor is the date and the time, as both rules and profiles are parametrised by them.

VI. Conclusion

Testing is an important step in developing EMS, especially when systems work in a microgrid environment, where small changes in load have a big impact on overall balance. To have statistically significant data about microgrid operation, a large number of long-term tests has to be made. A real infrastructure for testing purposes is often not available. Detailed profiles of energy usage of devices can be measured, but they do not reflect the way people use devices. User behaviour is very varied and influenced by many factors. Simulator of energy consumption has to mimic this behaviour with all its impreciseness and unpredictabilities, which requires using probabilistic distribution combined with fixed profiles. Presented energy consumption simulator requires rules and profiles that define device's behaviour. Based on that it creates time series of energy consumption aggregated per node, which is a tool for EMS testing.

It is clear that more efforts should be made to examine the nature of different energy consumers to obtain the statistical distribution of loads considering different social and environmental factors. That would also help to find where energy is wasted and how to avoid it. The next stage of the research is exhaustive testing of the EMS and then connecting it to real devices.

Acknowledgment

The research of W. Radziszewska was supported by the Polish Ministry of Science and Higher Education under the grant N N519 580238, and by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

References

[1] "Atlas de la demanda eléctrica española," RED Eléctrica de españa, Tech. Rep., 1999.

[2] M. Vasirani and S. Ossowski, "A collaborative model for participatory load management in the smart grid," in *Proc. 1st Intnl. Conf. on Agreement Technologies*. CEUR, 2012, pp. 57–70.

[3] P. Pałka, W. Radziszewska, and Z. Nahorski, "Balancing electric power in a microgrid via programmable agents auctions," *Control and Cybernetics*, vol. 4, no. 41, pp. 777–797, 2012.

[4] M. J. Wooldridge, *Introduction to Multiagent Systems*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[5] S. McArthur, E. Davidson, V. Catterson, A. Dimeas, N. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-agent systems for power engineering applications Part I: Concepts, Approaches, and Technical challenges," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1743–1752, 2007.

[6] ——, "Multi-agent systems for power engineering applications Part II: Technologies, Standards, and Tools for building multi-agent systems," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1753–1759, 2007.

[7] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings, "Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid," *J. Artif. Int. Res.*, vol. 42, no. 1, pp. 765–813, Sep. 2011.

[8] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[9] T. Hesterberg, "Matched-block bootstrap for long memory processes," MathSoft, Inc., Tech. Rep., 1997.

[10] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings, "Agent-based micro-storage management for the smart grid," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1*, ser. AAMAS '10. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 39–46.

[11] J. Wasilewski, M. Parol, T. Wojtowicz, and Z. Nahorski, "A microgrid structure supplying a research and education centre - Polish case," in *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, 2012, pp. 1–8.

[12] H. Vogt, H. Weiss, P. Spiess, and A. Karduck, "Market-based prosumer participation in the smart grid," in *4th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*. IEEE, 2010, pp. 592–597.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 683–685

Evolutionary Nonlinear Data Transformation for Visualization and Classification Tasks

Kamil Ząbkiewicz
Vilnius University, Institute of Mathematics and Informatics
Akademijos str. 4, LT-08663 Vilnius, Lithuania
Email: Kamil.Zabkiewicz@mii.vu.lt

Polish Academy of Sciences, Systems Research Institute
ul. Newelska 6, 01-447 Warsaw, Poland
Email: K.Zabkiewicz@ibspan.waw.pl

University of Bialystok, Faculty of Economics and Informatics
in Vilnius
Kalvariju str. 135, LT-08221 Vilnius, Lithuania
Email: K.Zabkiewicz@uwb.edu.pl

***Abstract*—In this paper we propose new approach in data set dimensionality reduction. We use classical principal component analysis transformation. Instead of rejecting features we generate new one by using nonlinear feature transformation. The values of transformation weights are changed evolutionary by using genetic algorithms. Results show better classification rates in smaller feature space. Visualization results also look better.**

I. Introduction

NOWADAYS data visualization is one of the important fields in knowledge discovery. Human beings can perceive data up to three dimensions. Although dimensionality of datasets is often enormous. To solve this issue, dimensionality reduction methods were proposed. There are two main directions: feature selection and feature extraction. The subject of this paper will be related only with second approach. In this work we propose new approach by extending functionality of PCA transformation.

The paper is organized as follows. In section II the most popular data visualization methods will be described. In section III proposed nonlinear data transformation method will be presented. In section IV the results of experiments with UCI repository data will be shown. Finally in the section V main conclusions and directions for future research will be presented.

II. Data visualization methods

There are many different methods to visualize multidimensional data. One of the oldest and popular also in nowadays is principal component analysis. First proposed by Karl Pearson in 1901. Good review of this method is given by Jollife [2]. Another data visualization technique is multidimensional scaling, which firstly was used in psychometry. It is based on minimization of squared differences between distances of points in original and reduced feature space. Overview of this method is given in [4].There is also neural network approach called self-oragnizing maps (SOM) introduced by Kohonen in [3]. Good overview of recent visualization techniques is also given in [1].

III. Proposed nonlinear transformation

Main idea of this work is extension of PCA transformation with additional information provided with new feature. Our goal is to encode rejected features after principal component transformation into one number and append it to reduced data set. Encoding scheme is based on principles of fundamental theorem of arithmetic, i.e. that every natural number can be written as a product of prime numbers rised to appropriate power. This product is unique

$$a = \prod_{i=1} p_i^{k_i},$$

where p_i is prime number and k_i is power. In this case powers are natural numbers (including zero). It can be extended to rational powers due to the same cardinality. In this case numbers are also unique. In this work we will simplify initial requirement by using rational fractions between 0 and 1 instead of prime numbers. These numbers will be called weights. The initial formula is now written in such way

$$a = \prod_{i=1} w_i^{x_i},$$

Due to the fact that we do not know values of weigths we have to solve optimization problem. We will use genetic algorithm to optimize weights. Our fitness function will be result of classification made with nearest neighbour classifier. The algorithm is presented on page 684.

Let us comment several moments of this algorithm. First we perform initial data normalization into interval [0,1]. It is done because of possible different scales of features. We have chosen the nearest neighbour classifier because of small number of

978-83-60810-53-8/$25.00

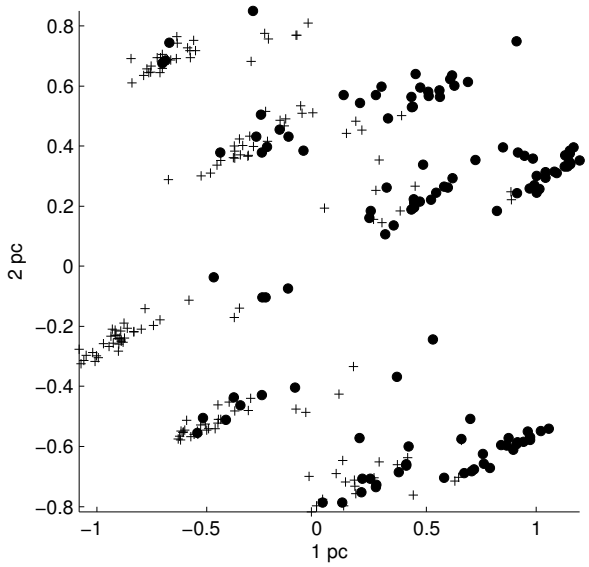
Algorithm 1 Classification and visualization using nonlinear evolutionary transformation

Require: $data$ - initial dataset, n_c - number of components for visualization, n_{iter} - number of iterations, n_{folds} - number of folds for crossvalidation, $n_{trfolds}$ - number of folds for internal crossvalidation
Ensure: $data_r$ - reduced transfomed dataset, w - vector of weights
normalize dataset features into interval [0,1] using min-max normalization
perform PCA transformation
split transformed dataset into n_{folds} folds
for $i = 1$ to n_{folds} **do**
  generate population of initial weights
  $iter = 1$ {current iteration}
  while $stopping\_criterium \neq$ true or $iter \leq n_{iter}$ **do**
    split training set into $n_{trfolds}$ folds
    for $j = 1$ to $n_{trfolds}$ **do**
      classify $fold_j$
      measure average classification error err_{iter}
    end for
    if $iter > 1$ **then**
      if $err_{iter} > err_{iter-1}$ **then**
        $stopping\_criterium \rightarrow$ true
      end if
    end if
    $iter \rightarrow iter + 1$
  end while
  perform evaluation on test fold tst_fold_i
end for
compute average algorithm performance $perf$
$w \leftarrow w_{best}$
$data_r \leftarrow$ transform$(data, w)$
return $perf, data_r$
visualize$(data_r)$

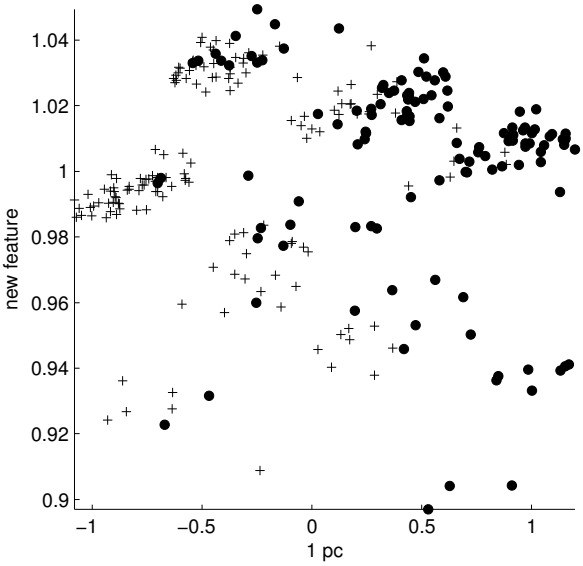
parameters (only number of neighbours and distance measure, in this case it is Euclidean distance). First tests of algorithm showed that proposed method has tendency to overfit the data. As a result new task appeared - to formulate early stop criterium. Brumen et al work [5] gave an idea how to solve this problem. In this work we use more simple criterium, i.e. "if error starts to grow break evolution process". To measure change of the error rate we use internal 5-fold crossvalidation. Only by using early stopping criterium we reduce number of iterations to perform. Internal crossvalidation provides also very important property, i.e. that classification error does not increase very rapidly. To prevent rapid increasing to infinity (decreasing to zero) of the generated feature value we let weights to vary only in interval $[1 - \varepsilon, 1 + \varepsilon]$, where $\varepsilon = \exp(1/number\_of\_rej\_comps) - 1$

IV. Experimental results

The experiments were performed on datasets from UCI Repository. These are: Glass, Teaching Assistant Evaluation, Heart and Chess. More details about each set is given in table I.



(a) Two principal components space



(b) First principal component and new feature space

Fig. 1. Heart dataset visualizations.

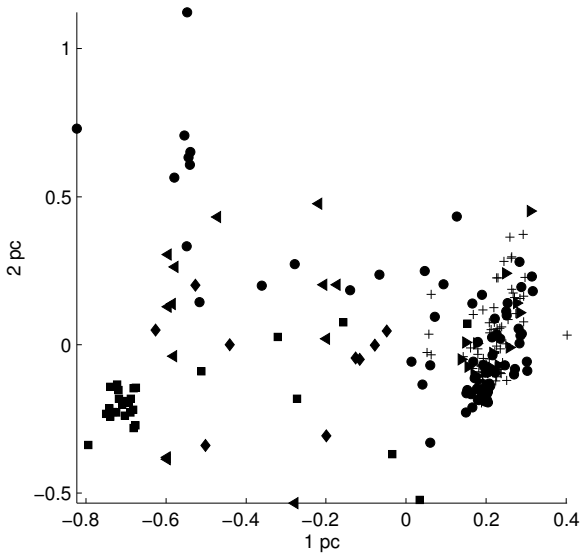
Datasets were chosen because the percent of total variance is not mostly covered by the several first principal components. The experiments were performed by reducing principal component space into 2 and 3 dimensions. Later the same experiments were made with the same datasets but with new additional feature appended. The number of features in both cases is equal. Results are shown in table II. We can notice that additional feature generated by nonlinear transformation in many cases gives better results.Visual analysis of datasets was also performed. Due to limited space we will show only several comparisons in two dimensional space. Let us look at plots of Heart dataset shown on figure 1. Classical PCA

TABLE I
DETAILS OF UCI DATASETS

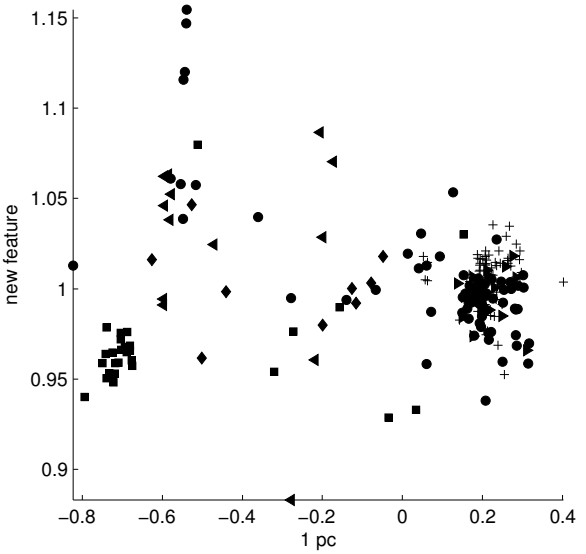
Dataset	Num. of feat.	Num. of classes	Num. of vectors
Glass	9	6	214
Teaching Assistant Evaluation	5	3	151
Heart	13	2	270
Chess	36	2	3196

TABLE II
UCI DATASETS CLASSIFICATION RESULTS

Dataset	1-NN				3-NN				5-NN			
	2 feat.		3 feat.		2 feat.		3 feat.		2 feat.		3 feat.	
	with	without	with	without	with	without	with	without	with	without	with	without
Glass	**0.5882**	0.5557	**0.5795**	0.5364	**0.5573**	0.5506	0.5672	**0.5688**	0.6053	**0.6136**	**0.6115**	0.6017
Teaching Assistant Evaluation	**0.6394**	0.6179	0.5763	**0.5911**	**0.5114**	0.4518	0.5143	**0.5241**	**0.5274**	0.4946	**0.4725**	0.4625
Heart	**0.7385**	0.7333	0.7330	**0.7481**	**0.7874**	0.7519	**0.7670**	0.7444	**0.8078**	0.7963	**0.8070**	0.7852
Chess	**0.6928**	0.5942	0.7020	**0.7030**	**0.6815**	0.6142	**0.7331**	0.6609	**0.6918**	0.6230	**0.7379**	0.6559



(a) Two principal components space



(b) First principal component and new feature space

Fig. 2. Glass dataset visualizations.

shows that classes are scattered, our approach on the other hand makes points of one class to be concentrated in certain part of space. Now let us analyze vizual data of Glass dataset. Both cases are shown on figure 2. As we can see objects of some classes are groupped more closer to each other. As a result using e.g. k-nearest neighbour classifier we can obtain better classification rates.

V. CONCLUSIONS

Proposed approach extends popular and quite simple data transformation. The computational cost of new feature is not that large as e.g. computing pairwise distance matrix (in case of multidimensional scaling) As it has been presented earlier proposed transformation made positive influence on analyzed data sets. First it increased classification rates with small number of components. In next step it provided more clear visualization results. Unfortunately this approach has one minus - it can be used only with numerical features. Of course we can encode categorical features into numbers, but this method is rather artificial and classification results can depend on way how features were encoded. The problem of big data sets is also actual. In this work small datasets were rather used. In future the proposed technique will be also tested on larger ones, e.g. KDD Cup 1999 data set. For solving such type of problem we will make some optimizations. One of them could be use of architectures that are based on parallel computing paradigm, such as clusters or GPUs.

REFERENCES

[1] G. Dzemyda, O. Kurasova and J. Zilinskas, *Multidimensional Data Visualization,* Springer, 2013.
[2] I. Jolliffe, *Principal Component Analysis, Second Edition,* Springer-Verlag New York, 2002.
[3] T. Kohonen, *Self-Organizing Maps, Third Edition,* Springer-Verlag, 2001.
[4] I. Borg, P. Groenen, *Modern Multidimensional Scaling: theory and applications (2nd ed.),* Springer-Verlag, New York, 2005.
[5] B. Brumen et al., "Learning process termination criteria," Informatica, vol. 23, No. 4, pp. 521-536, Vilnius University, 2012.

Information Systems Education & Curricula Workshop

ISEC goal is to promote the discussion about the convergence between Computer Science and Information Systems topics, so that researchers can present a complete and detailed specification of their educational curricula by means of these two topics. We inspect papers that contribute to the better understanding of emerging and important educational fields of Computer Science (CS) and Information Systems (IS). Authors are invited to submit their papers in English, presenting the results of original research or innovative practical applications in the field.

Regarding the selection process, we plan to perform a triple review process.

Topics

Key issues in this workshop will focus on (but are not limited to):

- Convergence between IS & CS in higher Education
- Specification of IS Education Curricula
- General IS Theory
- IS Scope
- IS Educational Fields
- Student participation in research
- Adaptation to the European Higher Education
- Assessment of students
- Innovative teaching methods
- Training for career and skills development
- Definition of "knowledge" in IS
- Quality and evaluation of teaching
- Social and environmental commitment
- Curriculum organization and curriculum
- The Fundamental Concepts Underpinning IS
- Merge from IS to CS and vice versa in higher Education

Event Chairs

Fardoun, Habib M., King Abdulazziz University, Saudi Arabia

Gallud, José A., University of Castilla-La Mancha, Spain

Tesoriero, Ricardo, University of Castilla-La Mancha, Spain

Program Committee

Abou-Tair, Dhiah el Diehn, German Jordanian University, Jordan

Aknin, Noura, Université Abdelmalek Essaadi, Morocco

De la Guía, Elena, University of Castilla-La Mancha, Spain

Garrido, Juan Enrique, University of Castilla-La Mancha, Spain

Giménez, Rafael, Barcelona Digital Technology Centre, Spain

Kempin, Nils, CGI, Germany

Majchrzak, Tim A., University of Münster, Germany

Mystakidis, Stylianos, University of Patras, UOC, UW, Greece

Tambo, Erick, United Nations University, Germany

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 689–693

Towards improved student placement and preparation methods on Information Technologies post-secondary education

Jaime Ramírez Castillo
ITKnowingness, Spain
Email: jaime.ram@gmail.com

Aldabbagh Ghadah , Habib M. Fardoun
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
Email: {galdabbagh,hfardoun}@kau.edu.sa}

***Abstract*—In this article we present the results of a pilot programme for student placement on the university, comprised of a preparation course, the GCE Ordinary level test in Computing, which students perform afterwards as a placement test, and a post-course questionnaire. The aim of the research programme is to identify weaknesses in the student placement tests and set the road map for improved first-time entrant placement and preparation. The study shows that common placement tests are far from giving strong predictions and they should be complemented with other metrics, such as high school grades or social factors. Test outcomes show that placement test results per se do not yield enough data to predict student success. However, we discovered it as a quite helpful tool for revealing anomalies at the institutional and methodological level, such as very different outcomes among campuses of the same university, or remarkable difficulty to answer certain questions. In order to enhance student placement accuracy and preparation for university, these issues will need to be addressed in forthcoming research.**

I. Introduction

UNIVERSITIES use placements tests to assess the students' academic abilities. In theory, test results give enough information about alumna skills to place them in the right level and determine which classes are suitable for each student. However, studies seem to indicate that accurate student placement is problematic [1].

At King Abdulaziz University we have started research in this area by requiring a set of first-time entrants to enrol a pilot programme for students placement, including the CPIT100 preparation course, the GCE Ordinary level test in Computing, which students perform after the course as a placement test, and a post-course questionnaire. Placement tests are broadly used to determine the level at which each student should be placed. However, studies agree in the fact that such exams perform poorly at proper student level placement. Therefore, we have carried out a review of studies about placement tests reliability, which is show later in the article.

The objective is to review placement tests as an instrument to predict student performance and to assess the outcomes to identify issues that need to be addressed and set the basis for the development of improved placement mechanisms.

The following section shows statistical data about study subjects (students). Section III describes instruments that have been used for the experiment, namely the course, the test and the questionnaire. Section IV describes the procedure for placement and preparation, which is basically taking the course, then the test and eventually filling in the questionnaire. In Section V, statistical data are extracted from test and questionnaire outcomes for analysis. Finally Section VI concludes with discussion of achieved results, studies reviews, discussion and requirements for further research.

II. Subjects

A total number of 2685 students distributed among three campuses took the course, specifically 1083 from Alsulaimanyah campus, 992 from Alsharafiyah campus and 610 from Alsalamah campus. Students also came from different educational backgrounds; 1387 of them came from a scientific scope, 1283 from administrative scope and 15 were previous regular students.

Finally, students were divided in two categories for the questionnaire: regular and distance students.

III. Instruments

A. GCE Ordinary Level in Computing

GCE Ordinary level in Computing is a qualification created by ©Pearson Education to encourage candidates to develop an understanding of computer systems, software, data and hardware and their implications for communications and people. It also aims to help students acquire necessary skills to apply computer-based solutions to problems. Candidates who successfully follow the syllabus will have a good practical understanding of computing and its applications. Namely, they will: develop an understanding of the main principles of using computers to solve problems; appreciate the range of applications of computers and the effects of their use; understand the organizations of computer systems, software, hardware and data and their implications; acquire the skill necessary to apply computer-based solutions to problems[2].

We considered that meeting the objectives of GCE Ordinary Level in Computing Test requires a skill level

978-83-60810-53-8/$25.00

suitable for students beginning post-secondary education. Therefore, we decided to use materials provided for the qualification as tools to prepare first-time entrants and also as a possible way to assess their knowledge and predict at which level should students be placed at, once the university courses start. This means not entirely leveraging GCE as at qualification itself, but as a placement test that would allow predicting students performance. Placement tests are discussed below.

B. CPIT100 Preparatory Course

CPIT100 is the preparatory course developed by King Abdulaziz University to help the students to achieve the assessment objectives of GCE Ordinary Level in Computing.

The course aims to provide the students with advanced skills to operate and make use of a personal computer in different environments such as in academia, in business, and at home. It introduces the students to the main concepts and terminologies of information technology, and equipped them with the knowledge to administer one of widely/used operating systems. The course also aims to provide the students with the practical skills to utilize an office productivity package for different purposes. The course will prepare the students for new learning methodologies, namely distance learning and e/learning. The delivery of the course contents will be based on a hands/on approach.

Apart from preparing students from the GCE placement test, our purpose was to reinforce computing and information technology skills of first-time entrants to the university. For skilled students, the course was expected to refresh and stimulate their capabilities. Nevertheless, underprepared students were the key, since the course was specially intended to bring them to the adequate level for post-secondary education. We considered the course as a mixture of a qualification, placement and remedial course.

Remedial education is a kind of teaching that is bellow university level work. It is designed to bring unprepared students to the level expected of entrants to the university. This kind of education has received criticism due to the “double billing” problem, which for unprepared students means to spend more than the amount needed for students that perform successfully by just following regular courses [3].

Although remedial/preparatory education is a controversial issue because of its expenditures, supporting arguments present remedial education spending as an investment. The hypothesis made here is that, in the long term, educating underprepared entrants will decrease the likelihood of their future dependency on social programs and as students that eventually obtain the degree, they would potentially contribute to the state taxes in the future [4].

C. Placement Tests

Universities use placements tests to assess the academic abilities of the students. In theory, the results of these tests give enough information about students to place them in the right level. However, studies seem to indicate that the accurate student placement is problematic. Many students are misplaced in remedial courses or wrong levels where they should be not in.

Some studies indicate that any model that uses final course grade as the criterion for the validity of a placement rule would likely fail [5].

The nature and validation of placement tests is rarely discussed in the language testing literature, yet placement tests are probably one of the commonest forms of tests used within institutions which are not designed by individual teachers and which are used to make decisions across the institution rather than within individual classes[1].

The predictive power of placement exams is still quite good given how short they are. But overall the correlation between scores and later course outcomes is relatively weak, especially in light of the high stakes to which they are attached. Given that students ultimately succeed or fail in college-level courses for many reasons beyond just their performance on placement exams, it is questionable whether their use as the only determinant of placement can be justified on the basis of anything other than consistency and efficiency. Allowing more students directly into college-level coursework (but perhaps offering different sections of college-level courses, some of which might include supplementary instruction or extra tutoring), could increase the numbers of students who complete university-level coursework in the first semester, even if pass rates in those courses decline [6].

Summing up, testing student abilities needs to be complemented with other several indicators such as high school grades and good teaching standards. A more suitable metric seems to be the combination of placement tests with other metrics such as high school grades, social factors, years since graduation in high schools, etc.[7],[8].

D. Questionnaire

At the end of the course, students were asked to fill in a questionnaire were they were asked a series of questions about their daily use of new technologies and their level of satisfaction with the course to evaluate teaching quality. Since course experience questionnaires have proven to be a useful tool to evaluate teaching performance [9], we included it to help us improve future courses and also to get an indicator of the access students have to new technologies, which would give as an idea of whether they need more equipment at the campuses or they can carry out some work at home.

IV. PROCEDURE

Before the academic course started, CPIT100 course was imparted to first-time entrants as preparation teaching for the GCE Ordinary Level in computing test. Individuals taking the course were classified by campus (Alsulamaniyah, Alsharafiyah and Alsalamah), by educational background (scientific, administrative or regular previous students) and by delivery mode (distance/external and regular students).

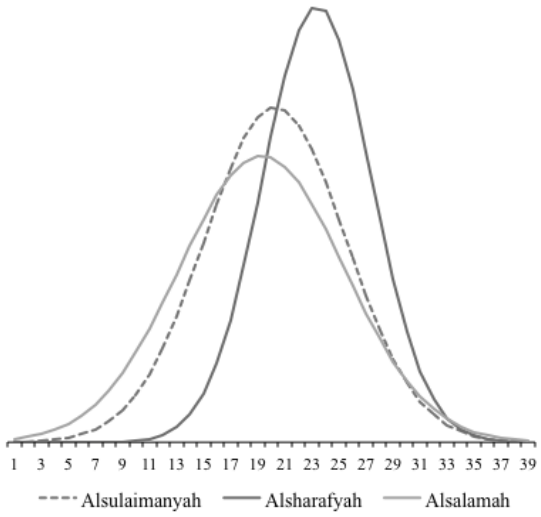


Fig 1 *Average score by campus*

After test administration, results were collected and analysed with Qualtrics research software [10]. We monitored mean score and number of correct answers by campus to assess how they perform individually.

Eventually, the post-course questionnaire was delivered to students to collect information about their personal opinion and their access to new technologies.

V. Results

Test outcomes data are categorized into three campuses, Alsulamaniyah, Alsharafiyah and Alsalamah. Fig 1 shows the normal distribution of scores by campus. Alsharafyah campus clearly outperforms the other two with an average score of 17.31 and a smaller deviation.

Alsulaimanyah campus achieved an average score of 14.23 and Alsalamah comes right after with an average score

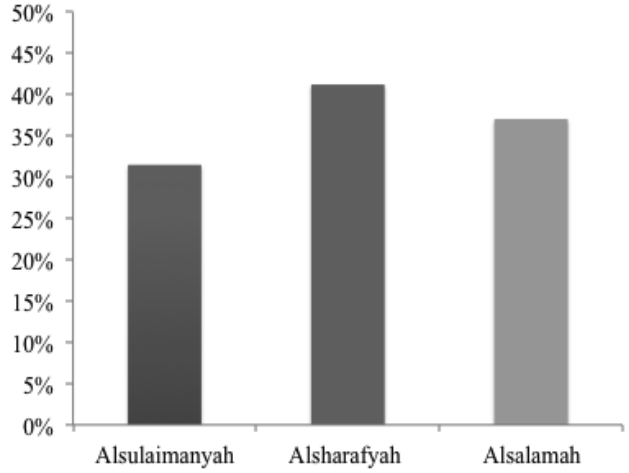


Fig 2 *Percentage of correct answers by campus*

of 13,23, although the latter shows more dispersion. Having such a different average score among campuses from the same university reveals an anomaly that we need to pay attention to.

Another interesting metric we monitored is the number of correct answers, shown in Fig 3, also distributed by campus and showing the average of all students as well.

Section A (questions 1 to 7) shows higher scores than section B (from question 8 on). Clearly average score goes down when the section B starts with question 8. Lower scores in section B might be caused by either by an increased difficulty or by different question type; while section A contains short questions, section B is comprised by a set of longer questions based on a case of study.

Some questions have been especially difficult for students. For instance, question 2.3 "*Data entry clerks use a keyboard to enter text into a computer. Another method of entering text into a computer is.*" only shows a correct answer percentage of 11%. Few students were able to state alternative text entering methods.

The figure also shows Alsharafyah campus performs better than the two other campuses, regardless of the

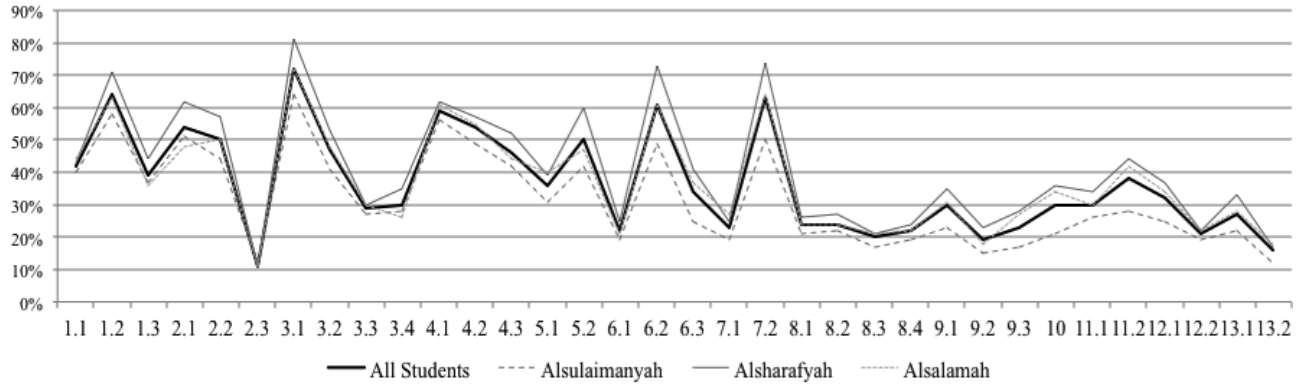


Fig 3 *Average score by question for each campus*

question. On the contrary, Alsulaimanyah campus gets the lowest score in almost all the questions. This can also be seen in Fig 2

With regard to the questionnaire, students were divided in three groups: scientific, administrative, and previous regular students. The questionnaire evaluates the following questions: use of computers in the student's daily life, student's expectation fulfilment, course outcomes achievements, students' command of MS Windows, student's command of MS Office. The latter three ask the students whether they think they would be capable of resolving a problem or they would need to ask for help.

Near 100% of students are used to perform tasks with their computers at home and at the school. Also almost all of the students have Internet connection in their houses.

With regard to course outcomes and aims, regular students do not show much confidence. Despite of the fact that 57% got the basic concepts and knowledge, only 21% acquired system administration and maintenance concepts, such as downloading and properly installing software without help, and only 22% thought he or she had fully understood computer programming.

Distance and external students show similar results, although performing a bit better on system administration concepts (27%) and not doing so well in basic knowledge (51%).

Course expectations were met for the 44% of the students, partly for the 48% and not met for the 8%.

In general, command of MS Windows shows good results. Regular students perform better than external & distance students. Basic Windows usage and tasks are well performed by more than 80% of the users, except for formatting the hard-drive and re-installing the operating system, which only was executed by a 48% of the students without help.

Microsoft Office also shows good results but not as high as Windows results. Easy tasks, such as writing an essay and using tables, charts and pictures are executed without help by more than 90% of the students. Working with analytical data and preparing presentations is done without help by more than 73% of the students. Finally, the hardest tasks for the students where database management, arranging appointments for meetings, create email groups.

Lastly, Fig 4 shows a slightly lower percentage of positive indicators in external and distance students, which can occur because of a lower engagement factor.

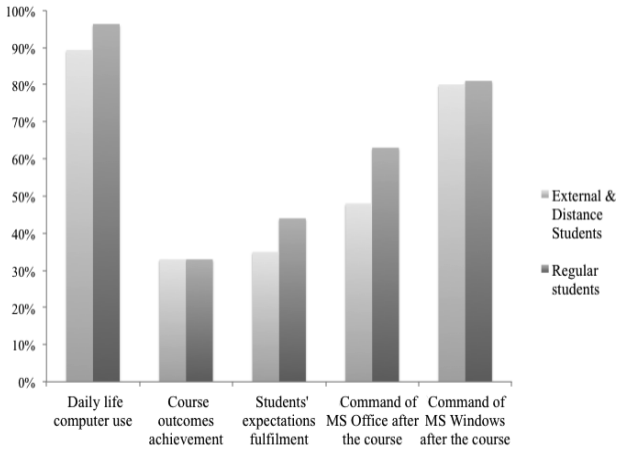


Fig 4 Final positive indicators

VI. CONCLUSIONS

In this paper we described how we employed the GCE Ordinary level in computing qualification as a placement test. Also, a questionnaire has been given to the students to gather information about their daily use of new technologies and opinion about the process.

First question to address is why does one campus clearly performs better than others. It is important to perform a more exhaustive study on students' skills for each campus. High school grades need to be part of the equation to effectively know if different campuses are receiving students with different levels. Social environment data would be also helpful to adapt campuses pedagogical methods if they are receiving students from areas with lower high school grades.

Moreover, Pearson GCE Ordinary level certificates are being replaced by the new International GCSEs, available from 2009. The equivalent for the old Computing certificate is the GCSE in Information and Communication Technology, updated to the requirements of modern society. In the future we plan to adapt the CPIT100 course new GCSE qualification.

With regard to placement test predictions, they need to be complemented with secondary education grades in subjects related to computing (namely, math, physics and computing) and students commitment during the preparatory course. Again, having information about the social environment will be important, especially to get more information about students daily use of information technologies. We must perform research on how to extract this data from the questionnaire and act accordingly. Thus, with such improved indicators we could know the kind and amount of homework that is given to students, based on the facilities they have at home.

As for level prediction rates we ought to assess the prediction rate of our approach by performing research on how students perform in the tasks that have been covered in the course. A good measure would be to correlate tests data with students skills on the specific topics covered during the course. However, before carrying out this assessment better student classification is needed for future exams, namely, get statistical data for scientific and administrative background students.

Finally, our GCE tests should be compared with results from other institutions that have also used the same tests. The comparison should be accompanied with social, environmental and educational information about the students of both institutions and consequently draw the correct conclusions. The main aim of this is to share knowledge and be aware of the areas we need to improve.

REFERENCES

[1] Dianne Wall, Caroline Clapham, and J. Charles Alderson. Evaluating a placement test. Language Testing November 1994 11: 321-344, doi:10.1177/026553229401100305.

[2] GCE O level in Computing. EdExcel (Pearson). http://www.edexcel.com/quals/olevel/7105/Pages/default.aspx

[3] Jamie P. Merisotis, Ronald A. Phipps. Remedial Education in Colleges and Universities: What's Really Going On? The Review of Higher Education Volume 24, Number 1, Fall 2000.

[4] Saxon, D. P., & Boylan, H. R. (2001). The cost of remedial education in higher education. *Journal of Developmental Education*, *25*(2), 2-9.

[5] Armstrong, William B. Validating Placement Tests in the Community College: The Role of Test Scores, Biographical Data, and GradingVariation. Annual Forum of the Association for Institutional Research (35th, Boston, MA, May 28-31, 1995).

[6] Scott-Clayton, J 2012, Do high-stakes placement exams predict college success? CCRC working paper no. 41, CCRC, New York.

[7] Belfield, C., & Crosta, P. (2012). Predicting success in college: The importance of placement tests and high school transcripts. Working paper no. 42. New York, NY: Community College Research Center, Teachers College, Columbia University.

[8] High-Stakes Placement Exams Predict College Success?" and "Predicting Success in College: The Importance of Placement Tests and High School Transcripts". ADAMS, CARALEE // Education Week; 3/14/2012, Vol. 31 Issue 24, p5.

[9] The development, validation and application of the Course Experience Questionnaire. Keithia L. Wilson, Alf Lizzio, Paul Ramsden Studies in Higher Education Vol. 22, Iss. 1, 1997.

[10] Qualtrics home page. http://www.qualtrics.com/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 695–701

Reduction of the Students´ Evaluation of Education Quality Questionnaire

Montserrat Corbalan
R&D&I EduQTech group. Polytechnic University of Catalonia (UPC), Terrassa 08222, Spain)
Email: montserrat.corbalan@upc.edu

Inmaculada Plaza
R&D&I EduQTech group University of Zaragoza, Teruel 44003, Spain
Email: inmap@unizar.es

Eva Hervas
R&D&I EduQTech group University of Zaragoza, Teruel 44003, Spain
Email: inmap@unizar.es

Emiliano Aldabas-Jordi Zaragoza
R&D&I EduQTech group Polytechnic University of Catalonia (UPC), Terrassa 08222, Spain
Email: emiliano.aldabas@u

Francisco Arcega
R&D&I EduQTech group University of Zaragoza, Zaragoza 55018, Spain
Email: arcegafj@unizar.es

***Abstract*—Assessment of students and the evaluation of their satisfaction has been an important element in the improvement of teaching quality in all the Higher Education areas. Specifically the student participation in Computer Science (CS) and Information System (IS) has been highlight valued. Thus a large number of methodologies and standard tools regarding student evaluation has been developed. Specifically, the Students´ Evaluation of Education Quality (SEEQ) is a tool that is validated for international use. But its use leads to several problems, such as the low voluntary participation of students. To solve these problems, a short version of this questionnaire developed using statistic tools is proposed. After using the proposed new version, the voluntary participation of students increased. The reduction of the number of questions facilitates the analysis of data, improving the flow of information and feedback between professors and students.**

I. Introduction: Quality in Higher Education

CONCEPTS of quality from the entrepreneurial world are increasingly being incorporated into the university field [1]. According to the "Declaration of Prague" (2001), quality should be an important and determinant aspect of Europe´s international attractiveness and competitiveness [2]. In 2003 in Berlin (Berlin, 2003), the Ministers responsible for Higher Education stressed that "the quality of Higher Education has proven to be the heart of the setting up of a European Higher Education Area" [2].

When these ideas have been translated into action, the "teaching quality" concept has become outstanding. According to Kember, "it might be noted that concern about Teaching Quality is growing at the national level. This appears to be a worldwide phenomenon" [3]. In 1992, Stones pointed out [4] that "quality teaching is more properly conceived of as a unified field embracing both theory and practice in which teachers, teacher educators and researchers are jointly responsible for the development of theoretical understanding and the improvement of teaching".

Currently, there remains a high degree of concern about the improvement of teaching quality at the higher education level. Specifically, the improvement of teaching quality is one of the primary matters that must be addressed continuously by universities. The question is how to achieve this continuous improvement.

Continuous improvement is not a tool or a technique but rather a way of life (or at least a cultural approach to quality improvement) [5]. According to the UNE 66178:20004 model [6], there are three steps that should be taken into account during an improvement process:

1. analysis of the information for the improvement;
2. the improvement project; and
3. monitoring, evaluating and reviewing the improvement.

One method of attaining continuous improvement is the PDCA Cycle (also known as the Shewhart or Deming cycle: Plan-Do-Check-Act) [7], which emphasizes the continuous, never-ending nature of process improvement. The PDCA Cycle highlights and demonstrates that improvement programs must start with careful planning, result in effective action, and move back to careful planning in a continuous cycle. Figure 1 shows this global idea.

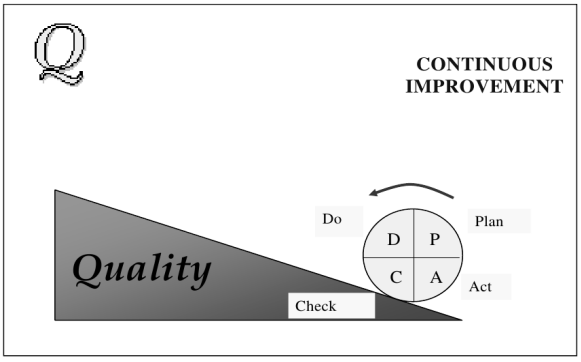


Fig. 1 The PDCA Cycle as a tool to continuous improvement [8], [9]

Clearly, it is necessary to collect and evaluate data in order to obtain conclusions. Improvement actions should be based on the data and the conclusions of the evaluation.

This point of view is completely applicable to the university environment, especially to lecture theatres. To improve the teaching quality, it is necessary to evaluate the teaching process and its results. In this evaluation, the students' opinions about learning and the teaching process are crucial due

to the students' roles as the primary consumers in higher education [10].

This paper focuses on students' evaluations of the quality of higher education. The extensive body of research regarding student evaluation of teaching leads us to look for standard tools and methodologies (section 2). Section 3 details a widespread method that can be used internationally, the Students' Evaluation of Educational Quality SEEQ) questionnaire, and considers the problems that teachers identified after its application. To solve these problems, section IV presents a proposal for the reduction of this questionnaire. Sections V and VI show an example of its application and give several guidelines regarding the use of this short version of the SEEQ questionnaire. The final conclusions are explained in section VII.

II. Review of the Standards

Spooren and Mortelmans underlined the value of student evaluations of teaching. They found that students reward good teachers with higher ratings on several scales of teacher performance [11]. The literature contains an overwhelming number of data collection instruments and scales. Several authors chose to develop their own form (see, for instance, [12] or [13]). In some cases, the forms were developed by faculty committees [14]. In general, there is an extensive body of research regarding student evaluation of teaching and how students contribute to assessments of teaching effectiveness [15], [16]. Thus, it is possible to conclude that there is a need to unify and standardize the approaches and specific tools used to evaluate the teaching quality.

Standards are public technical documents that establish common terminology in a field (in this paper, quality). They set specifications extracted from experience, knowledge, and available technology [17], [18] and [19].

There are several international and national standards developed for the teaching field that are applicable to this work [20].

* UNE 66931 is a Spanish standard that aims to provide guidelines for the application of the ISO 9001 model in educational organizations. It is equivalent to the document IWA 2 published by ISO [21].

Section 8.2 (concerning monitoring and measurement) points out that the educational organization must have reliable methods to measure and control the satisfaction of the client (8.2.1.). Moreover, the educational organization should define and use methods to monitor the results of the educational product (8.2.4). Furthermore, the customer and stakeholder satisfaction surveys are described as important data (section 8.4).

* The ISO/IEC 19796 family, focused on the information technology field, is an international standard under the general title "Information technology – Learning, education and training – Quality management, assurance and metrics" [22] and [23]. This family is a framework used to describe, compare, analyse, and implement quality management and quality assurance approaches. It will serve to compare different existing approaches and to harmonise these approaches towards a common quality model. [17] [22]. The ISO/IEC 19796-3 [23] is an instrument for the implementation and adaptation of the first quality standard ISO/IEC 19796-1 and, in particular, for the specification of the individual process descriptions [24].

According to Campo [25], this family has been defined abstractly and without specific guidelines to provide a mechanism to its implementation.

Thus, it is possible to conclude that international and national standards specially developed for the teaching field provide professionals with frameworks and guidelines to improve the quality of education. However, these standards do not provide specific tools or mechanisms to evaluate teaching quality.

In addition, other models and standards regarding quality that are widely used in the entrepreneurial environment can be applied to lecture theatres. The use of these standards helps institutions respond to the EHEA (European Higher Education Area) quality requirements [26]. In the present case, two standards could be applicable: UNE 66178:2004 and UNE 66176:2005.

* UNE 66178:2004

The standard entitled "Quality management systems. Guide for the management of process for improvement" is focused on continuous improvement. This standard is based on the idea that every organization needs to improve. Its capability to satisfy the requirements of its stakeholders (such as the customers, staff and social environment) determines the survival of the organization. Furthermore, these needs are changeable [27]. This point of view is completely applicable to the university environment, especially to lecture theatres [26]. In Appendix A, teachers can find a list of techniques and tools that can be used in the improvement process. The cycle of Deming is listed in this Appendix.

* UNE 66176:2005

This Spanish standard is titled "Quality management systems. Guide for measuring, monitoring and analyzing customer satisfaction". According to its title, this standard specifies guidelines for the definition and development of a measuring process for customer satisfaction [6]. Its guidelines are generic and can be applicable to any organisation, regardless of its size or activity. Of particular note is Table 1 of the standard, which contains different techniques for data collection and indicates their advantages and disadvantages. Professors interested in its application will find appendices A to E significant [26].

The analysis of standards UNE 66176 and UNE 66178 produces the same results as the previous models (UNE 66931 and ISO 19796): they provide professionals with frameworks and guidelines but do not provide specific tools or mechanisms to control the satisfaction of students.

The global conclusion of this section is that there is a high number of collection instruments for obtaining data about student satisfaction of teaching quality. Thus, it is necessary to look for other standard tools. The review of international and national standards can provide guidelines and global methodologies. Their application will allow for the extraction of ideas for improving traditional working

methods at the university level [26]. However, to apply the previously cited standards, professors who wish to use them must carry out a tailoring process. Thus, the standards do not provide any specific tool to evaluate the students' opinions about teaching quality.

III. Description of the SEEQ

The lack of a standardised tool to evaluate students' opinions about teaching quality leads us to look for a widespread method that can be applied internationally.

In terms of the more formal, internationally validated questionnaires in use in higher education, five are particularly worth mentioning [28]:

- The Students' Evaluation of Educational Quality (SEEQ),
- The Course Experience Questionnaire (CEQ),
- The Module Experience Questionnaire (MEQ),
- The Postgraduate Research Experience Questionnaire (PREQ), and
- The Experiences of Teaching and Learning Questionnaire (ETLQ).

Richardson recommended using either the SEEQ or the CEQ, as both have been validated for international use through research studies [29]. Keane pointed out that the SEEQ and MEQ have potential because they have been statistically validated [28].

In the present study, the SEEQ questionnaire was chosen. There are several reasons for this choice: a robust factor structure, excellent reliability, and reasonable validity [30]. Furthermore, as has been mentioned, the SEEQ has been validated for use internationally. It is possible to find universities in many different countries that use the SEEQ. Although not exhaustive, the following list presents several examples: the Universities of Manitoba [31], Saint Mary's [32] Mount Allison [33] and Saskatchewan [34] in Canada; Fordham University [35] and the Schreyer Institute for Teaching Excellence [36] in the U.S.; Oxford University [37] and University of Leicester [38] in the U.K.; Semnan University [39] in Iran; Curtin University [40] in Australia; and the Universities of Navarra [41] and Vigo [42] and the Polytechnic University of Catalonia [43] in Spain. This international use will enable the development of comparative analyses in the future.

The SEEQ was developed by Dr. Herbert Marsh of the University of Western Sydney [44]. Dr. Marsh is an internationally recognised expert in the area of psychometrics. Now in the public domain, the SEEQ has been extensively tested and used in more than 50,000 courses with over one million students at both the graduate and undergraduate levels [33].

Using a five-point scale, the SEEQ questionnaire examines different characteristics of effective teaching. Each of these categories contains three or four questions.

It is possible to find different versions of the SEEQ [45], [46]. The version used in this work consist of 37 questions. It is detailed in references [43] [33].

* The questionnaire finishes with three open questions.

This SEEQ questionnaire was used in seven subjects in three cities and three different centres: Polytechnic School of Teruel (Spain), School of Engineering of Terrassa and Faculty of Engineering and Architecture of Zaragoza (University of Zaragoza (Spain).

The use of the SEEQ enabled professors to identify strengths and weaknesses and to improve the teaching-learning process. However, the main problem was the low participation of students: a high number of students did not answer the questionnaire or answered only the first questions on the form [47]. Moreover, Verdugo and Cal remarked upon the fact that rapid feedback is needed. These authors explained that teachers could be overloaded with work and not provide rapid feedback [46]. In order to solve these problems, this paper proposes a reduction of the SEEQ questionnaire developed in collaboration with a student [48] in order to encourage the students' research.

IV. Proposal of Reduction

To maintain the reliability of the short SEEQ questionnaire, statistical parameters should be used. Computing tools can help researchers in this process. In this work, analyses have been developed with SPSS® [49]. Specifically, this statistical package enables the calculation of the Cronbach's Alpha and Pearson's r parameters [50]. The values of Cronbach's Alpha and Pearson's r Correlation Coefficient for the long SEEQ questionnaire (37 questions-items) were calculated. To develop the analysis, a sample of 111 polls was used and items with different scales were recoded [48]. Table I shows the value of the global Cronbach´s Alpha.

TABLE I.
STATISTICAL VALUES OF RELIABILITY FOR THE LONG SEEQ QUESTIONAIRE (37 ITEMS)

Number of Items	Cronbach's Alpha
37	0.920

The Cronbach's Alpha can be used in the reduction process, as Cronbach's Alpha is an index of reliability.

According to Santos [51], this coefficient ranges in value from 0 to 1 and may be used to describe multi-point formatted questionnaires or scales (i.e., in this rating scale, 1 = poor and 5 = excellent). The higher the score is, the more reliable the generated scale. Nunnaly [52] (cited by Santos) indicated 0.7 to be an acceptable reliability coefficient, but lower thresholds are sometimes used in the literature [51]. Table II show the same data after the reduction process, using Cronbach's Alpha as reduction criteria.

TABLE II.
STATISTICAL VALUES OF RELIABILITY FOR THE SHORT SEEQ QUESTIONAIRE (22 ITEMS)

Reduction criteria: Cronbach's Alpha

Number of Items	Cronbach's Alpha
22	0.936

The global reliability for the initial long SEEQ questionnaire was 0.920, and the value for the short SEEQ is 0.936.

Thus, there is an improvement of the internal consistency of the questionnaire.

Several statistics experts recommend developing the reduction process using two criteria: Cronbach's Alpha and Pearson's r Correlation Coefficient. Although both analyses tend to give similar results, the combination is useful because it provides more information with which to make decisions [50].

With the initial data, a new reduction process was developed using the correlation as new criteria (third column in our tables). Table III shows the statistical results after completing the process.

TABLE III.
STATISTICAL VALUES OF RELIABILITY FOR THE SHORT SEEQ QUESTIONAIRE (22 ITEMS)

Number of Items	Cronbach's Alpha
22	0.936

Both analyses lead to the same results:

- the fifteen deleted items are the same;
- there are no negative correlations; and
- the final value of reliability (global Cronbach´s Alpha) is the same, 0.936, and there is an improvement in the questionnaire's internal consistency.

The short version of the SEEQ questionnaire proposed in this work is detailed below (the item number from the long version is in brackets).

*** Learning.**

1 (1) - I find the course intellectually challenging and stimulating.

2 (2) - I have learned something that I consider valuable.

3 (3) - My interest in the subject has increased as a consequence of this course.

*** Enthusiasm**

4 (5) – The instructor is enthusiastic about teaching the course.

5 (6) – The instructor is dynamic and energetic in conducting the course.

6 (7) – The instructor enhances presentations with the use of humour.

7 (8) – The instructor's style of presentation holds your interest during class.

*** Organisation**

8 (9) – The instructor's explanations are clear.

9 (10) – The course materials are well prepared and carefully explained.

10 (12) – The instructor gives lectures that facilitate taking notes.

*** Group Interaction**

11 (13) – Students are encouraged to participate in class discussions.

12 (14) – Students are invited to share their ideas and knowledge.

13 (15) – Students are encouraged to ask questions and are given meaningful answers.

14 (16) – Students are encouraged to express their own ideas and/or question the instructor.

*** Individual Rapport**

15 (17) – The instructor is friendly towards individual students.

16 (18) – The instructor makes students feel welcome in seeking help/advice in or outside of class.

17 (19) – The instructor has a genuine interest in individual students.

*** Breadth**

18 (21) – The instructor contrasts the implications of various theories.

19 (22) – The instructor presents the background or origin of ideas/concepts developed in class.

20 (23) – The instructor presents points of view other than his/her own when appropriate.

21 (24) – The instructor adequately discusses current developments in the field.

*** Examinations**

22 (27) – The examinations/graded materials test the course content that is emphasized by the instructor.

In conclusion, fifteen questions, approximately 40%, have been eliminated with this method.

V. EXAMPLE OF APPLICATION

The new short version of the SEEQ questionnaire was used in different university subjects. For instance, in Circuits and Electric Drives the previous academic year, students filled out the long version of the SEEQ questionnaire. As the teacher is the same, a comparison is possible. Table VI shows the level of student participation (number of completed questionnaires versus number of registered students). The percentage of voluntary involved students has increased.

In table IV, the items corresponding to the four lowest-scoring questions from both versions are listed. In this case, two items are equal. In addition, higher-valued items are listed, and there are three equal items.

TABLE IV.
PARTICIPATION LEVEL

Version	% Participation
Long	53.16%
Short	69.23%

Thus, the SEEQ questionnaire helps professors to detect weaknesses and strengths. These results can be used as the starting point of the following improvement plan.

The proposed short version of the SEEQ questionnaire has been used in other subjects in which the teachers were not the same as those in previous years. Nevertheless, in this academic year, the long version was used voluntarily in three subjects. The average percentage of participation was 69.41%. In addition, the short version was used voluntarily in three other subjects, with an average percentage of participation of 86.84%.

The conclusion is clear: the use of the proposed version increases the voluntary participation of students.

VI. Guidelines for the use of the proposed short version

Several guidelines can help professors use the proposed short version of the SEEQ questionnaire.

- As explained in the introduction, the SEEQ questionnaire can be used as a tool to develop improvement processes.
  Professors can distribute the questionnaire in the middle of the academic year.
- It is advisable to give the questionnaire to the teachers before the analysis of the data [53], [47].
- After the analysis of the data, instructors should identify the lowest- and highest-scoring questions on the questionnaire. These results can help instructors develop a plan to improve the weaker points while maintaining the strengths. The next step is to implement the plan. It is recommended that the questionnaire is used again at the end of the academic year and that the lessons learned be recorded in order to remember or to explain them to colleagues.

Table V.
The detected strengths and weakness

Version	Weakness
Long	– Students are encouraged to participate in class discussions. – Students are invited to share their ideas and knowledge. - How does this course compare with other courses you have had at this university? – Your level of interest in the subject prior to this course
Short	- My interest in the subject has increased as a consequence of this course. - The instructor enhances presentations with the use of humour. - The instructor's style of presentation holds your interest during class. – Students are encouraged to participate in class discussions. – Students are invited to share their ideas and knowledge.
Strengths	
Long	– The instructor is friendly towards individual students. – The instructor makes students feel welcome in seeking help/advice in or outside of class. – The instructor has a genuine interest in individual students. – The instructor is adequately accessible to students during office hours or after class. – The examinations/graded materials test the course content that is emphasised by the instructor.
Short	- The instructor is enthusiastic about teaching the course. – The instructor is friendly towards individual students. – The instructor makes students feel welcome in seeking help/advice in or outside of class. – The instructor has a genuine interest in individual students. – The examinations/graded materials test the course content that is emphasised by the instructor.

- Students require feedback as an element of motivation. As Chen remarked [54], "This study employs expectancy theory to evaluate some key factors that motivate students to participate in the teaching evaluation process. The results show that students generally consider an improvement in teaching to be the most attractive outcome of a teaching evaluation system. The second most attractive outcome was using teaching evaluations to improve course content and format. (…) Students' motivation to participate in teaching evaluations is also impacted significantly by their expectation that they will be able to provide meaningful feedback." Thus, teachers can give students feedback on the evaluation results.
- According to Centra [55], student evaluations of teaching can only facilitate improvement when professors are able to access new and valuable information from them. Teachers must then understand how to translate the new evidence into action and must be motivated to do so [55] and [56].
- The short SEEQ questionnaire can be used in combination with other methods [57] [28].

VII. Conclusions.

Concepts of quality taken from the entrepreneurial world are increasingly being incorporated into the academic field. Currently, there remains a high degree of concern about the improvement of teaching quality at the higher education level. Specifically, the improvement of teaching quality is one of the primary matters that should be addressed continuously by universities.

To improve teaching quality, it is necessary to evaluate the teaching process and its results. In this evaluation, students' opinions about learning and the teaching process are crucial because the students are the primary consumers in higher education.

The extensive body of research regarding student evaluation of teaching leads us to look for standard tools and standard methodologies. The review of national and international standards enables us to obtain guidelines and global methodologies. Their application will allow for the extraction of ideas for improving universities' traditional working methods. However, they do not provide any specific tool to evaluate the opinion of students regarding teaching quality. International organizations should work together to define standard tools and methodologies to evaluate students' opinions.

The SEEQ, developed by Dr Herbert Marsh, is a tool validated for use internationally. It has a robust factor structure, excellent reliability and reasonable validity. However, there are two problems with its use: the low participation of students (there are 37 questions to be answered) and the teachers' sense of being overloaded if they try to provide rapid feedback.

To solve these problems, a short version of the questionnaire is presented. It was possible to reduce the form using statistical methods. The proposed version consists of twenty-two questions. After using this new short version, the voluntary participation of students increased.

The short SEEQ questionnaire can be used as a tool to develop a teaching improvement process as its use detects teaching weaknesses and strengths. It is recommended that the questionnaire be used in the middle and at the end of the academic year in order to establish an improvement cycle. The reduction of the number of questions facilitates data

collection and the analysis of the data, in both cases, with software tools. Also it improves the flow of information. In this way, the proposed version of the SEEQ questionnaire helps produce feedback intended to motivate students' participation in the teaching evaluation process. In addition, the short questionnaire helps professors receive new and valuable information about their teaching from student evaluations more quickly. Professors can use the short SEEQ questionnaire with other tools.

The proposed short SEEQ questionnaire can be used by other university professors regardless of the subject or the degree course.

ACKNOWLEDGMENT

The authors would like to acknowledge the "Chair in Innovation and Technological Quality" and the CTP for their help. Thanks to the IEEE Foundation "Gobierno de Aragón" and to the "Fondo Social Europeo" for their support to the EduQTech group.

REFERENCES

[1] R. G. Lewis and D. H. Smith, "Total Quality in Higher Education". Delray Beach, FL: St. Lucie Press, 1994, Total Quality Series.

[2] F. Jurado, et al, "A review of the Accreditation Bodies and Processes in Europe. A vision from the Engineering", *35th ASEE/IEEE Frontiers in Education Conference*. F2D, 2005, pp. 13 - 18.

[3] D. Kember "Action Learning and Action Research: Improving the Quality of Teaching and Learning" Ed. Routledge. 2000.

[4] E. Stones "Quality Teaching: a sample of cases". Taylor & Francis, 1992.

[5] I. Plaza and C. Medrano "Continuous Improvement in Electronic Engineering Education" *IEEE Transactions on Education*, Vol. 50, Nº. 3, 2007 pp. 259 – 265.

[6] UNE 66176:2005. Quality management systems. Guide for measuring, monitoring and analysing customer satisfaction. 2005.

[7] W. E. Deming, "Out of the Crisis". *MIT Center for Advanced Engineering Study*. ISBN 0-911379-01-0. 1986.

[8] Plaza, I. et al (2008) "Code of good teaching practices based on quality criteria" (Original in Spanish) *Workshop of Educational Innovation, ITC and Educational Research* (2ª Jornadas de Innovación Docente, TIC e Investigación educative). CD of the workshop.

[9] JJ. Marcuello et al "Code of good teaching practices based on quality criteria". *EAEEIE Annual Conference*, 2008 19th, 2008. pp: 70-75. Digital Object Identifier 10.1109/EAEEIE.2008.4610161. Available at IEEEXplorer. Extended version of the previous article.

[10] F.M. Hill "Managing service quality in higher education: the role of the student as primary consumer", *Quality Assurance in Education*, Vol. 3 Iss: 3, 1995, pp.10 – 21.

[11] P. Spooren and D. Mortelmans "Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings?" *Educational Studies*, Vol. 32, No. 2, June 2006, 2006, pp. 201–214.

[12] E. Coşkun and M. Alkan "Evaluation of learning and teaching process in Turkish courses" *International Electronic Journal of Elementary Education* Vol. 2, Issue 3, July, 2010.

[13] P. Ramsden "A performance indicator of teaching quality in higher education: The Course Experience Questionnaire" *Studies in Higher Education*, Volume 16, Number 2, 1991, pp. 129-150(22).

[14] Joint Committee: The California State University, California Faculty Association and Academic Senate CSU. "Report on Student Evaluation of Teaching". 2008. Available at: http://www.calstate.edu/AcadSen/Records/Reports/documents/Report_on_Student_Evaluations_of_Teaching.pdf. Last visit: April 2013.

[15] H. T. Tagomori and L.A. Bishop "Student Evaluation of Teaching: Flaws in the Instruments". *Thought & Action*, v11 n1 1995, pp 63-78. Spr

[16] Z. Zerihun, W.Van Os and J. Beishuizen "Re-conceptualising approaches to the evaluation of teaching quality". Chapter of the book: Access & Expansion: Challenges for Higher Education Improvement in Developing Countries. Cantrell, M., Kool, R. & W. Kouwenhoven (Eds.) VU University Press, Amsterdam, The Netherlands, 221 pp. 2010. Available at: http://hdl.handle.net/1871/15816. Last visit: April 2013.

[17] http://www.iso.org/iso/home.html Website of the International Organization for Standardization. Last visit: April 2013.

[18] www.iec.ch/ Website of the International Electrotechnical Commission. Last visit: April 2013.

[19] http://standards.ieee.org/ Website of the IEEE Standards Association. Last visit: April 2013.

[20] [Plaza, 2010] I. Plaza et al "Quality and innovation in Higher Education: Code of Good Practices" *40th ASEE/IEEE Frontiers in Education Conference*. October 27 - 30, 2010, Washington, DC. 2010.

[21] ISO/IWA 2:2007 "Quality management systems -- Guidelines for the application of ISO 9001:2000 in education". ISO.

[22] ISO/IEC 19796-1:2005, "Information technology -- Learning, education and training -- Quality management, assurance and metrics -- Part 1: General approach". 2005.

[23] ISO/IEC 19796-3:2009 "Information technology -- Learning, education and training -- Quality management, assurance and metrics -- Part 3: Reference methods and metrics". 2009.

[24] C.M. Stracke, (2009): "Quality Development and Standards in e-Learning: Benefits and Guidelines for Implementations"; in: *Proceedings of the ASEM Lifelong Learning Conference: e-Learning and Workplace Learning*. Bangkok (Thailand). [Also online available on: http://www.qed-info.de/downloads. Last visit: April 2013].

[25] E. Campo et al. "La evolución y adopción de estándares en la formación virtual". *CompDes* 2010. 28-30 July. 2010. Available at http://www.redusoi.org Website. Last visit: April 2013.

[26] I. Plaza et al. "The use of Quality Standards as Element of Innovation in Higher Education." *4ª Conferência Ibérica de Sistemas e Tecnologias de Informaçâo. Póvoa de Varzim* - Portugal, 17 - 20 June. Proceedings book, 2009, pp. 567-572. ISBN:978-989-96247-0-2.

[27] UNE 66178:2004. "Quality management systems. Guide for the management of process for improvement". 2004.

[28] E. Keane and IM. Labhrainn "Obtaining Student Feedback on Teaching & Course Quality" National University of Ireland, Galway. 2005. Available at www.nuigalway.ie/celt/documents/evaluation_ofteaching.pdf. Last visit: August 2011.

[29] J. Richardson, "Instruments for obtaining student feedback: a review of the literature" *Assessment and evaluation in higher education*, Vol. 30, No. 4, 2005.pp. 387-415.

[30] M. and G. Graham "The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education. Research Note" *Assessment & Evaluation in Higher Education*, v26 n1 Feb 2001, pp 89-93.

[31] http://umanitoba.ca/computing/ist/teaching/seeqinfo.html Website of University of Manitoba (Manibota, Canada). Last visit: April 2013.

[32] http://www.smu.ca/ Website of the Saint Mary's University (Nova Scotia, Canada). Last visit: April 2013.

[33] www.mta.ca/ Website of the Mount Allison University (New Brunswick, Canada). Last visit: April 2013.

[34] http://www.usask.ca/. Website of University of Saskatchewan.). Last visit: April 2013.

[35] http://www.fordham.edu/ Website of the Fordham University (New York – U.S.) Last visit: April 2013.

[36] http://www.schreyerinstitute.psu.edu/Tools/SEEQ Website of the Schreyer Institute for Teaching Excellence (The Pennsylvania State University – U.S.) Last visit: April 2013.

[37] www.ox.ac.uk Website of the University of Oxford (U.K.). Last visit: April 2013.

[38] http://www2.le.ac.uk/ Website of the University of Leicester. Last visit: August 2013.

[39] http://english.semnan.ac.ir/. Website of the Semnan University (Iran) Last visit: April 2013.

[40] http://www.curtin.edu.au/ Website of the Curtin University (Australia) Last visit: April 2013.

[41] www.unav.es Website of the University of Navarra (Spain) Last visit: August 2013.

[42] www.uvigo.es/ the University of Vigo (Spain) Last visit: August 2013.

[43] http://www.upc.edu/ Website of the Polytechnic University of Catalonia (Spain). Last visit: August 2013.

[44] H. W. Marsh "SEEQ: A Reliable, Valid, And Useful Instrument for collecting Students' Evaluations of University Teaching" *British Journal of Educational Psychology*, Volume 52, Issue 1, pages 77–95,

February 1982. Article first published online: 13 May 2011 (http://onlinelibrary.wiley.com. Last visit: April 2013).

[45] Tale S, Nazifi M. and, Bigdeli I. "Validation of the Iranian version of student's evaluation of educational quality questionnaire". *Journal of Behavioral Sciences*, Vol. 3, No. 2, 2009 pp. 127-134.

[46] M.V. Verdugo and M.I. Cal "(Teaching Assessment: SEEQ) Valoración de la enseñanza: SEEQ" *Revista de Formación e Innovación Educativa Universitaria*. Vol. 3, Nº 4, 2010.182-193.

[47] M. Corbalan et al. "(Adaptation and Reduction of SEEQ questionnaire to know the opinion of students about teaching received) Reducción y adaptación del cuestionario SEEQ para conocer la opinión del alumnado sobre la docencia que recibe" CIDUI *2010 - New Areas of Quality in Higher Education - A Comparative and Trend Analysis*. 30 June – 2 July. Barcelona (Spain). 2010.

[48] E. Hervas "(SEEQ and GESTEST. Proposal of adaptation, reduction and computerization to engineering) SEEQ y GESTEST Propuesta de adaptación, reducción e informatización para ingeniería" Final Degree Project. EUPT. University of Zaragoza.

[49] http://www-01.ibm.com/software/analytics/spss/ Website of IBM® SPSS® Statistics. Las visit: April 2013.

[50] J.P. Lévy and J. Varela "(Multivariate Analysis for the Social Sciences) Análisis Multivariable para las Ciencias Sociales. Ed. Pearson-Prentice Hall

[51] J.R.A. Santos "Cronbach's Alpha: A Tool for Assessing the Reliability of Scales" Vol. 37, Number 2 , *Tools of the Trade - 2TOT3*. Available at the URL: http://www.joe.org/joe/1999april/tt3.php?ref=Klasistanbul.com. Las visit: April 2013.

[52] J. Nunnaly, "Psychometric theory". New York: McGraw-Hill. 1978.

[53] M. Valero-García, et al. "(Is it possible to do something else with the teaching polls?) ¿Se puede hacer algo más con las encuestas docentes? *2nd. Congreso Internacional: Docencia Universitaria e Innovación; (CUIEET Conference)* Tarragona (Spain). July 2002.

[54] Y. Chen and L. B. Hoshower "Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation" *Assessment & Evaluation in Higher Education*, Volume 28, Issue 1, pages 71-88. 2003.

[55] J. A. Centra, "Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness" Jossey-Bass - San Francisco, CA, Jossey-Bass. 1993.

[56] D. Cobb and V. Scott "Report of the 2010-2011 AQIP Student Evaluation of Teaching Committee". Available at the URL: http://www.siue.edu/innovation/assessment/set/pdf/SET_Report_In_processv5_FINAL.pdf. Last visit: April 2013.

[57] G. Gibbs and M. Coffey "The Impact Of Training Of University Teachers on their Teaching Skills, their Approach to Teaching and the Approach to Learning of their Students" *Active Learning in Higher Education* March 2004 vol. 5 no. 1, 2004.pp. 87-100.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 703–707

Tutor Platform for Vocational Students Education

Abdullah Saad AL-Malaise AL-Ghamdi, Habib M. Fardoun
Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah, Saudi Arabia
Email: {aalmalaise, hfardoun}@kau.edu.sa}

Antonio Paules Cipres
University of Castilla-La Mancha, Information Systems Department, Albacete, Spain
Email: apcipres@gmail.com

***Abstract*—Current vocational education characteristics as well as the dropout rates and compared to students' motivation, suggest the need for a system that allows and supports students to work towards the ultimate goal of their training: "Incorporating vocational students into the workforce". In this paper, we present a learning platform designed and developed to empower vocational and college students; a second aim is to provide students, faculty and staff a system to control and acquire new qualifications based on the official curriculum and work experience. The curricula-based improvement of the individual student's curriculum can be strengthened and improved when using the system, for the areas teachers determine necessary for each student. Consequently, the end system supports and contains official curricula information based on workers' professional qualifications, skills and competencies.**

I. Introduction

VOCATIONAL education operates to serve the labor market needs for specialized technicians. Currently, vocational education in Spain is far from the enrollment rates recorded within the European Union's average. The rate involves early school leavers who do not get a post-compulsory degree or Bachelor or vocational education; these dropout rates have increased in 2000; the rate was 29.1% of the total, while in 2007 it was increased to 31% and 31.9 in 2008, compared to 14.9% of the EU average. We also found a graduation rate in intermediate vocational education in Spain at 39%, far from the 51% of the European average, and from the 45% of the average within the European Economic Community [2].

Currently there are 639,887 pupils within vocational education, of whom 312,441 students study middle level vocational training, 288,861 students study upper undergraduate studies, and 385,86 students study distance vocational education. With some recent changes in 2012, 15.7% of the students, compared to the previous year 2011, have decided to pursue distance learning and dropout data with previous years has declined to 26.5% [3]. Based upon such data, we found that the 20.5% of the jobs in Spain are actually technical derived from vocational education. In fact, the industry has experienced an increase of 1.32 points and 9.62% of this specific labor supply related to technical staff from vocational education. Finally, the number of jobs for new graduates has been increased by 35.22% and 9.62% of these offers refers to technical jobs again obtained from vocational education [4].

The vocational education cycles are related to the technical and professional skill levels degree obtained in Spain. Thus, the educational ministry is in a process of adapting the current titles to the professional qualifications in Europe (European Qualifications Framework, EQF).

For the titles to be valid, the professional qualifications are structured based upon associated curricula, and thus, they directly affect the teachers' schedules. To follow this curriculum, the students are on a six years' formal training, to achieve professional qualifications specified by the training cycles. The structure associated with the vocational education teachers, facilitates the creation of activities tailored to the current companies' needs, the in situ immediate environment, and sometimes it involves a car provided to the students depending on the labor market needs. Currently, one of the real problems is that the professional qualification a student acquires is not related to the real knowledge, skills and competencies he needs, compared to the ones acquired in the academic courses (i.e. improve the level of basic skills in foreign languages).

In this paper, we conduct a study on tutors' needs for a platform to support vocational education students, workers and college students. This tutor-based platform is seeing from two perspectives, the official curricula and each student's personal curriculum, in order to meet college, students, and companies' needs. First, we present the current situation in vocational education in Spain for LOMCE[1] [1] and the ways it affects Spanish vocational education, the curriculum structure as well as the associated training cycles. Secondly, based on needs analysis, we define the targets to achieve. Having established the aims and objectives, and based on cloud architecture, we propose the design and development of a platform as well as the ways to use it. Lastly, we finish with conclusions and future work.

II. Structure And Classification of Professional Qualifications

Professional qualifications structures with associated curriculum provide the basis for comparison as well as the inter-relationships so to establish the main features of the vocational education units in relation to a system design and development. The professional qualifications characteristics and structure are described below [5] [6] [7]:

[1] LOMCE: Ley Orgánica de Mejora de la Calidad Educativa

978-83-60810-53-8/$25.00

I. Identifying properties:

- **Family:** It identifies the "Family" feature by name, and consists of 26 professional families, of which every family has 5 levels classification, depending on the professional competence required for productive activities in accordance with knowledge required for the specific level.
- **Name:** It identifies the professional qualification that is within each of the families, also following specific levels.
- **Code:** It is a unique identifier, an alphanumeric code, consisting of the first letters of the family and the numeric code identifier.

II. Properties that define objectives and scope:

- **General Competition:** Competition is defined as "the set of knowledge and skills that allow exercising professional activity in accordance with the requirements of production and employment."
- **Professional Scope:** The productive sectors and occupations or jobs are related to specific professional qualification.
- **Productive Sectors:** This is a sector in the economy that produces a material. The productive sector includes mining, forestry, fisheries, agriculture, industry and energy, but excludes government activity and social services.
- **Units of competence:** Competence is to perform each qualification functionality. Therefore, the unit of competency is the minimum aggregate susceptible to recognition, evaluation and accreditation part.

III. Identification Data:

- **Name**: It identifies the Competence Unit (UC). It is related to one of the aspects that describes the general competence of the professional qualification, which is associated with the UC.
- **Level of qualification:** The level of professional competence qualification to which UC is associated with, according to the degree of complexity, autonomy and responsibility required to perform a work activity (these are five levels).
- **Alphanumeric code:** It allows locating systematically the units in the qualification competition with which they are associated.

IV. Professional Achievements:

The Professional Achievements are accomplishments as competition elements that set an individual's expected behavior, in the form of consequences or results of the functions or activities s/he performs.

V. Performance Criteria:

They express the professional performance acceptable level of a certain function, so as to meet the associated productive organizations' objectives. Thus, they construct a guide towards the professional competence assessment.

VI. Professional Context:

Describe, as a mentor, production media, products and results of the work, information used or generated and the ways several items of a similar nature are considered necessary to frame the unit of competency.

This professional qualifications system applies on vocational education to ensure that students receive formal training according to a regulated system within the European Union. There is no tracking application in the market for vocational education students, unemployed, and job seekers to support the professional qualifications management according to the rules and the training they receive.

III. System Description

VII. Definition and Objectives:

Having dealt with the qualifications system and its relationship with the educational system and vocational education in particular, and in addition to understanding the professional qualifications system structure, we present the design and development of a supporting platform. This platform automatically establishes a student's acquired professional qualifications via a combination of competency units. An example is provided for students' new professional qualifications acquisition as follows:

- Workers: Describes the student's successful work expertise and competence units during his/her vocational education, in addition to several more free settings competency units s/he may acquire new professional qualifications.
- Student training: The student can acquire new qualifications in addition to previous ones via training units adapted to student's needs. In this way, students can complete their degree with more professional qualifications susceptible to recognition within the training courses that are in the same professional family, see section II-point A.
- College students: Students can gain relevant professional qualifications through the courses curricula corresponding to professional families, or students may opt for an acquisition of professional qualifications advised by their teachers or following their own criteria.

The system also identifies and distinguishes student diversity and competence; gifted students may develop more skills and students who do not meet the minimum requirements for the qualification have a professional qualification on a level they could complete. As a result, an increase on students' motivation is possible as well as they can improve their professional qualifications through formal training in universities and also in schools where vocational education is provided.

VIII. Teaching Methodology:

In regard to the teaching methodology and the minimum knowledge students should acquire in associated taught subjects, specific teaching methods were developed. The ones that allow student's abilities maximum development using learning technologies are related to the constructivist methodology; the student builds upon their own methods using provided tools to solve a given situation or problem [8] [9].

To achieve this, the main aim of the platform is to design an annual training plan for each student built upon the objectives set at the beginning of their academic year. Thus, they can be aware of what the qualifications acquired once evaluated, in addition to what content they should study and work with the help of their teacher or tutor. Therefore, our duty as instructional designers is to build a training plan that includes all the educational stages a student can follow, or an academic plan. In case of a worker, a tutor academic plan is provided, that encompasses professional qualifications completed over a period of time to achieve these objectives.

IX. Architecture:

A system in a cloud can enable and support the students, teachers and workers to interact. This system allows the inclusion of ancillary services and associated performances. We have chosen cloud architecture, because, in addition to supporting phased development over time and adding new services, it facilitates their deployment and growth and can be integrated within diverse systems built by other agencies.

Figure 1 shows the system architecture: it operates in two hybrid clouds that receive information from outside, and in turn, it communicates such information within these clouds.

The platform begins with data collection, based on data from institutions and actors who interact in the platform; these are the teachers, students and workers.

In green color The "Qualifications CTutor" cloud is depicted with green color; it provides services for data collection and maintenance for professional qualifications. Also the data "Integration Services" is represented; these services collect data from students' curricula in order to normalize the data and adapt it to specific needs. What we ensure is data normalization and structure by following the structured criteria, previously presented in section two. Thus, once the data is structured and validated, the curriculum profile for students and workers is created.

Separated into two distinct groups, students and workers, the connection to obtain new skills and create the workers' curricula is established; also a map of the obtained qualifications exists as well as the professional qualifications they may obtain from their studies (student or worker). The tool also indicates the specific qualifications to update or enhance. This cloud contains students' curriculum profile adapted to the national qualifications system, which allows and supports the growth of students' professional qualifications, keeps updated these professional qualifications on the profile, and adapts it to possible changes.

- Students: It contains the data obtained by the official curriculum and the objectives that the students have achieved in each of the official training courses; this is in order to have enough data to seek the needed professional qualifications according to their training needs and establish a curriculum to obtain professional qualifications that can be obtained by their studies.

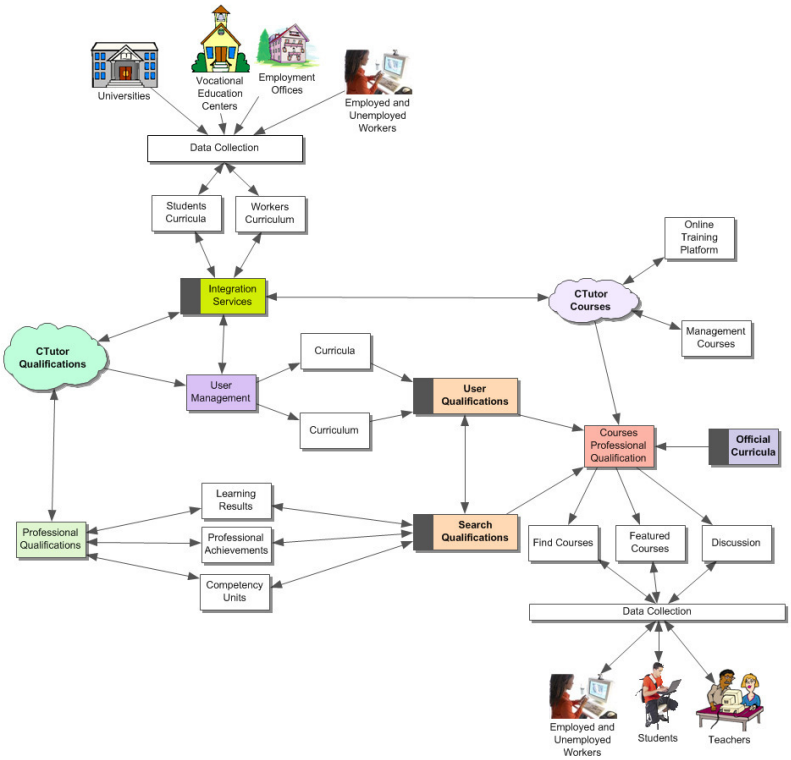


Fig. 1 System Architecture

- Workers: The data come from their curricula vitae; the CV must certify and accredit workers to obtain professional qualifications and training required for other professional qualifications; we also need to keep updating qualifications achieved to date.

The cloud "CTutor Courses" in purple color contains the services that facilitate the teaching and learning process. This cloud is fed by the data sent from the "Qualifications CTutor" cloud, and contains the necessary parameters to find the courses and users' profiles. The tool offers three options for choosing courses:

- **Find Courses:** Users can search for courses that best suit to their profile and then they can choose and start the registration process in the course, so to obtain the associated skills.
- **Featured Courses:** Users can access the recommended courses from their profile; this is the interface where teachers and human resource centers can recommend students what the way forward should be for training in accordance with the criteria to establish and improve the curriculum for pupils according to their official studies. In the case of workers, it is related to the needs that the company has about the qualification the company's employees must have, and also new expertise acquisition.
- **Discussion:** Students, teachers and community members enter into the debate educational course selection; as such, students can follow and solve any doubts on the process of choosing new courses. Sometimes it can be the case that a student or a worker is not aware of the comprehensive curriculum, so they can take advantage of the system and proceed to required corrections. At this level tests are included to accurately assess the student's knowledge, also including the necessary tests for languages found as required for the workers' skills as well as students' reinforcement in their language subjects.

It is important to keep the system updated, for that we include the official curricula of vocational education courses. As for the inclusion of qualifications in training courses, the target is to obtain formal qualification training at the end of several professional qualifications collection.

IV. How The Systems Works

The ways a student acts on the systems is represented here: after reviewing his/her profile, s/he has access to existing courses within the platform. The student detects that the platform recommends more courses than he already has, so to obtain better professional qualification. The platform offers him/her the possibility to start the acquisition process with these new courses, by, first, selecting them, and later, request the tutor's approval.

Fig. 2 Courses Screen

Figure 2 shows the information and actions that a student can visualize i.e.: students can view the detailed information about specific courses, a list of recommended courses and expected professional achievements. Therefore, based on such information, the student can work to obtain the best possible professional qualifications. The platform also offers a set of actions like:

- Consult personal data: The student can visualize his personal data and request the needed changes.
- Course Registration: The student can ask for registering with a set of courses at the beginning of each semester.
- Registration Errors Correction: If the student identifies registration errors, s/he may ask for corrections, and follow the request process.
- Courses Schedule: It shows the student his/her academic calendar (list of courses, rooms, time, etc.)
- Selecting favorite courses: The student may identify a course as a favorite course, so that both the tutors and the platform can list associated recommended courses for the student to study, so to improve his professional qualification specifically for this specific field.
- Searching for recommended courses: The student may ask for a list of recommended courses associated with his/her degrees.

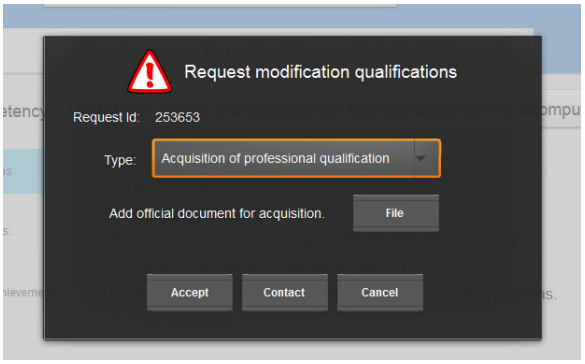


Fig. 3 Data Acquisition

Once a student observes that s/he already has the capabilities to acquire a competency unit, s/he may start the process of adding it to his/her professional qualifications, as in Figure 3. The student also has to argue about the merits with the administrators so to launch the validation process.

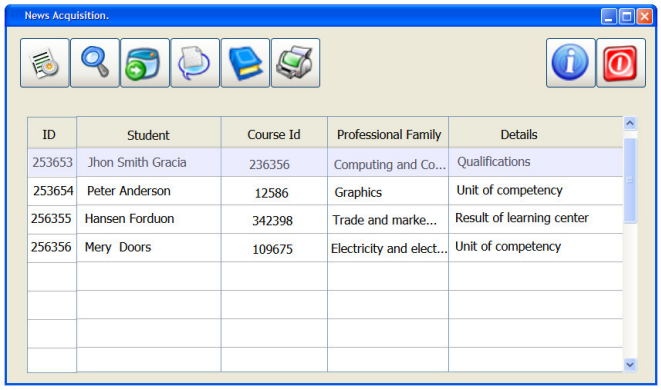


Fig. 4 System, new acquisition

During the acquisition process, the system sends a notification to inform the student about the process and the steps s/he has to follow, as in Figure 4. Once the student performs the requested information, s/he receives the notification informing him/her that the request is resolved; this service uses a message center.

V. Conclusion

The presented system aims to ensure students' and workers' improvement on the curriculum under the qualifications set within the European framework. Such system can keep the curriculum updated and allow the students to develop new qualifications based on skills acquired in formal training courses. The system interacts with users and provides all necessities related to the academic teaching methodology supporting the curriculum internationalization according to the European Qualifications Framework (EQF); an example is the language courses as they are becoming preparatory in Spain. As for the educational process, the students are able to perform additional studies, and thus, achieve more skills and be empowered in their curriculum subjects.

Following the platform design and development, future possible legislative changes are considered to adjust the platform to any new legislation structures and professional qualifications, as well as to adapt it according to the public academic administrations.

References

[1] Borrado Lomce http://www.mecd.gob.es/servicios-al-ciudadano-mecd/participacion-publica/lomce.html

[2] Bolívar Botía, A., & López Calvo, L. (2009). Las grandes cifras del fracaso y los riesgos de exclusión educativa. Profesorado: Revista de curriculum y formación del profesorado, 13(3), 51-78.

[3] Ministerio de Educación, Cultura y Deporte. Datos y cifras Curso escolar 2012/2013. http://www.mecd.gob.es/dctm/ministerio/horizontales/estadisticas/indicadores-publicaciones/datos-cifras/datos-y-cifras-2012-2013-web.pdf?documentId=0901e72b81416daf

[4] Adecco, soluciones de recursos humanos. III Informe Empleabilidad y Formación Profesional http://www.adecco.es/_data/NotasPrensa/pdf/410.pdf

[5] Cualificaciones http://servicios.aragon.es/eac/webgcp/cualificacionProfPublico.editar.do

[6] Ministerio de Educación, Pólitica Social y Deporte. Instituto Nacional de las Cualificaciones. Informe del Sistema Nacional de Cualificaciones y Formación Profesional. Edita Subsecretaria General Técnica.

[7] LEY ORGÁNICA 5/2002, de 19 de junio, de las Cualificaciones y de la Formación Profesional.

[8] Santángelo, H. N. (2000). Modelos pedagógicos en los sistemas de enseñanza no presencial basados en nuevas tecnologías y redes de comunicación. revista Iberoamericana de Educación, (24), 135-162.

[9] http://www.csems.uady.mx/media/docs/Formacion%20docente/Constructivismo%20y%20Competencias.PDF

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 709–712

New Subject to Improve the Educational System: Through a Communication Channel betwen Educational Institution-Company

Habib M. Fardoun, Abdulfattah S. Mashat, Lorenzo Gonzalez
Faculty of Computing and Information Technology, King Abdulaziz University,
Jeddah, Saudi Arabia
Email: {hfardoun asmashat, lgonzalez}@kau.edu.sa

***Abstract*—At the time students' finish their academic educational level within a current educational system, a big gap appears to exist between the acquired knowledge and the needed knowledge to enter the job world. This situation provokes certain disconcert about the necessity for adaptation period to be properly integrated in the working environment. For this reason, the inclusion of a new subject into the educational system is proposed serving as a nexus and a channel between the academic institution and the company. This subject focuses on the company employees' educational level required to be hired and that of the students needed to enter the working field. Thus, based on the specific job characteristics, the educational institutions as well as companies collaborate to establish a plan to follow. A company actually proposes the basic milestones a student must have to become a worker with the required skills. In this way, we, as university teachers, support students who, at the end of their academic educational stage have the needed knowledge, skills and competencies to begin their working life without the problems related to ignorance about the work to do.**

***Keywords*—Subject; Educational System; Educational Institution-Company Relation; Work; University.**

I. Introduction

IN THE current educational system, and in conjunction to the subsequent studies on the secondary education level, a problem appears to exist related to the students' preparation when starting their working life. To be more specific, the student needs an adaptation period so to work in any position of a specific company when s/he finishes high schools and universities (not as much at the vocational education). This situation provokes a feeling of insecurity to a student due to not getting the acquired knowledge and skills that are required and appropriate so to immediately perform at work. On the other hand, the students do not only suffer the consequences but also, the contracting companies have to estimate a time of adaptation for the new workers, which is costly in both time and money.

As a result, this paper describes a solution consisting of including a new subject at the last phase of the academic studies focusing on the company's work. The subject serves as a bridge and communication channel and collaboration between the educational institutions and companies; in this way, at the end of the student academic education, students can be incorporated in the working field without the need of any type of adaptation period or with a minimum one, which, in turn, does not cost to the contracting company as much. Thus, students, educational institutions, and companies take advantage of such subject, and also, the country's overall economy is improved as a response and consequence of a drastic diminution of the young people's unemployed rate. This is because companies do not require dedicating human and financial resources for the adaptation period and consequently, hiring is dramatically more effective.

There is increased importance in the aforementioned as nowadays, there are numerous and diverse factors that badly affect a country's economical growth. Among these facts, the unemployed rate is called the brain drain or the capable people's unhappiness caused by the low salary. One of the consequences is loss of interest towards completing their education to a university degree. This is probably because they know from the beginning that they will spend a great amount of money they possibly do not have and/or have to acquire support from their families or a loan from the banks. Moreover, for the time needed to study, the student will not have monetary benefits but problems with, for example, housing. Last but not least, in case of acquiring the degree, such certificate does not guarantee a job, which is rather a disappointing concept in regard to higher educational levels.

For these reasons, we propose the inclusion of this new subject to increase students' interest on higher educational levels, in order to becoming more familiar with the working life and with the job they have been studying for. This generates expectations and hope for the students and if hiring is more probable then there will be added value benefits.

The following sections are presented as follows: the current educational system situation, a global vision of how the subject can be included into the academic plan, the

978-83-60810-53-8/$25.00

channels of communication between institutions and companies, the evaluation criteria for the students in that subject, and finally, and the work that we are doing to improve this proposal.

II. State of Art

With respect to university academic education, a solution exists to reduce the problem previously highlighted; this is the gap between the finishing the academic education and starting the working life. This solution is called "company practices". It consists of a student working in a company, who asks for this kind of practice. This type of work is remunerated, in some cases, and it is a bridge between the student, as a worker, and the company, which helps the student. On the other hand, in many occasions, the work done by the students inside a company does not reflect students' expectations. In other words, students think that they are going to learn about the working conditions and skills needed in a company within the field that they have been studying, however, they usually acquire a job unrelated to their academic education. Definitely, in some cases, companies use this kind of opportunities to hire people at low salary or no salary to perform very low-level tasks or tasks that nobody of the current workers wants to do.

With respect to the vocational education, the gap is not as wide as with university academic education. It is rather more focused on supporting the students entering the work world and for this reason; a closer relation with the local companies exists. In addition, there are studies to prepare the students who would like to enter the working life as soon as possible. Those studies are called "Programas de Cualificación ProfesionalInicial" (PCPI), in English, Initial Professional Qualification Programs; and previously called "Programas de Garantía Social" (PGS), in English, Social Guarantee Programs [1].

Despite the aforementioned problems, any help that the educational system could provide to improve the integration of the students in the working life would be well received by the society. For that, we are going to explain our proposed idea to ease this process.

III. Global Vision

The proposed subject is included in the last course of the corresponding studies plan. Thus, the students have the academic concepts as fresh as possible and the companies can identify a person to cover a vacant in short notice. This subject is optional, as it occurs within the company practices, and prepares the student with the needed knowledge to start working in a specific job. For that, the student identifies the kind of position he wants to occupy, when he finishes his studies, his spurting tutor and mentor etc. as the support provided by the university has to be as realistic as possible. After that, it will be clearer for the student to have a list of themes related to the desired job. This list of themes is established in a base shared between the institution and the company in order to know the essential milestones the student has to achieve to obtain the desired job.

As this is a subject in constant evolution, the companies will not be always the same and for this reason, the required essential milestones have to change occasionally. For example, two companies establish an agreement with the university where the needed criteria are established for a student to work in a vacant job as a Computer Science Engineer. One of the companies' jobs requires good knowledge of the CMS[1] Joomla![2] and the other .NET[3] Technology. Both vacancies require a Computer Science Engineer to develop the work. However, it is highly likely that the student at the end of his/her studies does not have enough knowledge and skills on any of the previous technologies. In addition, those technologies are different, so it is impossible to establish a common list of themes for both because the student will not be specialised to the required degree.

Taking into account the previous concepts, we are creating two different types of lists of themes associated with the current company necessities, which can be selected by the students who would choose the subject. One of the lists of themes is focused on the work on Joomla! where the student must study themes related with PHP[4], free software licenses, databases with MySQL[5], managing of a Apache server, etc. The other list of themes focuses, differently, on the programming language C#, the study of the .NET Framework, the architecture in layers, the client side programming language JavaScript[6], the SQL Server[7] database, etc.

Once we know the different lists of themes available for a specific course, the academic institution passes the information to the students so they can choose the closest to their necessities. If there is not enough number of students, more academic load will be dedicated to a subject with one

1CMS. Content Management System.
2Joomla!. Open code CMS, developed on PHP and under GPL license.
3.NET. Microsoft Framework for developing applications.
4 PHP. Server side programming language usually used for web development.
5 MySQL. Relational database management system of Oracle.
6 JavaScript. Client side programming language used to improve the performance and the interface visualization.
7 SQL Server. Relational database management system of Microsoft.

list of themes, and another list of themes will be rejected as non-viable (as occurs at the current educational system).

At the end of the academic stage, the students will have the required skills for the current working life for a set of companies that focus on specific technologies. In addition, the companies collaborating with the university can provide the bases for future workers to have at their availability qualified personnel so to work on the moment as they enter the company; this increases the students' or ex-students hiring rate by the specific institution. Such targets combination avoids the existing gap between the student and the company due to the lack of information and/or resources.

IV. Educational Institution-Company Communication

Finally, a communication channel is needed to bridge educational institutions and companies. This communication channel is of great importance because the successful collaboration between them determines the success or the failure of the whole project. So, if we establish the correct methods and level of understanding, the results obtained will be beneficial for all, the students, the institution, and the companies.

Therefore, different meetings between the interested parties can provide, as a result, the bases to establish the knowledge and capacities that students have to acquire to reach a determined job in a specific company. Moreover, we obtain real work-related profiles of qualified personnel to occupy a vacant job when finishing his/her education, which is nowadays not offered by the university.

This type of relationship is beneficial for everyone involved and even for not explicit parts, as for example, the employment system. Below we detail the resulting possibilities:

- For the students. The proposition improves the student self- esteem and attitude towards a future job. The student knows that the subject will be really useful for his/her working future. In addition, due to the fact that the company is already continuously immersed in this process, it increases the possibilities of hiring a student by that company or another, which works with the specific kind of technologies and/or methodologies.
- For the university. Improving students' attitude results on choosing the subject that will open the working world, and increase the number of students deciding to register for that subject, which in turn, directly affects the university's strongbox. Moreover, if there are a lot of companies that present their technologies and methodologies and due to this several plans created, an increase in the number of job vacancies can be established. Thus, this increases the number of subjects to teach, which would be great for teachers with problems appeared as lack of opportunities. Lastly, however, not less important, is the increase in possibilities of hiring students, therefore, the university's prestige would grow, raising the "graduated student-hired student" rate.
- For the company. As a consequence of this relationship, the company needing specific employees does not overspend on human and financial resources searching for suitable candidates who fit the job requirements; this is because the potential future employees have already studied the subject that is in accordance with the company plan. In addition, the adaptation period for a new worker is removed or drastically reduced. This increases the initial productivity and, consequently, the profits.
- For the employment system. The employment system is indirectly related to the communication channel between the educational institution and the company. This is due to different reasons; for example, and as aforementioned, is the number of plans about the subject increases, the number of vacancies for teachers needed also increases, which in turn decreases the number of unemployed people. Another factor is related to the direct relation between a company, the university and the subject with the specific list of themes for a job in that company and this is due to the communication channel; as the number of hired students recently graduated increases, the unemployed young people rate decreases.

To conclude, if we establish an effective, organized and structured communication channel between the educational institutions and the companies, more global benefits are obtained. These benefits are applied to every part and improve the educational and working system as such.

V. Evaluation

Finally, evaluation of this proposition is needed to assess that the studied concepts by the students are rooted enough and they have been useful to them. For this, and as a result of the conversations between the university and the company, evaluation criteria are needed to impart the subject and to obtain qualitative conclusions referred to the capacities of each student.

The evaluation need to be composed of theoretical parts as well as practical parts, depending on where the company assigns weight to the most important parts; these are the parts which have more weight at the time the teacher evaluates the work presented by the student. As for the degrees, for instance, in engineering or architecture, the evaluation criteria can be related to practices. However, for the rest of the degrees, if for example related to Arts, practical applications appear to be better and more positive

for the student because as they refer to real cases; these can be found and implemented inside the company, related with the study plan for the subject or other similar competences.

To increase the students' interest, we can establish agreements; companies are obliged to hire a specific and minimum number of students (if the students have finished the subject with appropriate qualification). Thus, at the end of students' studies, there will be interviews between the company and the selected students.

VI. Conclusions and future work

Nowadays, serious problems exist with respect to the rate of unemployed young people and the difficulty of the recently graduated people related to getting a job based on their studies inside of the frontier. For that reason, in this paper we proposed the inclusion of a new subject to the study plan of the educational institutions. This subject foments the communication channel between companies and educational institutions to achieve beneficial goals for everyone. Thus, the result is efficient and effective for all, students, company, educational institution and the public employment system. This is due to the creation of new work vacancies that help companies to find adequate people fitting with the company profile. Furthermore, students' efforts are rewarded with the possibility of being hired by the company, which manages the plan for the subject, or by companies with methodologies and/or technologies similar to that plan.

It is also necessary to mention that we are working towards managing different generated plans for adapting the conversations and agreements between educational institution and companies having a big part of their contents common to others. Thus, not only do the students achieve the goals of a particular company and their own but also within many companies, and thus, increasing the possibilities of hiring them. Getting more prepared students whose skills are open to a wider number working possibilities. Therefore, the improvement of the studies model plan of a subject for higher studies have to be considered, and also, aspects on studies of a lesser grade. For this, we are studying different forms for the students who do not want to follow higher studies; however, they can acquire a good career opportunity following the same paradigm of bridging subjects and communication between educational institutions and companies.

References

[1] Ministerio de Educación, Cultura y Deporte. Oferta formativa referida al CNCP (Catálogo Nacional de Cualificaciones Profesionales).Ministry of Education, Culture and Sports. Formative offer referred to the National Catalog of Professional Qualifications. http://www.educacion.gob.es/educa/incual/ice_OfertaFormativa_CNCP.html#PCPIs.

[2] Cataloging teaching units: Resources, evaluation and collaboration. Antonio Paules Ciprés, Habib M Fardoun, Abdulfattah Mashat. 2012. Computer Science and Information Systems (FedCSIS). E-ISBN : 978-83-60810-51-4. Print ISBN: 978-1-4673-0708-6.

[3] School Community Relations (3rd Edition). Douglas J. Fiore, Ph.D. 2011. ISBN 978-1-59667-161-4.

[4] Ivory Tower to Concrete Jungle: The Difficult Transition from the Academy to the Workplace as Learning Environments. P. C. Candy and R. G. Crebert. The Journal of Higher Education Vol. 62, No. 5 (Sep. - Oct., 1991), pp. 570-592. Ohio State University Press.

[5] Delivering the promise: the transition from higher education to work. Colin Graham, Alasdair McKenzie. 1995. Education + Training, Vol. 37 Iss: 1, pp.4 – 11.

[6] Experiences of Recent High School Graduates. The Transition to Work or Postsecondary Education. Nolfi, George J.; And Others. 1978. Lexington Books.

[7] The Entry-Level Engineer: Problems in Transition from Student to Professional. Susan M. Katz. 1993. Journal of Engineering Education Volume 82, Issue 3, pages 171–174.

[8] Pedagogical perspectives on the relationships between higher education and working life. Päivi Tynjälä, Jussi Välimaa, Anneli Sarja. 2003. Institute for Educational Research, University of Jyväskylä, Finland.

[9] Trends 2010: A decade of change in European Higher Education. Andrée Sursock, Hanne Smidt. 2010. European University Association.

[10] Reforming the labour market in Spain. Anita Wölfl, Juan S. Mora-Sanguinetti. 2011. OECD Economics Department Working Paper No. 845.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 713–716

Improving learning methods through student's opinion into teacher's curricula Using graphical representations

Habib M. Fardoun, Daniyal M. Alghazzawi, Lorenzo Gonzalez
Faculty of Computing and Information Technology, King Abdulaziz University,
Jeddah, Saudi Arabia
Email: {hfardoun dghazzawi, lgonzalez}@kau.edu.sa

***Abstract*—Nowadays in curricula of university teachers doesn't appear anything related with the way to do the classes or the efficiency of the methods used with the students. We must have to take into account that a teacher is not only defined by his knowledge about a specific field but he is defined also by the way to pass it. For that reason we propose the inclusion inside of the teacher curricula of a section expressly dedicated to the student's opinion about the execution of his classes. This information is obtained through surveys and it will be displayed graphically with the goal of localizing the aspects which the teacher must improve and of maintaining an historic register that helps to check the progression. In addition, it will serve to the competent educational organisms to know the different skills of their employees.**

Keywords—Graphical Stadistics; Curriculum; Usability; Evaluation; Study; Student's opinion; Surveys.

I. Introduction

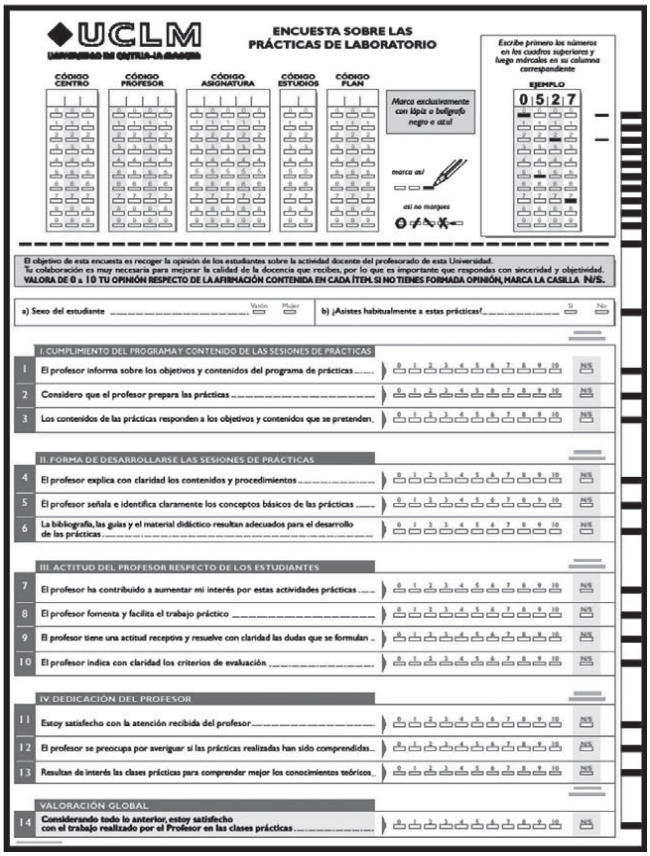
IN THE current higher education system in Castilla-La Mancha the teachers are evaluated by the students by mean of surveys which will be taken into account later for future bonuses or others [1]. These types of surveys performed by the students evaluate different aspects of the learning of a determined teacher. This kind of proposal is a good idea for keeping a good teachers attitude with the students inside the system. In addition, it means that not only it's needed a good base of knowledge of the teacher but a good way to express them and to show them to the students for a better understanding and acquisition of them. However the results of the surveys don't have enough influence inside the system in general and the teachers in particular.

In this paper we propose the study and inclusion of student surveys inside the teacher's curricula. Thus, a historic related with the attitudes and aspects to improve and improved, is kept in the system. But this is not only useful to study the learning aspects related with the teacher but also it forces the teacher to get interest for his work of knowledge transference of a more active form. That is because it will be displayed at the curricula and visible for the competent institutions.

The inclusion of this type of elements will be captured of a graphic manner inside the curricula to ease their understanding. Thus, with a quick sight we will know the strongest and weakest points of a determined teacher in particular. Moreover as the obtained results by the surveys won't be ephemeral (stored in a history), this foments a care for part of the teachers to give the classes. With all of it we don't only get to take into account the experience and knowledge of a determined teacher to evaluate his skills but also to take into account his work with the students and the way to foment the learning.

II. State of Art

Nowadays the evaluation of the teachers is performed by mean of surveys. These surveys were passed to the students,


UCLM

ENCUESTA SOBRE LAS PRÁCTICAS DE LABORATORIO

Escribe primero los números en los cuadros superiores y luego márcalos en su columna correspondiente

EJEMPLO 0 5 2 7

CÓDIGO CENTRO | CÓDIGO PROFESOR | CÓDIGO ASIGNATURA | CÓDIGO ESTUDIOS | CÓDIGO PLAN

Marca exclusivamente con lápiz o bolígrafo negro o azul

marca así

así no marques

El objetivo de esta encuesta es recoger la opinión de los estudiantes sobre la actividad docente del profesorado de esta Universidad.
Tu colaboración es muy necesaria para mejorar la calidad de la docencia que recibes, por lo que es importante que respondas con sinceridad y objetividad.
VALORA DE 0 a 10 TU OPINIÓN RESPECTO DE LA AFIRMACIÓN CONTENIDA EN CADA ÍTEM. SI NO TIENES FORMADA OPINIÓN, MARCA LA CASILLA N/S.

a) Sexo del estudiante ... Varón / Mujer

b) ¿Asistes habitualmente a estas prácticas? ... Sí / No

I. CUMPLIMIENTO DEL PROGRAMA Y CONTENIDO DE LAS SESIONES DE PRÁCTICAS

1 El profesor informa sobre los objetivos y contenidos del programa de prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S
2 Considero que el profesor prepara las prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S
3 Los contenidos de las prácticas responden a los objetivos y contenidos que se pretenden ... 0 1 2 3 4 5 6 7 8 9 10 N/S

II. FORMA DE DESARROLLARSE LAS SESIONES DE PRÁCTICAS

4 El profesor explica con claridad los contenidos y procedimientos ... 0 1 2 3 4 5 6 7 8 9 10 N/S
5 El profesor señala e identifica claramente los conceptos básicos de las prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S
6 La bibliografía, las guías y el material didáctico resultan adecuados para el desarrollo de las prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S

III. ACTITUD DEL PROFESOR RESPECTO DE LOS ESTUDIANTES

7 El profesor ha contribuido a aumentar mi interés por estas actividades prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S
8 El profesor fomenta y facilita el trabajo práctico ... 0 1 2 3 4 5 6 7 8 9 10 N/S
9 El profesor tiene una actitud receptiva y resuelve con claridad las dudas que se formulan ... 0 1 2 3 4 5 6 7 8 9 10 N/S
10 El profesor indica con claridad los criterios de evaluación ... 0 1 2 3 4 5 6 7 8 9 10 N/S

IV. DEDICACIÓN DEL PROFESOR

11 Estoy satisfecho con la atención recibida del profesor ... 0 1 2 3 4 5 6 7 8 9 10 N/S
12 El profesor se preocupa por averiguar si las prácticas realizadas han sido comprendidas ... 0 1 2 3 4 5 6 7 8 9 10 N/S
13 Resultan de interés las clases prácticas para comprender mejor los conocimientos teóricos ... 0 1 2 3 4 5 6 7 8 9 10 N/S

VALORACIÓN GLOBAL

14 Considerando todo lo anterior, estoy satisfecho con el trabajo realizado por el Profesor en las clases prácticas ... 0 1 2 3 4 5 6 7 8 9 10 N/S

Fig. 1. Survey of practical part.

978-83-60810-53-8/$25.00

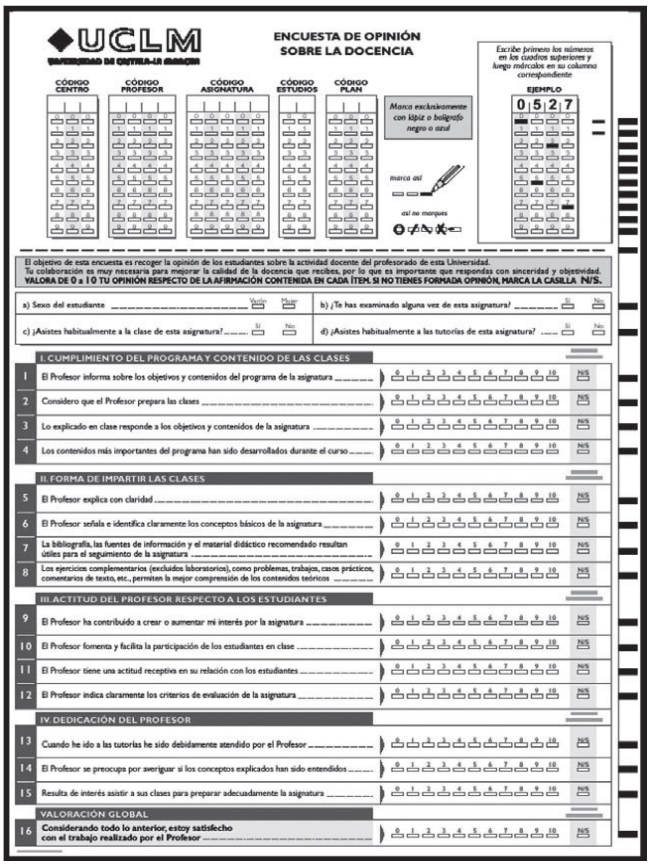


Fig. 2. Survey of theoretical part.

until this kind of fact started to be done electronically in 2011, at the final part of the teaching of each subject and in relation with a determined teacher. In that moment the students qualified the different attitudes of the teacher in an anonymous form. Thus, the results referred to the teaching period of a specific subject are obtained.

Surveys were determined by formularies which the students must paint the specific value inside of each category. At the following images we can see a form related with a theoretical part and another related with a practical part of a subject.

Actually the process can be developed by mean a web application. Thus, we reduce the costs and we promote the use of the electronically media. In addition, through this application the teachers may download the reports referred to the students' opinion to check if the methodological aspects of the teaching are adapted to the current students.

By observing these data, we propose the inclusion of the surveys' results in the personal curricula of each teacher. With it everyone can know the progression of a determined teacher and if the teacher's capacities satisfy to the students of the subject. Thus, we get various goals:

- To foment the dedication of the teaching stuff to improve their learning skills.
- To guarantee that the students are agree with the learning and with the taught knowledge.
- To obtain a history for the teachers can see what facets to improve.
- To include in the teacher's curricula not only the technical skills but also the learning skills.
- A control by the educational institution of their employees, by checking if they reach the expected expectatives.

III. Evaluation and Study

If we take into account the figures 1 and 2, we can see that there exist four main factors to evaluate inside of the capacities of a determined teacher. In function of it is theory or practices respectively, these are:

- Compliance of the program and content of the classes (or practice sessions).
- Way to teach the class (or to develop practice sessions).
- Teacher attitude with students.
- Teacher dedication.

Each of these capacities comes also determined by more specific aspects inside of them which form the global. These aspects will be evaluated by mean of the students' opinion and will be treated to avoid extreme values which go out of the real estimation. Knowing that the valuations of the students usually have a normal distribution [2], different mathematical and statistical methods are used, as for example the empiric rule which use the standard deviation as selector element of the data range used to the evaluation [3].

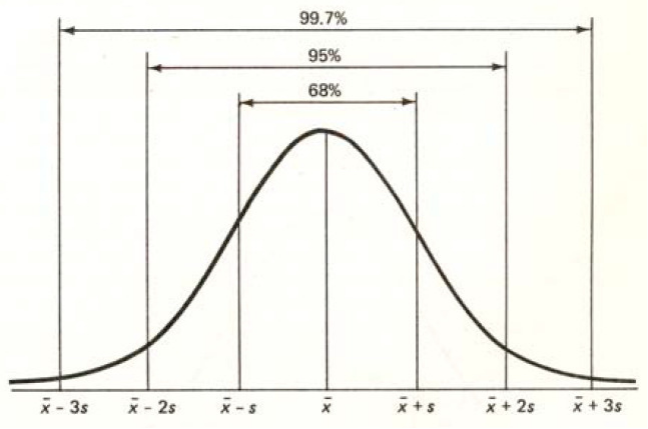


Fig. 3. Empiric rule.

Once we have the validate data inside of the sample set, we will proceed with the calculation of the mean of each of these data in reference to each question of the survey. After that we will evaluate at the same form the global capacity in function of these specific data. With it, we will obtain tangible data to be used by teachers and pertinent educational institutions.

IV. Inclusion into teacher's curriculum

Knowing the punctuation provided by the students to a teacher in a determined subject, next thing to do is the inclusion of these data inside of the teacher's curricula. Nowadays this information is available through a web application, thus the teacher may download the reports when he wants. However this type of methodology doesn't foment a increment on the teacher's interest to improve his classes. For it, in this paper we propose the insertion of the data inside of the teacher's curricula for giving it importance and this is taken into account. With this option, students and educational institutions will have more opportunities to visualize it.

All those things drive the teacher attitude to a restructuration of his classes in case of these classes don't give the expected result in a specific field or to a maintenance of the form of doing them in case of the evaluations were successful. Thus, we obtain a greater control over the learning for the knowledge arrives in good conditions to the students, getting a quality education. These data will serve as a history to see the evolution of a determined teacher and as a supporting point to improve on the fields where he is failing. However, the fact of to include the data doesn't have too much influence on the attitude for visualizing them. For that reason, the best form for these data call the reader's attention (the teacher or anyone else) is to show them by mean of intuitive and easy graphs, understandable in a first look.

This kind of graphs will call the attention and will show the evaluations of the four main components of the previously treated surveys at the state of art.

V. Graphical representation

There exist different forms to display data graphically to the user. In this particular case we need to show a graph which information contains the four main facets related with the learning of a determined teacher. In addition, we must to take into account that for each of them there exist other aspects which constitute them. These aspects will be also represented.

Following the specifications described in the article [4] where the graphs of football games are taken as reference, we are going to use for this case the same form of representation but with the particularity of using a square rather than an octagon. Thus, the representation of the data would be displayed as follows at the figure 4.

Once observed the data referred to these four main elements which constitute the evaluation of a teacher, for the case of wanting to look deeper in each of them, we only need to pass over one of them to see the associated data. With it, we can watch the information related with the evaluation of a determined aspect of the teacher learning. At the figure 5 we can observe how that information is shown (theoretical part).

With all of these data added to the teacher's curricula, the information contained in it will be completed. Of this form

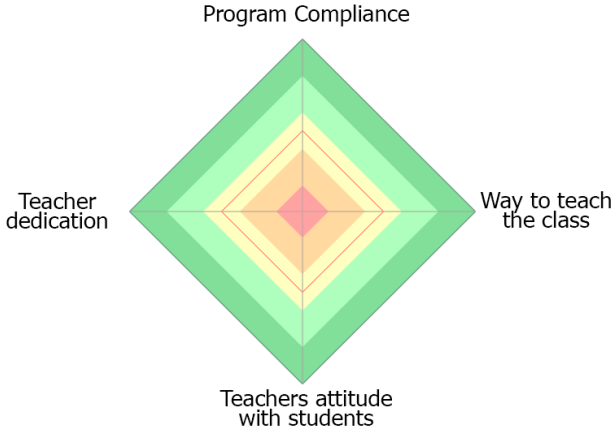


Fig. 4. Graphic representation of the teacher's attitudes.

there won't be only technical aspects but also aspects referred to the learning attitudes.

VI. Conclusions and future work

With the inclusion of the students' opinion into a teacher's curricula we achieve that the teacher pays more attention to his work of teaching because that information will be available for the educational institutions and students can see it when they want. Moreover, as that data is displayed graphically, the user who consults these data will have it easier and in one look he will know the attitudes of a determined teacher. But it is not only used as a consult method to external agents to the teacher but the teacher will be able to consult the data to see the areas which need a improvement for the efficiency in the educational labours was more productive. With it, the opinion of people who must receive the knowledge for the educational system, gets the deserved importance. In addition we foment the good practices and the self-criticism of the teachers to improve their learning capacities. All of this is resumed in the aspects previously commented:

- To foment the dedication of the teaching stuff to improve their learning skills.
- To guarantee that the students are agree with the learning and with the taught knowledge.

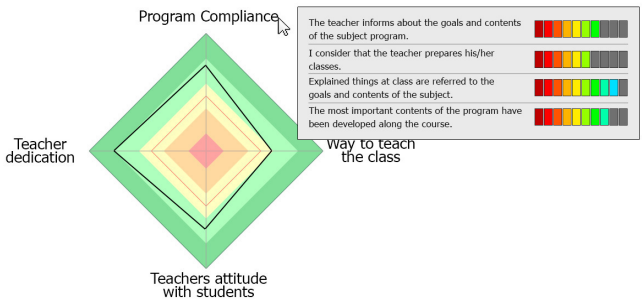


Fig. 5. Contained information inside of a specific aspect of the teacher's learning.

- To obtain a history for the teachers can see what facets to improve.
- To include in the teacher's curricula not only the technical skills but also the learning skills.
- A control by the educational institution of their employees, by checking if they reach the expected expectatives.

The following challenges that we have proposed to us are referred, among others, to foment that the participation in the surveys to be the more realistic as possible, trying to avoid that the external factors have influence on the students at the time of qualify the attitudes of a determined teacher and on a specific subject. This type of factors can be determined by comments of the teachers due to a fear to a negative evaluation or by the students trying to do wrong evaluations without take into account that this actions can have a negative influence into the curricula of the teacher to evaluate.

References

[1] Guía metodológica de la encuesta de "opinión de los estudiantes sobre la docencia del profesorado" en la Universidad de Castilla-La Mancha (Survey Methodology Guide "opinion of students on the teaching skills of the faculty" at the University of Castilla-La Mancha). Andrés Vázquez Morcillo, Jesús Santos del Cerro, Ángel Manuel Patiño García, Elena Silva Gutiérrez. 2005.

[2] Handbook of the Normal Distribution, Second Edition (Statistics: A Series of Textbooks and Monographs). Jagdish K. Patel, Campbell B. Read. 1996.

[3] Introductory Statistics. Douglas S. Shafer, Zhiyi Zhang. 2012. ISBN-13: 978-1453344873.

[4] Graphical Evaluation of Expedients. Lorenzo Carretero González, Habib M. Fardoun. 2013. Sent to IDEE 2013 Workshop (ICEIS 2013 Congress).

[5] HTML5 Graphing and Data Visualization Cookbook. Ben Fhala. 2012. Packt Publishing. ISBN 978-1-84969-370-7.

[6] 62 Tips on Graphic Design, UI/UX Design, and Visualization for eLearning. 2012. The eLearning Guild.

[7] Asking Students about Teaching: Student Perception Surveys and Their Implementation. MET Project. 2012. Bill & Melinda Gates Foundation.

[8] Best Practices for Including Multiple Measures in Teacher Evaluations. 2012. Hanover Research.

[9] Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. Laura Goe, D. Courtney Bell, D. Olivia Little. 2008. National Comprehensive Center for Teacher Quality.

[10] Teacher Evaluation: A Conceptual Framework and Examples of Country Practices. Paulo Santiago, Francisco Benavides. 2009. prepared for presentation at the OECD-Mexico Workshop Towards a Teacher Evaluation Framework in Mexico: International Practices, Criteria and Mechanisms, held in Mexico City on 1-2 December 2009.

[11] Handbook on Teacher Evaluation: Assessing and Improving Performance. James H. Stronge, Pamela D. Tucker. 2003. ISBN: 9781930556584.

[12] Teacher Evaluation: A Comprehensive Guide to New Directions and Practices. Peterson, K. 2000. 2nd edition, Thousand Oaks, CA: Corwin Press.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 717–720

IS (ICT) and CS in Civil Engineering Curricula: Case Study

R. Robert Gajewski
Warsaw University of Technology,
Faculty of Civil Engineering
Armii Ludowej 16, 00-637
Warszawa, Poland
Email: rg@il.pw.edu.pl

Lech Własak
Warsaw University of Technology,
Faculty of Civil Engineering
Armii Ludowej 16, 00-637
Warszawa, Poland
Email: lw@il.pw.edu.pl

Marcin Jaczewski
Warsaw University of Technology,
Faculty of Civil Engineering
Armii Ludowej 16, 00-637
Warszawa, Poland
Email: mjacz@il.pw.edu.pl

***Abstract*—The paper presents case study – Information Systems and Computer Science in Civil Engineering curricula. Introduction gives historical background of present role and position of IS, Information & Communication Technologies (ICT) and CS. Later details of the course of Information Technologies (IT) are presented in which elements of IS (ICT) and CS are combined. The use of spreadsheet and its Solver as well as Computer Algebra System (CAS) are stressed. Special attention is put on lectures which give necessary theoretical background for classes. Additionally there are presented some remarks about the subject Computing in Civil Engineering (CCE) which is natural successor of IT course. The paper is illustrated by the results of questionnaires performed in the beginning and in the end of semester. The main purpose of the article is to present changes and convergence between mentioned above subjects.**

I. Introduction

THE place and role of Information Systems (IS), which can be treated in narrow sense as a term referring mainly to ICT, and Computer Sciences (CS) in curricula of Civil Engineering (CE) studies changed a lot in last few decades. In ’70 and ’80 mainframe computers were used and students were taught Algol and FORTRAN as programming languages. In ’90 first PC laboratories were created so Algol and FORTRAN were replaced by Turbo Pascal. Students were also taught software like spreadsheets and text processors. Moreover for the purpose of Numerical Methods Computer Algebra System (CAS) MathCAD was introduced. In XXI century Windows environment and its applications dominates in curricula of studies. Additionally to these changes IS and CS are more separated than before. They are taught in different subjects namely Information Technologies and Informatics because Visual Basic for Applications (VBA) as sample programming language was in 2005 replaced by C++. The place and role of IS and CS in curricula of CE studies are subject of never ending discussions. The most orthodox engineers see no room for such subjects. In their opinion students’ knowledge gained from secondary school on the level of European Computer Driving License (ECDL) described in [1] is absolutely satisfactory for future civil engineers. Their opponents see growing role of IS and CS in all engineering fields. Existing curricula is to some extend a kind of compromise between these two opposite opinions. In the rest of the paper laboratories and lectures in the field of IT, curricula of CCE and attempts to flip the education will be presented.

II. Information Technologies: Laboratories

Because in common opinion C++ taught on 2nd semester of BSc studies in subject named Informatics is not the best idea curricula of the subject Information Technologies (IT) taught on 1st semester is a trial to make a reasonable combination of IS and CS. Curricula of IT subject is based on compromise between the level of students’ knowledge and foreseen needs of other subjects. While first point can be easily measured by questionnaires second one is hard to tackle with because of mentioned above conservative attitude of many teachers to the role of IS and CS in engineering.

A. Results of questionnaires

Questionnaires has been conducted regularly since 2011.

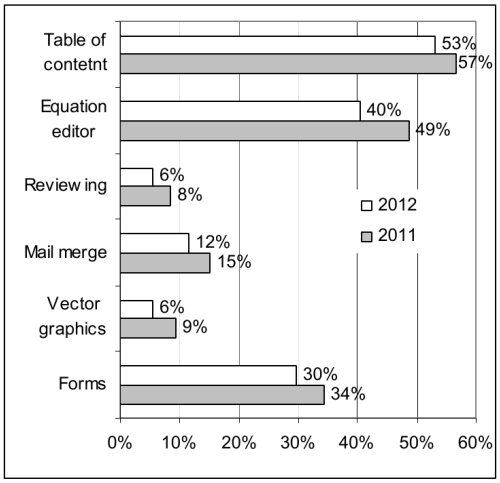


Fig. 1 Sample questions concerning text editor

This work was supported by Warsaw University of Technology

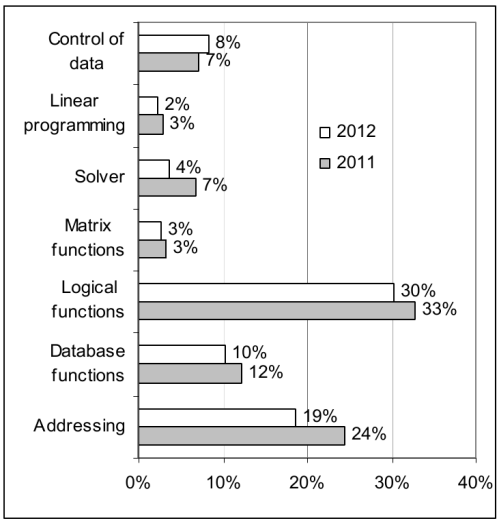


Fig. 2 Sample questions concerning spreadsheet

Their results are worse than expected. Students know how to run software like text editor or spreadsheet but they do not know how to use it in order to solve particular problem in effective way. Both Fig. 1 and Fig. 2 show that the knowledge of more advanced functions in text editor and in spreadsheet decreases. This means that material from IT on the level of ECDL stil should be present in curricula of studies.

B. Block 1: First things first…

First block consisting of three classes can be named first things first. Students are definitely very skilled mainly in dragging, dropping and tapping, so topics like file systems, file transfer and rights are quite new for some of them. Similar situation happens in the field of presentation graphics. Ten slides and 20 MB files are common problems because terms like resolution and appropriate graphic file format are in students' opinion not necessary for them. Last but not least text editors are also used in rather ineffective way. Styles, table of content, bibliography tools as well as mail merge are used very rarely. Last part of this short block consists of elements of Hyper Text Markup Language (HTML) and Cascade Style Sheet (CSS). People who are against this say that nowadays nobody creates web pages in pure HTML and CSS using only text editor. The purpose of these classes is different. First of all this is one of the simplest examples in which it is possible to presents to students how something from the field of IT works. Moreover in contradiction to clicking, dragging, dropping and tapping HTML and CSS require precise thinking which can be treated as an example of algorithmic thinking necessary in programming. Last on the list of pros is the fact that due to existence of numerous validators students can easily check results of their work.

C. Block 2: Spreadsheet - in between…

This block of classes is also three weeks long. There are three major points in this block: logical functions and conditional statements, database functions and their usage and Solver described precisely in [2]. Solving engineering problems especially during design process means using conditions. Simple spreadsheet IF function combined with OR, AND and NOT functions are excellent introduction to programming. Table databases are not real relative databases but they can at least give to students a flavor of database systems and give to them an opportunity to learn how to formulate queries using specialized database functions. Solver opens the opportunity to solve more complicated problems like optimization. Students are taught two things: how to solve linear and nonlinear maximization and minimization problems and how to create appropriate mathematical model of a given problem. Fig. 3 shows Solver window for linear programming problem.

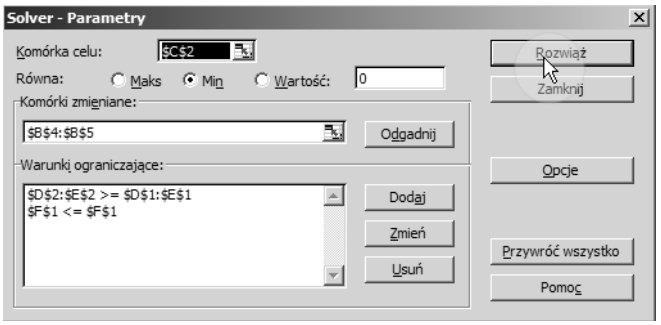


Fig. 3 Solver parameters for linear problem

D. Block 3: CAS – towards algorithmics and programming

The last and the biggest block of classes is devoted to CAS namely to MathCAD which is described in many books like: [3], [4] and [5]. Its presence in curricula of studies is a source of never ending discussions. In the opinion of many teachers students overuse MathCAD while preparing their design homework using it. It is enough that one person creates a file and all remaining can simply enter only data.

First part of MathCAD classes is devoted to solving classical mathematical problems:

- Symbolic calculations
- Definition of variables and functions
- Calculus: integrals, derivatives, limits.
- Matrix and vector operators and functions.
- Solving problems: linear and nonlinear equations, minimization and maximization.

Second part is devoted to programming. In the first part basic instructions (if, for, while) and control statements (return, continue, break) are introduced. The idea of this subject was inspired by the book [6]. List of algorithms is based on two Polish books from this field: [7] and [8]:

- Numerical algorithms (bisection, regula falsi, Newton method)
- Classical algorithms (Euclid's algorithm for greater common divider, Fibonacci numbers)
- Sorting algorithms (insertion, selection, bubble)

These algorithmic problems are solved together with classical programming problems including:

- Matrix and vector operations and
- Sums of series.

Fig. 4 presents sample MathCAD code of insertion sorting algorithm.

```
ins(w) := | n ← length(w)
          | for j ∈ n − 1, n − 2.. 1
          |     rob ← w_j
          |     i ← j + 1
          |     while i ≤ n ∧ rob > w_i
          |         w_{i−1} ← w_i
          |         i ← i + 1
          |     w_{i−1} ← rob
          | w
```

Fig. 4 Sample MathCAD code

III. Information Technologies: Lectures

Lectures are also very controversial part of the subject Information Technologies. There is quite common opinion that they are not necessary because in the second decade of XXI century everybody is a specialist in the field of IS and CS. Lectures are mainly based on four books: [9], [10], [11] and [12]. First one is accompanied by excellent web site enabling students to learn independently. Subjects of subsequent lectures are as follows:

- Introductory remarks: layers of computing system, the history of hardware and software.
- Binary values, number systems, conversions, floating point arithmetic.
- Data representation of text, graphics, audio and video; compression.
- Boole algebra, gates and circuits; von Neumann architecture;
- Elements and parameters of computing system.
- Algorithms and their representation: searching and sorting algorithms, recursive algorithms.
- Programming languages: translation, compilation, interpretation, basic programming structures, programming paradigms and languages.
- Types and data structures: stacks, queues, lists; subprograms.
- Operating system: role, memory and process management.
- File system and directories.
- Information systems and applications: spreadsheets and databases.
- Computer networks and their security: network addresses, cloud computing.
- Internet and the World Wide Web: HTTP, HTML.
- Introduction to artificial intelligence and expert systems.
- Limitations in the field of hardware and software; the untouched problems; questions and answers.

IV. Computing in CE

This subject plays supplementary role to IS and CS. Its curriculum consists of three blocks. First is devoted to Directs Stiffness Method (DSM) enabling students to solve trusses, beams and frames. Second part is devoted to stationary heat transfer problems – students learn how to solve set of Partial Differential Equations (PDE) using Finite Difference Method (FDE) or Finite Element Method (FEM). The last block devoted to optimal design is mainly based on the book [13]. Student learn from this block how to solve Operation Research (OR) problems as well as structural optimization problems. Tools used in curricula of this subject are mainly known from IT classes – this is spreadsheet and CAS software MathCAD. The stress is put on building appropriate mathematical model. In many cases for simple problems solution can be verified by hand calculations. Fig. 6 presents solver window for transportation problem.

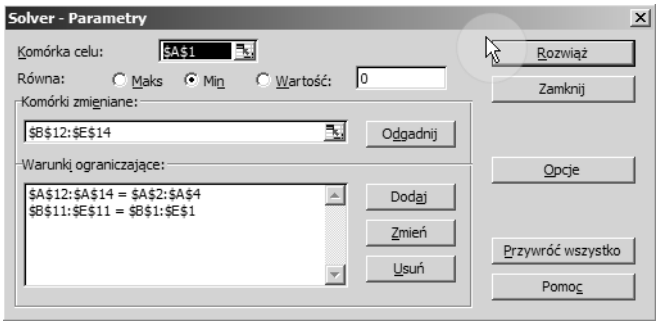


Fig. 5 Solver parameters for transportation problem

V. Towards Flipped Education

For ten years students have been provided with different multimedia materials in the form of podcasts – personal on demand broadcasts. First podcasts prepared in the Division of Information Technologies (DoIT) had the form of screencasts – "digital recordings of computer screen output often containing audio narration". Screencasts contain software animations helping students to learn how to use software. Second kind of podcasts are slidecasts – "audio podcasts combined with slideshow". Slidecasts have the form of knowledge clips – short explanatory presentations of particular problem and its solution. Last kind of multimedia materials prepared by DoIT are webcasts – "media presentations distributed over the Internet using streaming media technology to many simultaneous viewers". In fact webcasts were lecture captures which were recorded and later distributed as podcasts. Tenths of hours of different podcasts stored on ed-

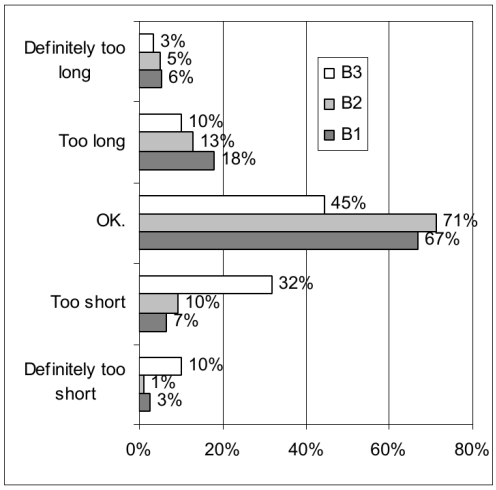


Fig. 6 Opinion about the length of blocks

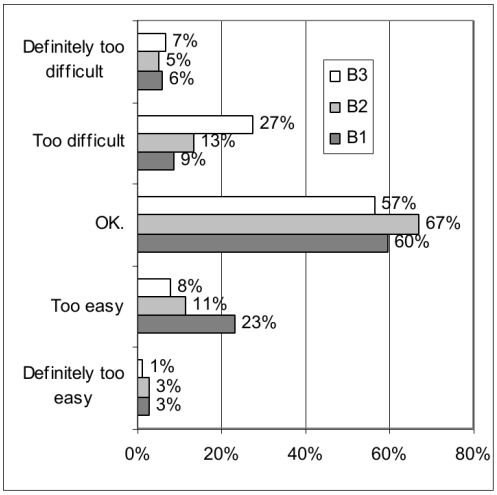


Fig. 7 Opinion about the difficulty of blocks

ucational portal helped a lot during classes but did not have expected impact on quality of learning process measured in terms of grades obtained by students.

Starting from academic year 2012-2013 in some of the groups podcasts are used in different way. Students are asked to watch podcasts at home. During classes they should be prepared to use software without any problems and to solve using it particular problems. This idea is known as flipped classroom and is described precisely in [14] and [15]. First results of this experiment are to some extend promising - students gain better scores in flipped mode. Students are not very keen to spend time at home watching podcasts. They do prefer to "be taught" during classes. This problem can be easily solved by adding simple point to subject regulations – students should be prepared to computer laboratories and this fact is checked by means of test before the class. According to European Credit Transfer System (ECTS) in fact the same amount of time as at the university average student should spend learning at home. It is much more effective to watch passive in nature screencasts at home and solve problems with tutor in class than the other way round.

VI. Conclusions

The question of the place and role of IS (ICT) and CS in engineering curricula belongs to the category of ill posed problems. From the point of view of majority of teachers the most important for engineers is their intuition, so they neglect the knowledge from the fields of IS (ICT) and CS. Results from all computational programs should be verified and this is out of the question. One can say that the role of computations is to prove hand calculations. But on the other hand IS and CS can make engineers work more efficient.

From students' perspective IS tools are treated as specific black box. In order to use these tools in proper way at least basic knowledge form CS area is required.

Results of questionnaires from the end of the course presented on Fig. 6 and Fig. 7 show that the idea of sustainable presence of IS and CS in curriculum of civil engineering studies is accepted by students.

Presented paper shows that further convergence between IS and CS in higher education is necessary, because tools used by engineers are more and more complicated. New and innovative teaching methods like podcasts and evaluation of teaching can help to merge from IS to CS and vice versa. Curricula of studies should be continuously improved by adding new elements like for example Geographic Information Systems (GIS) or Building Information Modeling (BIM) software and systems.

References

[1] J. Kennedy, *Complete Ecdl 5*, Gill & Macmillan Ltd, 2012.
[2] M. Harmon, *Step-By-Step Optimization With Excel Solver - The Excel Statistical Master*, Excel Master Series, 2012.
[3] B. Maxfield, *Essential Mathcad for Engineering, Science, and Math*, Academic Press, 2009.
[4] P. Pritchard, *Mathcad: A Tool for Engineering Problem Solving*, McGraw-Hill Science/Engineering/Math, 2011.
[5] R. W. Larsen, *Introduction to Mathcad 15*, Prentice Hall, 2010.
[6] D. Harel & Y. Feldman, *Algorithmics: The Spirit of Computing*, Addison-Wesley, 2004.
[7] M. M. Syslo, *Algorytmy*. Warszawa: Wydawnictwa Szkolone i Pedagogiczne, 1997.
[8] M. M. Syslo, *Piramidy, szyszki i inne konstrukcje algorytmiczne*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne, 1998.
[9] N. Dale & J. Lewis, *Computer Science Illuminated*, Jones & Bartlett Learning, 2012.
[10] J. G. Brookshear, *Computer Science: An Overview*, Prentice Hall, 2011.
[11] R. T. Watson, *Information Systems*. CreateSpace Independent Publishing Platform, 2012.
[12] R. Stair & G. Reynolds, *Principles of Information Systems*, Cengage Learning, 2011.
[13] J. Arora, *Introduction to Optimum Design*, Academic Press, 2011.
[14] J. Bergmann & A. Sams, *Flip Your Classroom: Reach Every Student in Every Class Every Day*. International Society for Technology in Education, 2012.
[15] J. Gerstein, *The Flipped Classroom: The Full Picture*, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 721–726

Testing the perception of time, state and causality to predict programming aptitude

José Paulo Leal
DCC/FCUP & CRACS/INESC-TEC
University of Porto, Portugal
Email: zp@dcc.fc.up.pt

***Abstract*—The aim of the research presented in this paper is the development of a novel approach to predict programming aptitude. The existing programming aptitude tests rely on the past academic performance of students, on their psychological features or on a combination of both. The novelty of the proposed approach is that it attempts to measure student capabilities to manipulate abstract concepts that are related with programming, namely time, state and causality. These concepts were captured in OhBalls - a physical simulation of the path taken by a sequence of balls through an apparatus of conveyor belts and levers. An engine for this kind of simulation was implemented and deployed as a web application, creating a self-contained test that was applied to a cohort of first-year undergraduate students to validate the proposed approach. This paper describes the proposed type of programming aptitude test, a software engine implementing it, a validation experiment, discusses the results obtained so far and points out future research.**

I. Introduction

PROGRAMMING is hard to learn for most people [9], [12]. Although some students seem to learn it without apparent difficulty, this is not the case of most of them, especially those majoring in subjects other than computer science or software engineering. Even students in those areas are not immune to these problems, as many educators feel that they do not acquire the necessary programming skills in introductory courses [10]. Probably some students are not cut out to be programmers and knowing it in advance would a great advantage [7].

Predicting which students are likely to succeed in learning programming is not easy, some even say it is unfeasible [9]. Nevertheless, this is precisely the goal of the research described in this paper: to create a programming aptitude test, and especially to explore new ways to predict programming aptitude.

Several studies suggest that the high school performance in mathematics is the best indicator of a programming aptitude in college [3], [6], although the correlation between both is usually small. The standard explanation is that both mathematics and programing require abstract reasoning, thus a student with a good performance in mathematics during high school is bound to succeed also in college programming courses.

It is indisputable that a connection exists between programming and mathematics, at least in a broad sense. Although it is much more difficult to define mathematics than programming, if we accept that the realm of mathematics is patterns of thought them clearly computer science in general, and computer programming in particular, are deeply connected to this discipline. However, the mathematical knowledge of a high school student is essentially algebra, calculus and basic logic, and these fields lack a number of abstract concepts that are essential in programming, namely time, state and causality.

The concept of *time* is central to computing. Computers have an internal clock that regulates how instructions are executed. Programs are sequences of instructions that are executed in a flow over a period of time. All the elements of a computer, or of a program, are in a certain *state* that evolves over time. The execution of instructions *causes* changes to these states, that are influenced by their previous states.

This type of reasoning, however abstract, is in general absent from mathematics, especially from the branches of mathematics taught in high school. Take calculus for instance. Although the variable of a function may be interpreted as time, the function itself exists beyond time; it always existed and never changes. A derivative may be seen as the amount of change of the function's value but it has no discernible causes or consequences. Or take Boolean logic as another instance, where A implies B that does not mean that A precedes B or caused by B. If A is true then it has always been true and will never change. This timelessness exists also in algebra where "variables" are actually "unknowns", values that can and eventually will be determined by a computation, not values that actually change over time.

It can be argued that some programming languages, namely declarative languages, are closer to mathematics and above these concepts of time, state and causality. Although this is true, these languages are not widely used and certainly not used as much as Java, C/C++ and Python in introductory programming courses [11]. Moreover, although the denotational semantics of these languages may be independent from time and state, their operational semantics depends on them and the programmer must understand them, if not for anything else, to be able to debug programs.

An algorithm is probably the mathematical concept learned before college that most closely resembles to a computer program. However, students learn specific algorithms that they execute, for instance the division algorithm, rather than study algorithms as a topic, without actually creating new algorithms. These differences between mathematics and programming are possibly the reason why a student may reveal

978-83-60810-53-8/$25.00

aptitude for maths but none for programming and vice versa, and why math grades are insufficient to predict proficiency in programming.

The motivation for this research comes from the intuition that there is a kind of reasoning that is specific to programming, that is different from the reasoning required in mathematics. Based on this insight, the goal of this research is to develop a new kind of test based on the concepts of time, state and causality. The test should be self-contained, in the sense that it should not require a person to administer it and should not require any previous knowledge of programming concepts.

The remainder of this paper is organized as follows. Section II reviews the related work on predicting programming aptitude. Section III presents the proposed type of aptitude test and describes a JavaScript engine implementing it. Section IV reports on an experiment to evaluate the proposed type of test. The final section summarizes the research conducted so far and identifies paths for future research.

II. Related work

Predicting programming aptitude is still a challenging task, although this topic is being studied for more than 40 years [1] and its relevance has been well establish for almost a quarter of a century [7].

Most of the recent research in the literature attempts to correlate programming aptitude with factors that are unrelated with programming, either the student academic record or psychological features. As part of the academic performance the most relevant factor is previous math grades [2], [3], [8] although science grades and even average grades have also been investigated. Other plausible factors were also investigated, such as creativity, problem solving aptitude, attitude toward computers, with even lower correlations [6]. None of these factors has yet provided a good predictor of programming aptitude.

The test recently proposed by Dehnadi [5] differs from the previous since it is actually related to programming; it is a sequence of questions on Java assignment statements. Dehnadi claims that *consistency* in the interpretation rather than its correctness is the main indicator of programming aptitude. The purpose of the test is not to discriminate students who answered correctly, which would assume prior knowledge of programming. The purpose is actually to determine those who answered consistently according to a single mental model. This would reveal the ability to create a meaningful rule to interpreter the assignment command, from which the aptitude to learn a programming language could be inferred. Unfortunately, this experiment was later repeated by Caspersen [4] and also by Wray [12] and none was able to reproduce the results of Dehnadi.

Wray [12] explored the known link between autism and descendants of mathematicians and scientists. He proposed an alternative method for predicting programming aptitude based on mild autistic-spectrum related questionnaires from the Autism Research Center. These questionnaires tests two facets of autism: the level of understanding of systems of objects (SQ) and the level of understanding of other people emotions (EQ). Individually these tests are moderately correlated, but combined they provide a good correlation (r=.67) with programming aptitude. However, the test was applied only to 17 students after they have completed an introductory programming course, and no subsequent results were published on the use of this method to predict programming aptitude.

III. Testing programming aptitude

The goal of this research is to develop a new kind of test to estimate programming aptitude. This kind of test intends to estimate the perception of time, state and causality, under the assumption that these concepts are present in programming reasoning and should reveal an aptitude to program. Also, the test must not require any programming knowledge and be self-contained, in the sense that anyone should be able to take it alone, without or with minimal supervision.

The test that is being developed is based on a set of physical simulations with a common scenario called *OhBalls*. This name comes both from the blue ball than moves on a screen, and the interjection that participants pronounce when it doest not behave exactly as expected. Since it is based on a simulation the test must taken on a computer, which nowadays is hardly a difficulty. In its current implementation the OhBalls test is deployed on the web hence it can be taken virtually anywhere.

A test with the OhBalls scenario is composed by a sequence of panels. Each panel presents a physical simulation set in a room where balls are dropped from a pipe on the ceiling, move through a system of conveyor belts and eventually fall in one of several buckets on the floor. The participant must predict the number of balls that will land on each bucket when the simulation is executed. Figure 1 depicts a example of those panels.

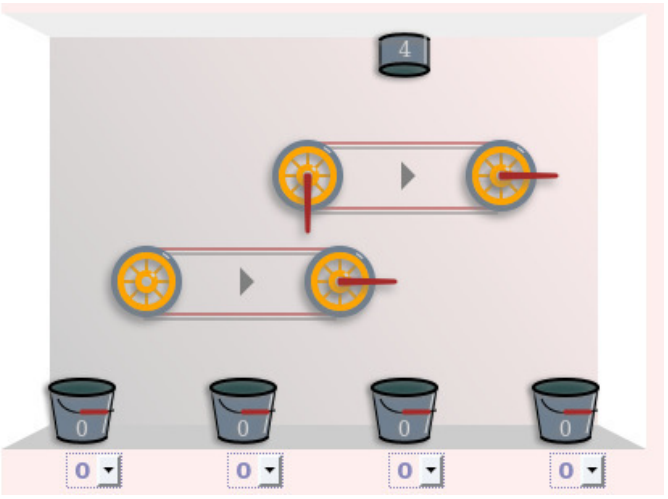


Fig. 1. Example of an OhBalls panel

Balls falling from the pipe land on a conveyor belt that carries them either to the left or to the right. The direction in which the upper side of the belt is moving and the ball would be carried is shown by the arrow head in the middle of the conveyor belt. This direction can be reversed by one of the levers connected to the wheels of the conveyor belt when a ball pushes it. For instance, the first ball falling from the pipe in Figure 1 will be carried to the right by the top conveyor belt and fall in the rightmost bucket. When falling to the bucket it will activate the right lever of the top conveyor belt, reversing its direction, and thus the second ball will go to the left.

The levers have always the same effect – reverse the conveyor belt to which they are connected – but are activated in different ways. For instance, the right lever on the top belt is activated when the ball falls out of the belt, while the left lever on the same belt is activated when the ball is carried to the right by the bottom belt. Hence, the second ball will fall on the second bucket from the right, and will reverses the direction of both belts.

The levers change the state of the system from one ball to the other. In the panel of Figure 1, when the third ball falls from the pipe the top belt will be moving again to the right, while the bottom one will be moving to the left. Hence, this ball will eventually land on the rightmost bucket, raising its count to 2, while reversing the top belt. At this moment both belts will be moving to the left. Thus, the fourth and last ball will be carried to the left, and is dropped on the leftmost bucket. Any subsequent ball would also end in the leftmost bucket but this simulation stops after 4 balls.

The number initially on the pipe indicates the number of balls that will fall during the simulation, and each bucket counts the number of balls that reached it. Under each bucket there is a selector where the participant predicts the number of balls that will reach it when the simulation is completed. The simulation is started by pressing a button (not shown in Figure 1) that is activated when the sum of these selectors equals the number in the pipe.

During the simulation balls fall from the pipe on the ceiling one after another. When a ball is dropped the counter in the pipe is decreased. When the ball reaches a bucket its counter is increased. Then a new ball is dropped from the pipe if this counter has not yet reached 0. When all balls reach a bucket the simulation stops and the number in each bucket is compared with the number in the selector beneath it. The participant is considered to have successfully predicted the outcome of the simulation if these figures match for all buckets. The participant may replay each simulation several times but will not be able to change the answer made before executing the simulation for the first time.

It should be noted that levers do not stop balls by themselves. They are activated by balls and change the direction of belts that carry them through the apparatus. For instance, the left lever on the top belt will be activated when a ball is moving to the right on the lower belt. The lever will not stop the ball in this belt but will reverse the direction of the top belt, affecting only the following balls. If this particular lever (in the left on the top belt) was pointing up instead of down then it would affect balls coming to the left in the top belt. It would not stop them but it would reverse the belt before they reach the point where they would fall.

In summary, the apparatus is composed of one or more conveyor belts that carry a ball falling from the pipe to the buckets on the floor. The motion of each belt is given by a pair of wheels. Levers are always bound to a belt, more precisely to one of its wheels, and can be in 4 possible directions (left, right, up and down).

It is obvious that many panels of this kind can be created with a different number of belt and lever settings. For instance, with a single belt there are 16 different combinations. On each side there are 4 possibilities for placing a lever: no lever at all, pointing up, pointing down, left or right, depending on the position of the wheel[1]. More than one lever per wheel would be possible but would also be too confusing. In a panel with 2 belts these must be arranged so that falling balls go either to another belt or to a bucket.

The easiest way to achieve this is to use a *grid* to place the center of the belts, the buckets and the pipe. The distance between the wheels of a belt must be set in a way that balls are dropped in alignment with the center of other belts and buckets positioned bellow them. Also, the distance between consecutive rows must take in consideration that a certain gap between belts, large enough for the ball to move between them, and small enough for the ball carried by the lower belt to activate a lever in upper belt. The pipe should be at the center and have a belt aligned bellow.

With this approach a setup with 2 belts in two different ways can be created, with the lower belt either to left or to the right, with a total of 512 possibilities. It would not make sense to place belts exactly over each other, or any relative position where balls would not go from one to the other.

An engine to execute this kind of simulation was implemented in JavaScript using the HTML 5 canvas element with a 2D context. This engine runs on a recent version of all major web browser. It can be parametrized with any number of panels following the approach described above. Currently the panels are limited to a grid of 5 rows by 7 columns, which is large enough to place 4 belts, a pipe and bottom row of buckets in each column of the grid.

The current version of engine supports only the prediction of the simulation outcome. In a future version it should allow the participant to place levers in order to achieve a certain configuration. Obviously, this is much closer to the reasoning involved in programming than the current implementation, which is comparable to tracing a program for debugging it. Implementing this feature is not difficult. The main reason for not having it in the first version was lack of knowledge on how the participants would react to this kind of test and the possibility that they would find it too complex as it is.

[1] a lever pointing inwards would be senseless

IV. EXPERIMENT

An experiment was designed to investigate how potential participants perceive the OhBalls type of test and its effectiveness in predicting programming aptitude. For this experiment a number of students enrolled in an introductory programming course took an OhBalls test and the outcome was compared with the grades of their middle term exam. This section presents the web application developed as the main instrument for this experiment, analyses the data collected with it and discusses the obtained results.

The web application developed for the experiment is based in the simulation engine described in section III. It allows a considerable number of participants to take the test simultaneously, while it collects data for later processing.

The interaction of each participant with the web application proceeds in four stages: identification, questionnaire, tutorial and test. The total time of each participation is about 20 minutes. To start the participation each student introduces his or her ID that is checked against a list previously loaded into the application, ensuring that each student participates only once. After being identified, the participant completes a small questionnaire with demographic data, mathematics and average grades from high-school, and former experience with computer and programming.

The OhBalls test is preceded by a tutorial that explains how it works. The tutorial runs on the same type of interface and highlights each important part while a text in a message box provides the necessary details. It explains how the balls are carried by belts in the simulation, how they activate levers and these change the direction of belts, how they reach buckets and new balls repeat the simulation until a predefined number of balls is processed trough the simulation. This tutorial explains also that the participant must predict the number of balls ending in each bucket before running the simulation, and how to activate it and proceed to the next panel. This tutorial runs in a loop until the participant decides to start the test. During the test the participant may rerun this tutorial, if needed.

The OhBalls type of test configured for this instrument consists of a sequence of 30 panels. Each panels is accompanied by a small text that emphasizes a particular point that was not present in the previous ones, such as “note that levers may by activated while balls are falling”.

The first set of panels has a single conveyor belt, and each panel is increasingly more complex than the previous ones. The following set of panels has two belts also with increasing complexity. Nevertheless, the first panels with two belts are less complex than the last ones with a single belt, since they have no levers or just a single lever. The following two sets, with three and four belts, are ordered in the same fashion: the first panels are fairly easy and the last are more difficult.

After the simulation is run the participant is informed if she succeeded in predicting the outcome of the simulation, the time she took to complete it (the number of seconds the panel was shown before pressing the button to start the simulation) and the percentage of correct answers.

When the participant proceeds to the next panel the application sends the data it collected to the server. The data collected for each panel includes the time the student took to complete it, the number of balls the student expected in each bucket and a Boolean indicating if the outcome of the simulation was predicted with success or not.

The participants were students enrolled for the first time in an introductory programming course. This course is common to the computer science and computer engineering programs offered by the computer science department of the faculty of sciences at the university of Porto. The course syllabus is problem solving oriented and uses C as programming language.

The experiment took place in September of 2012 during their first practical class and the participation was optional. Although no student refused to participate in the experiment, those that were unable to complete the test due to timetable constraints were excluded. The students received a brief explanation on the purpose of the experiment and were assured that their participation would not have any impact on their course grades.

The number of participants in the OhBalls test was 153 of which 115 where considered valid. Of these students a considerable number decided they were not ready to take the middle term exam. Only the data referring to the 57 students that took also the middle term exam was used in this experiment. Of these 57 students considered in the experiment the number of females was 10 (17.5%) and 18.16 was the average age.

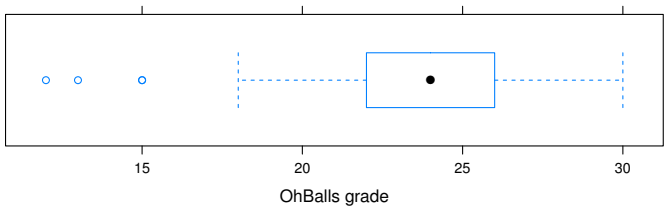


Fig. 2. Box plot of experiment grades

The time taken in each panel by the participants varied from 1 to 558 seconds, with a mean of 33.25 seconds. A possible inverse correlation between time spent analysing the panel and the a correct prediction was investigated, but it was not very high ($c = -0.24$).

To measure the outcome of each participant’s test a grade was computed by assigning 1 point to to each panel correctly answered and 0 otherwise. Thus, each participant had a grade with range 0 to 30 (the number of panels) assigned to her. Considering the series of test grades, the minimum grade was 12, the mean 23.3, the median 24 and the maximum 30. A 5 value summary of the ObBalls grades in the experiment is shown in Figure 2.

There was a good number of very simple panels, to make sure the participants understood the test, but most likely this difficulty was overestimated. In fact, taking 1 for a panel correctly answered and 0 otherwise, the overall median of was 1 and the mean 0.74; by panel, in 23 out 30 the median was 1. The data suggests that the OhBalls test used in the experiment is too simple and more complex panels are needed.

The correlation of the OhBalls grade with the middle term exam was not high ($c = 0.31$) and inferior to the correlation between the high school math grade reported by the students ($c = 0.39$). Nevertheless it was possible to identify a subset of 8 panels for which the correlation is comparatively high ($c = 0.54$). Figure IV is a plot of the the grade for this selected set of panels (as percentage) and the grade in the middle term exam (also as percentage).

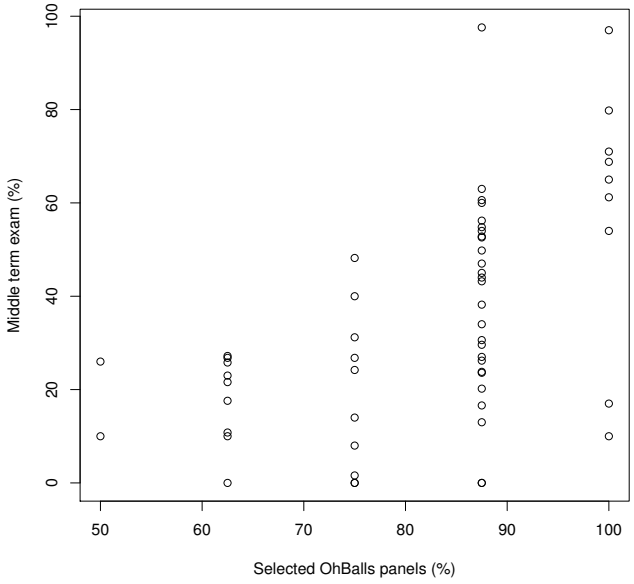


Fig. 3. Scatter Plot of selected panels and mid-term grades

Figure IV shows that the selected OhBalls panels predict the maximum grade the student will obtain, i.e. the student grade is almost always lower than the grade obtained in the selected panel of OhBalls.

The grade in the selected panels was also used in complement with other factors that are known to have influence in programming aptitude, namely math and other high school grades. The initial questionnaire collected the average grade used by Portuguese state universities to rank students applying to their programs. This average includes in equal parts the high school average grade and the national exams grades in particular subjects. In this case these subjects can be either math alone or math, physics and chemistry. A sum of equal parts of these 2 grades (selected OhBalls panels and average grade) reached a higher correlation (c=.64). By comparison, the average grade alone obtained a smaller correlation (c=0.57). The scattered plot of these combined grades with the mid-term grades, with the regression line in red, is presented in Figure 4.

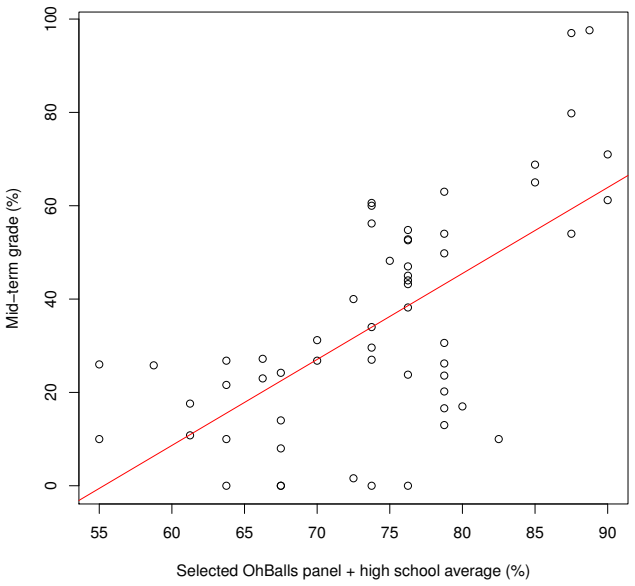


Fig. 4. Scatter plot of selected panels + average and mid-term grades

In any event, these are just preliminary results suggesting that there is room for improvement. The OhBalls panels used in this test were clearly too simple and more complex ones should be used. The capacity for the participants to understand the test seems to have been underestimated. With the current implementation of the OhBalls engine it is easy to create more complex panels. However, new types of panels where the participant must place levers in order to achieve a certain outcome (number of balls in each bucket) should also be used.

The experiment showed also that the OhBalls type of test is able to engage students. They were much more quiet and focused when they were taking the test than they were in the rest of the class. This kind of test has a game-like quality that motivates students to complete it, which is a requirement if students have to take the test on their own without supervision.

V. Conclusion and Future Work

This paper presents a novel approach to estimate programming aptitude based on the understanding of the concepts of time, state and causality. The proposed type of test is named OhBalls and does not require any prior knowledge of programming since it is based on physical simulations displayed on a a web application. The object of the simulation is the path of a sequence of balls trough an apparatus of conveyor belts and levers, until they reach a bucket, which is simple to understand by any undergraduate student. The test is self-applicable, in the sense that it does not require the presence of a person supervising its application.

An OhBalls test was applied to a cohort of computer science and software engineering undergraduate students. The initial results are promising but reveal that more work is still needed to fine tune the test. The average results are comparatively high, suggesting that larger number of panels, and more difficult panels ones, are necessary. Still, a subset of the panels from the current version has a reasonably high correlation with the student intermediary grades.

The main conclusion is that the OhBalls tests must have more panels and more difficult ones, in order to discriminate better the students with higher understanding of time, state and causality. Moreover a new class of panels will be added with a different type of challenge. Instead of simply predicting the number of balls reaching each bucket, considering the influence of the levers, the participant will have to position levers bound to the belts to achieve a certain configuration of balls in the buckets. The kind of reasoning involved will be closer to programming, since currently it can be considered closer to debugging.

In the continuation of this research the methodology of the experiments will have also to be changed. In the experiment presented in this paper only the students that took the middle term test were considered and the students that dropped out where ignored. This "negative" information will be taken in consideration when comparing with the final results.

The current results obtained with the OhBalls test used in the experiment need to be checked against not only the final course grades but also with other programming courses that these students are going to take in the following semesters. To prove the effectiveness of OhBalls test the kind of experiment presented in this paper must be repeated in computer science programs with different pedagogical approaches, in different universities and countries.

ACKNOWLEDGMENT

The author wishes to thank to the students that voluntarily participated in the test reported in this paper, as well as to the lecturer and teaching assistants of the introductory programming course where it took place. This work is in part funded by the ERDF/COMPETE Programme and by FCT within the FCOMP-01-0124-FEDER-022701 project.

REFERENCES

[1] Carol Ann Alspaugh. Identification of some components of computer programming aptitude. *Journal for Research in Mathematics Education*, 3(2):pp. 89–98, 1972.

[2] Susan Bergin and Ronan Reilly. Programming: factors that influence success. *SIGCSE Bull.*, 37(1):411–415, February 2005.

[3] Pat Byrne and Gerry Lyons. The effect of student attributes on success in programming. In *Proceedings of the 6th annual conference on Innovation and technology in computer science education*, ITiCSE '01, pages 49–52, New York, NY, USA, 2001. ACM.

[4] Michael E. Caspersen, Kasper Dalgaard Larsen, and Jens Bennedsen. Mental models and programming aptitude. In *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education*, ITiCSE '07, pages 206–210, New York, NY, USA, 2007. ACM.

[5] S. Dehnadi. Testing programming aptitude. In *Proceedings of the 18th Annual Workshop of the Psychology of Programming Interest Group*, pages 22–37, Brighton, UK, 2006.

[6] Yavuz Erdogan, Emin Aydin, and Tolga Kabaca. Exploring the psychological predictors of programming achievement. *Journal of Instructional Psychology*, 35(3):264–270, September 2008.

[7] Gerald E. Evans and Mark G. Simkin. What best predicts computer proficiency? *Commun. ACM*, 32(11):1322–1327, November 1989.

[8] Annagret Goold and Russell Rimmer. Factors affecting performance in first-year computing. *SIGCSE Bull.*, 32(2):39–43, June 2000.

[9] Tony Jenkins. On the Difficulty of Learning to Program. In *3rd annual Conference of LTSN-ICS,*, Loughbourgh, 2002.

[10] Michael McCracken, Vicki Almstrum, Danny Diaz, Mark Guzdial, Dianne Hagan, Yifat Ben-David Kolikant, Cary Laxer, Lynda Thomas, Ian Utting, and Tadeusz Wilusz. A multi-national, multi-institutional study of assessment of programming skills of first-year cs students. *SIGCSE Bull.*, 33(4):125–180, December 2001.

[11] Arnold Pears, Stephen Seidman, Lauri Malmi, Linda Mannila, Elizabeth Adams, Jens Bennedsen, Marie Devlin, and James Paterson. A survey of literature on the teaching of introductory programming. In *Working group reports on ITiCSE on Innovation and technology in computer science education*, ITiCSE-WGR '07, pages 204–223, New York, NY, USA, 2007. ACM.

[12] Stuart Wray. Sq minus eq can predict programming aptitude. In *PPIG 19th Annual Workshop*, University of Joensuu, Finland, July 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 727–734

Drawer: an Innovative Teaching Method for Blended Learning

Félix Albertos Marco
Computer Science Research Institute
University of Castilla-La Mancha
Albacete, Spain
felix.albertos@uclm.es

Víctor M.R. Penichet, José Antonio Gallud Lázaro
Computer Science Department
University of Castila-La Mancha
Albacete, Spain
{victor.penichet, jose.gallud}@uclm.es

***Abstract*—During the last decade there has been a shift in the way learning process is conducted. One of the main reasons is that technology is changing. Due to this fast movement, concepts like "class", "workgroup" and "learning process" are changing too. Learning processes are going beyond the boundaries of what was known as "class". Face-to-face models get mixed with online environments where students are remotely connected through the Internet. This new approach is called blended learning, and it is aimed at improving learning as well as bringing learning where it was impossible or complicated. Nevertheless, one of the main issues is that teachers need innovative tools that support these different learning models.**

As a consequence, this work is focused on the development of a tool for dealing with the main issues found in blended learning scenarios. It is divided in three phases. First, the blended learning experiences and models of the last decade are reviewed. In a second phase, a tool called Drawer, for supporting the main features of the design and use of blended learning experience is developed. In the last phase, an evaluation is made to assess the outcomes of the new tool.

I. Introduction

INFORMATION Systems are a widespread component of the current society. Computer science, multimedia technologies, telecommunications, Internet and other concepts of the "*digital age*" are essential in a wide range of fields. Information and communication technology allows the creation of tools and infrastructures for information management, data processing and communication with others, both individuals and groups. These tools can be used in almost every activity, including teaching and learning. But tools are only a part of the equation. How to use them and how to put in practice the related concepts may be firstly understood for a successful implementation of these activities. Therefore, experiences about their use are essential for the understanding of how and when to put them in practice.

In the case of education, the curricula are increasingly incorporating a combination of traditional face-to-face learning models and non-face-to-face models (mainly online through the Internet). Institutions and teachers are aware of the potential of using such approaches in the implementation of successful learning experiences. With the emergent technologies and the wide array of technological support at our disposal, there is no point in putting aside blended learning. But setting up blended learning environments is not a trivial task. There are a lot of things to take into account. All the involved stakeholders are crucial when setting up these learning experiences: teachers, students, institutions and academic staff, among others.

As a consequence, the first goal of this work, the study of the different perceptions of blended learning during the last decade, arises. In this period, technology has been leading the evolution of learning as well as teaching processes. But it is important to point out that technology is not the goal; it is only a tool to facilitate the connection between the different elements within the learning process. Pedagogical implications have to be always kept in mind.

When understanding how to use blended learning, teachers will have to choose and use the correct tools. But nowadays there is a gap between the perception of how these tools will look like and how they are built. This is why the understanding of blended learning is important in order to develop tools for helping teachers. And that is the second goal of this work: to gather the knowledge of the study of a decade of blended learning to develop a tool that combines the key elements for supporting blended learning environments. Finally, the third goal of this work is to make an evaluation to ensure that students get benefits through the use of the developed tool.

The paper is organized as follows: firstly, it is presented how blended learning has been understood and used during the last decade. Then, the lessons learned are highlighted. Next the developed tool is described. This tool follows the key elements previously found. Subsequently, a comparative evaluation of the developed tool is presented. And finally, the conclusions and the future work are presented.

II. A decade of Blended Learning

In this chapter different course experiences and models for the design of e-learning and blended learning in the last decade are reviewed. The main objective is to infer the key elements for the design of tools for supporting successful blended learning scenarios. Teachers, through this review, will know different experiences and approaches to apply in their education curricula. They also will be able to use exis-

This research is partially supported by the Ministerio de Economía y Competitividad´s TIN2011-27767-C02-01 project and the JCCM´s project with reference PPII10-0300-4174

tent tools in a different manner, by taking advantage of the depicted course models and experiences.

Valiathan in [1] categorized three different blended models. Skilldriven learning combines selfpaced learning with instructor support to develop specific knowledge and skills. Attitudedriven learning mixes various events and delivery media to develop specific behaviors. Competencydriven learning blends performance support tools with knowledge management resources and mentoring to develop workplace competencies.

Twigg presented new models for online learning improving learning and reducing costs [2]. Six characteristics were found when designing blended courses: 1) whole course redesign; 2)active learning: all of the redesign projects make the teaching-learning enterprise significantly more active and learner-centered; 3) computer-based learning resources; 4) rather than depending on class meetings, student pacing and progress are organized by the need to master specific learning objectives which are frequently in modular format; 5) on-demand help. Enhancing students feel that they are part of a learning community is critical in regard to persistence, learning, and satisfaction; 6) alternative staffing. Not all the tasks associated with a course require highly training and expertise. This work also identified five distinct approaches for course design: a) supplemental model: retains the basic structure of the traditional course, particularly the number of class meetings; b) replacement model: the key characteristic of the replacement model is a reduction in class-meeting time, substituting it with online, interactive learning activities for students; c) emporium model: the redesign model allows students to choose when to access course materials, eliminates all class meetings and replaces them with a learning resource center featuring online materials; d) fully online model: the instructor must be responsible for all interactions; e) buffet model: information technology in teaching and learning means that it can radically increase the array of learning possibilities presented to each individual student.

Aspden [3] asserted how a blended learning approach alters the dimensions of the relationships between the students and the other aspects of their learning experience. The findings reported indicate that the blend itself makes effective engagement in a range of possible situations, allowing students to fulfill their different activities together with more flexibility according to their particular circumstances.

Heinze [4] concluded that face-to-face, blended learning and e-learning are difficult to understand separately, mainly because there are overlaps between then, as depicted in Figure 1. So, the different learning strategies are represented in two axes: use of technology and time spent on online learning. Blended learning is located between face-to-face and online modalities.

Graham [5] described trends and future directions for blended learning systems. In Figure 2 it is depicted the progressive convergence of traditional face-to-face and distributed environments, by allowing the development of blended learning systems. Graham found six major issues relevant to designing blended learning systems: 1) the role of live interaction; 2) the role of learner choice and self-regulation; 3) models for support and training; 4) finding balance between innovation and production; 5) cultural adaptation; 6) dealing with the digital divide.

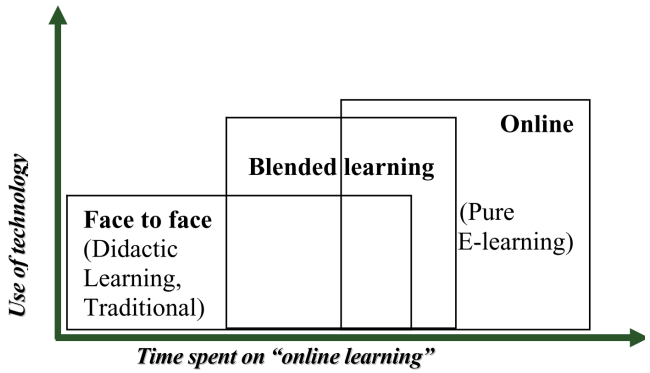


Fig. 1 Learning approaches

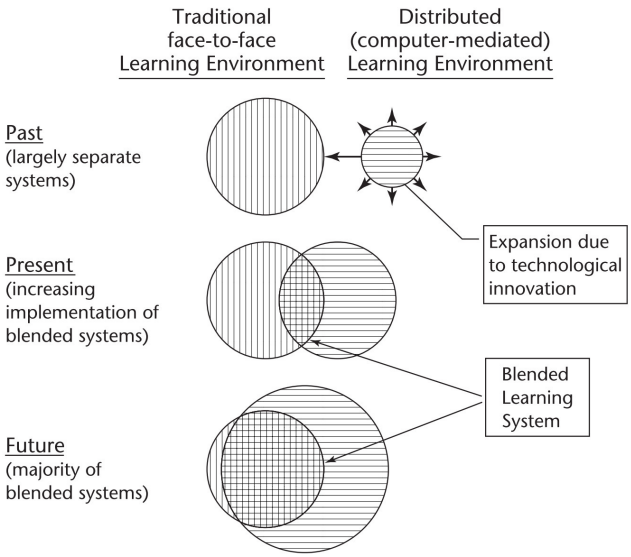


Fig. 2 Progressive convergence in blended learning systems

For Kulvietiene [6] the integration of a virtual classroom into learning managing systems has many advantages: a) opportunity is presented for provided blended learning; b) learning activities including both virtual classroom sessions and learning in a virtual classroom can be managed from a single location; c) information about learning activities is stored in a single location.

Draffan [7] identified the challenges for blended learning from two perspectives: the learner and the teacher. From the learners point of view, the main challenges are: skills, e-skills, preferences, content interaction and design, learning interactions and assistive technologies. From the teachers perspective, the main challenges are: the issue of context, learning design and to facilitate inclusive learning. It is needed to ensure that the students can interact successfully with the technologies, among themselves (through reflection), with their peers, with their teachers, with the support workers and with the learning materials.

Kim [8] presented a survey that found blended learning gained popularity in many organizations but also that several barriers exist in implementing it.

Wang [9] used asynchronous tools for online collaboration and offline interaction between students in blended learning. The offline atmosphere in carrying out the asynchronous computer media communication activities were sorted into five major categories: struggling with platform operations, handling technical problems, passive attitudes towards the procedure, tense atmosphere in class, and engagement in tasks. Blended learning does not automatically help students in their adoption of active learning strategies. The roles should be recognized to promote effective and efficient online/offline interaction.

Dziuban [10] reached a reasonable conclusion: students react generally to the course, the content, the instructor, the learning climate, and themselves. One remarkable and well-known conclusion is that the boundaries of what has been called the "class" are disappearing.

For Khan [11] assessment was, without any doubt, one of the major tools in the teaching and learning process. Assessment is considered an effective tool in determining student's knowledge gain in any particular course they enrolled. Traditional learning is more class oriented and less flexible in terms of class schedule, use of latest technology and learning methodology, while blended Learning is flexible and supports both classroom and online teaching.

Fleck [12] depicted the opportunities and challenges in blended learning communities. They presented the case of The Open University and its explicit social mission to provide educational access to those who were otherwise denied the opportunity for learning. To provide it, different models for blended learning arise throughout the time: a) correspondence and broadcast models: printed course materials sent by surface mail in a correspondence course style; b) purpose-designed quality distance education model: systematic consideration of pedagogic principles, professional editing, and explicit design for effective delivery over a variety of media; c) practice-based model: between learning materials and students; between tutors and students; between student peers; and above all between students and their work colleagues; d) learning community model: thanks largely to the web and readily accessible search procedures, raw information and data are available very easily and increasingly at little or no cost.

In his work, Fleck describes which features the next generation of learning models would have: a) the creation of a learning community; b) emphasis on process & activities, not content and assets; c) use of a wide range of existing & specially designed assets; d) focus on student-driven learning; e) use of Web 2.0 and mobile devices to support communication; f) design of face-to-face residential schools for business networking.

Moskal [13] proposed some questions for an initial blended learning: 1) Why should the institution engage in blended learning? What are our goals and what outcomes do we expect to achieve? 2) What student benefits do we seek? 3) What courses or programs will we offer in a blended format, and why? 4) How will we engage in and support our faculty in order to make them successful? 5) How will we roll out blended learning throughout the institution? 6) What levels of investments are we prepared to make and what returns do we expect?

Graham [14] presented six cases of institutional adoption of blended learning to examine the key issues that can guide university administrators interested in this endeavor. He describes three broad categories for the adoption of blended learning: strategy, structure, and support.

In [15] Bohle proposed four factors as crucial elements for a successful bottom-up change process in blended learning environments: the macro and the micro contexts, the project leader and the project members. He also revealed that bottom-up change process leads to three important outcomes: 1) the development of blended learning programs which match the needs of faculty and learner; 2) incentives for new task forces to solve institutional bottlenecks which only the faculty could have discovered; 3) new knowledge for the institutes.

Taylor [16] identified facilitators and barriers to systemic implementation of blended learning. It was found that, as teaching and learning environments are socially dynamic, strategic institutional change would only happen if there is a shared vision and energy that touches all parts of the organization.

Owston [17] examined the relationship between student perceptions in blended learning courses and their achievement. The overall conclusion of this study is that high achievers are very satisfied with the blended format. On the other hand, lower achievers may not be able to succeed in this learning environment as well.

Taplin [18] analyzed the monetary value students place on having access, via the Internet, to recorded lectures in a blended learning context. The principal results are that the average price students are willing to pay is approximately $30 per equivalent full time student.

III. Lessons Learned

The e-learning approach takes advantage of the benefits that information technologies provide to learning environments. They bring new opportunities in learning environments. But, contrary to the general believe of most teachers, distributing knowledge elements through electronic media is not always enough to take advantage of the e-Learning capabilities. It should not be only regarded as a cheap way of distributing resources to a big number of students. For understanding what blended learning is, how it has been used and how it be implemented in an effective way, next are presented blended learning definitions during the last decade:

- A solution that combines several different delivery methods, such as collaboration software, Webbased courses and knowledge management practices [1]
- Learning which combines online and face-to-face approaches [4]
- Learning that is facilitated by the effective combination of different modes of delivery, models of teaching and

styles of learning, and founded on transparent communication among all parties involved with a course [4] [7]
- Systems that combine face-to-face instruction with computer-mediated instruction [5]
- A combination of various networked technologies in a single learning package; a synthesis of various pedagogic methods that enables to achieve an optimal quality of learning process; a combination of various lecturing technologies (video cassettes, compact discs, internet material, etc.) together with direct lecturing by an instructor [6]
- An approach of combining face-to-face instruction with computer-mediated instruction is called blended learning [9]
- Blended approach studies how to join the best feature of face-to-face and online instruction [11]
- Instructional approach that substitutes online learning for a portion of the traditional face-to-face instructional time [17]
- A combination of online learning and face-to-face approaches to teaching [18]

Other lessons learned from the previous study are the main characteristics to keep in mind when designing blended learning approaches. These characteristics have been inferred from the review done in this research. They are intended to guide teachers, but they could be also interesting from other points of view, such us for measuring the quality and for evaluating courses based on blended learning scenarios. Main lessons learned are: a) Students learn by doing, not by listening to some one talk about doing. There is a wide range of learning approaches, from face-to-face to fully on-line. b) The "right way" to design a high-quality course depends entirely on the type of students involved. c) Students need to be treated like individuals, rather than homogenous groups. d) Effective blend of face-to-face and online learning opportunities have to take into account individual students' particular needs. e) Blended learning will be characterized on how they blend instead of whether. f) To guarantee inclusive learning is fundamental for creating successful blended learning. g) The term "class" goes beyond the boundaries of the physical location. h) One of the key issues is the role of technology. But technology is not an end in it itself, pedagogy must lead.

Through the review of the blended learning models and experiences, the weaknesses and the strengths have been also gathered.

The main weaknesses found in blended learning are:
- Need of effective guidance
- Technical issues
- Lack of communications
- Unsatisfactory use of the face-to-face session time
- Implementation
- Robust and reliable infrastructure is required
- Social interdependence among the participants, the tasks assigned, and the e-learning tools remain challenging for teachers

The main strengths found in blended learning are:
- Compatibility with working life
- Flexibility
- Good student support
- Improved pedagogy
- Increased access and flexibility
- Increased cost-effectiveness
- Information from the face-to-face activities to total online interactions is stored in a single place.
- Promotion of social interaction
- Quick feedback to learners which will help them in their learning process
- It provides collaborative activities among teacher and students
- It allows access to everyone who needs training by providing it in different ways

These characteristics produced a shift in teaching and learning from simple knowledge transmission in which "content" is transferred to the devising of processes and activities that enable deep learning following the "triple A" paradigm: Anytime, Anywhere, Anyone.

This change could be described by the Table I [11], where it is characterized the shift between traditional and blended learning from the point of view of the features of learning.

TABLE I.
SHIFTS BETWEEN TRADITIONAL AND BLENDED LEARNING

Characteristics of learning	**Traditional learning**	**BlendedLearning**
Place	Mainly in classrooms (Not flexible)	Combination of classroom / home, library (flexible)
Learning Methodology	Offline	Offline as well as Online Learning
Time of learning	Fixed as per the schedule (Not flexible)	Adjustable as per personal choice (Flexible)
Use of Technology	Not must up to the instructor to choose the teaching methodology	Latest use of technology is must

Creating scenarios that allow teachers in the process of setting blended environments is still challenging and complex. As a result of the aforementioned characteristics, collaboration and social factors are key aspects when designing these environments. In the Table II, Lambropoulos [19] describes social awareness requirements and propositions.

TABLE II.
SOCIAL AWARENESS REQUIREMENTS AND PROPOSITIONS

Social awareness requirements	Propositions
Embodied self & group presentation	Emoticons, avatars, group network representation
Visibility of social presence and connectedness, locality	Individual nodes, group ties and networks, online status
Social and cognitive awareness	Enhanced discussion forums, group network representation
Depiction of the individual and group locality to indicate the spatio-temporal relationship	Group network representation
Participation measurements	Participation graphs
Lightweightness & interoperability	PHP and JAVA programming languages
Simple to interpret and easy to use	User-centered design

IV. DRAWER

Drawer is a web application for helping teachers to design assessments. The main characteristics are the support for collaborative tasks, the sharing of information and the management of social interactions between users. Drawer has the advantage that all these functionalities are integrated in the same environment, making it easy to use. They are easily accessible, being easy to create learning experiences.

The application was made bearing in mind the findings made in the previous phases of the project. So, the key elements for developing successful blended learning experiences arise. To support them, the application manages the following elements: a) users and their relationships: user profiles, creation of groups, personalized shared workspaces; b) synchronous and asynchronous communication; c) information is stored in a single place: files, logs, conversations and other information.

The management of users and their relationships is made by the application. Drawer includes mechanisms for controlling the authentication and access to the application. Only registered users are allowed in the system. The main screen of the application is depicted in Figure 3. There, users are allowed to log in the system or to create a new account.

Once the user is logged in the system, the main screen for logged users is presented (Figure 4). There, there are two main areas. The main menu, that is located at the top of the screen, and the rest of the interface. The main menu has the following sections: start, alerts, mailbox, user information and search.

Fig. 3 Main screen of Drawer

The *"start"* section corresponds to the screen depicted in Figure 4. There, users view the available drawers. Drawers are stacked by categories for better organization or for grouping users. Users are allowed to create new drawers or stacks, simply by clicking the "plus" symbol located in the right side.

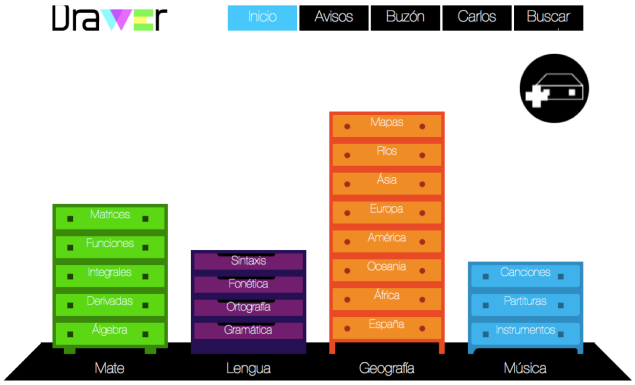


Fig. 4 Main screen for a logged user

One of the main characteristics of Drawer is the way it shares information among users. It follows the paradigm set in Drag&Share [20] allowing users to drag documents from his local devices to the shared workspace. As depicted in Figure 5, when users enter in a drawer, the shared workspace for that drawer is presented. There are allocated the resources and other users, as well as the synchronous communication means. All this in a single view, making it easy to follow others work and work in the tasks. Users are represented by their names over the shared workspace, showing their movements in real time. Each user has a representative color in the system for the chat and his cursor (the name over the shared workspace). Resources are represented in the shared workspace by the name and a representative icon. In the shared workspace, other users can be invited to perform collaborative tasks.

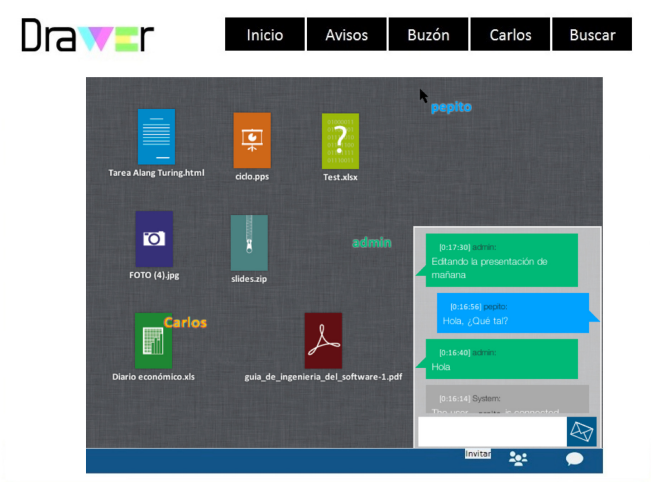


Fig. 5 Shared workspace

Users can perform actions over the resources by making right click or a double touch over them. These actions are depicted in the Figure 6. They can create new documents with the included rich text editor. The text editor allows the creation of documents with enriched text. Also images, videos and any multimedia resources can be inserted within these documents. The next option is to download documents from the shared workspace to the local device. The edit option allows users to edit existent documents in the shared workspace. The preview option shows the selected document. To delete documents, users only have to select the delete option. Finally, users can hide documents selecting the corresponding option, preventing other users to view the document in the shared workspace.

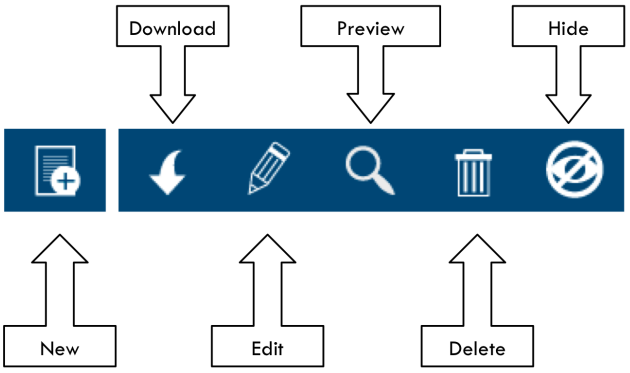


Fig. 6 Options in the shared workspace

The *"alerts"* section gives users information about the events produced in the system. There, they can see new friend requests, pending messages and request for resource sharing. The *"mailbox"* section, as depicted in Figure 7, allows the asynchronous communication between users. In this section, the active conversation between users or groups could be found. The section *"user"* contains personal information about the logged user. Also the information about the resources and interactions the user has made with the system is represented.

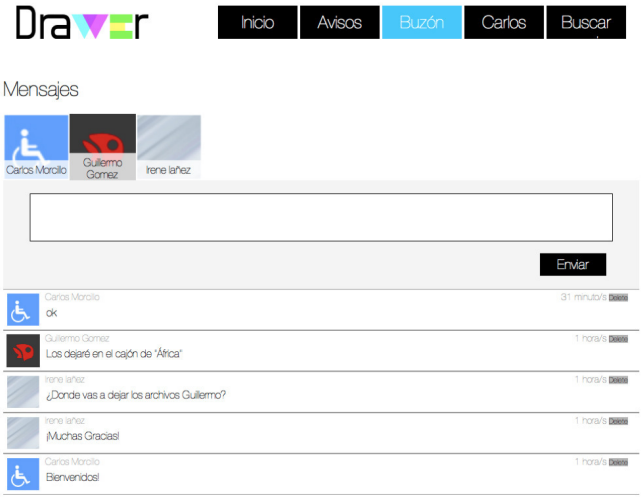


Fig. 7 Mailbox section

The section *"search"*, as depicted in Figure 8, allows searching for other users. This section is designed for viewing other users and to send them a friend request. This request may be accepted or rejected. Friends in Drawer are allowed to easily share information, send messages and perform collaborative tasks.

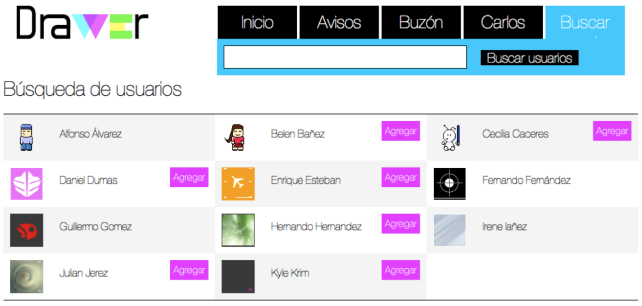


Fig. 8 Search users option and friendship management

V. Evaluation

Through this evaluation, we want to assess the impact in a blended learning activity when using Drawer. An activity is done by using the means provided by Moodle by default and using Drawer. The aspects to be assessed are how the proposal affects the productivity in the task, that is, how much time users expend to complete it, and how usability of the system is affected. Therefore, the evaluation is focused on the level of productivity and the user's satisfaction while using the system. Tasks time has been used to measure productivity and satisfaction has been measured using a questionnaire based on SUS (System Usability Scale) test [21]. Time, as productivity measure, has been selected because it provides a good insight on the impact when performing tasks using different systems. The SUS test has been chosen because it has proved to be a valuable evaluation tool, being robust and reliable.

The group of selected students to perform the evaluation has the following features. Seven students make up the

group. Four students are males and the other three are females. The participants are nearly 25 years old on average. The oldest and the youngest user are 28 and 22 years old, respectively.

To perform the evaluation it is selected a learning scenario where the teacher sends an assessment to a group of students in a blended learning scenario. Some of the students are in the same physical room, but other students aren't. They have to perform the task collaboratively. Students have a device for accessing the web application. The activity consists in an assessment where the students have to make a summary of a document provided by the teacher. Each student is responsible of making a part of the summary. They have to choose a leader responsible of joining the individual summaries. As a result, they will get the final summary. The task is divided into the following five subtasks: 1) the text is distributed, selecting the parts to be done by each student. At this time, the leader is selected among the students; 2) partial summaries are made and send to the leader; 3) the leader gathers the partial summaries into the final summary. It is send to all the members of the group; 4) the group reviews the final summary and decides if it is ended; 5) the leader sends the teacher the final summary.

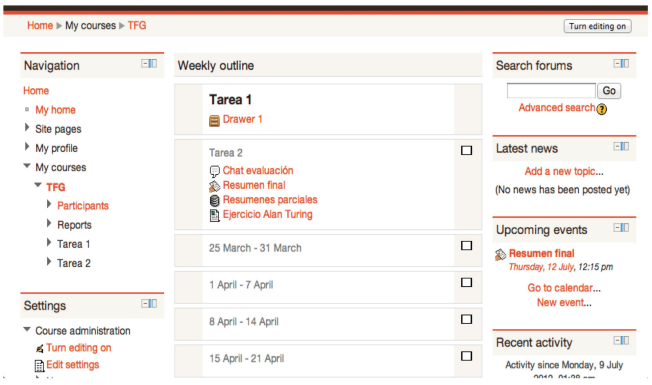


Fig. 9 Moodle setup to perform the task

The teacher is in charge of setting up the scenario for performing the activity. In Figure 9 it is depicted the scenario for performing the activity in Moodle. The scenario in Moodle is composed by the chat, a database for sharing partial summaries, and one link to upload the final summary. The scenario in Drawer is depicted in Figure 10. At a glance, this is simpler for users than the one used in Moodle.

The productivity of the system is analyzed based on the time spent to perform the collaborative task. This time is divided into the five tasks described above. The average time is shown in Figure 11 for the test developed in Moodle with default activities, and in the Figure 12 for the test developed in Drawer. The average time decreased drastically in four of the five measured tasks when Drawer was used.

Regarding student's satisfaction, the SUS satisfaction questionnaire has been used. In this test, users have to express their agreement with 10 sentences after performing the task. For each sentence, a score between 1 and 5 is given, meaning 1 strongly disagreement and 5 strongly agreement. Then, based on these values, the SUS satisfaction questionnaire final value is calculated. This value can be between 0 and 100. A final value near 100 indicates a complete satisfaction. In the performed test, the final value in Moodle with default activities was 33.10, which indicates that students were not satisfied with the system. The final value in Drawer was 84.2, which confirms that users were very satisfied when using it.

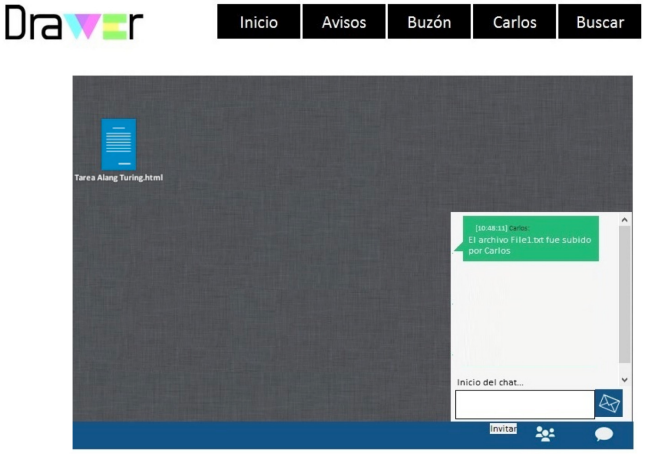


Fig. 10 Drawer setup to perform the task

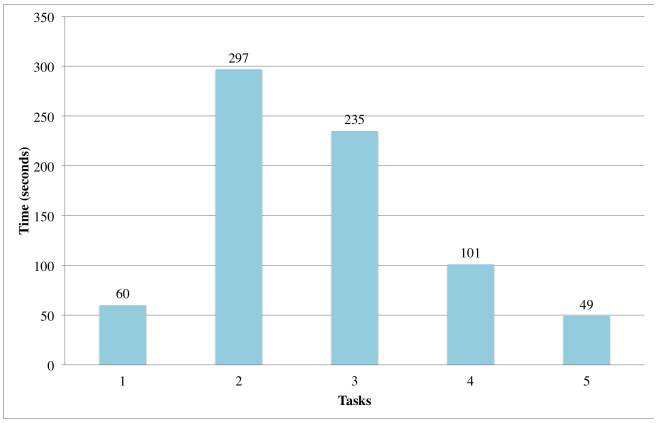


Fig. 11 Average times on each task with Moodle

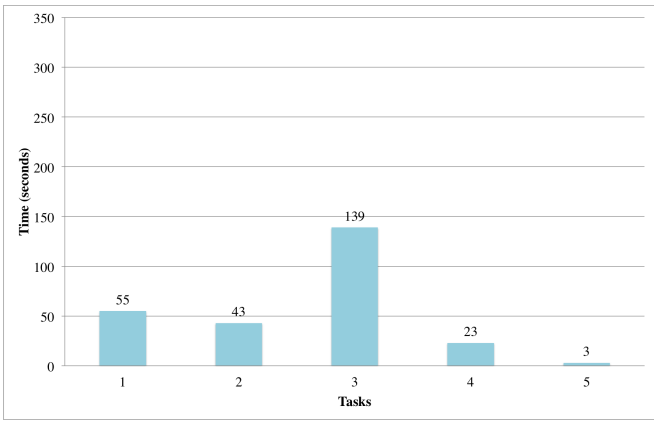


Fig. 12 Average times on each task with Drawer

At the end of the test, users were invited to write their personal impressions about the tools. Next, some of the comments made when using Moodle in the evaluation are included:

- "I don't know where I have to upload the file"
- "When I manage to upload the file, I don't know if I did it well. I have to search on the list where all the files were shown to see if mine were there"
- "It is tedious, you have to download the final file, search for it and then open it"
- "Because the chat was opened in other window, there was a moment when other users sent messages to me but I didn't realized it"
- "Workgroup was complicated"

In the other hand, comments made when using Drawer in the evaluation were:

- "This tool is very useful for workgroup"
- "Intuitive and easy to use"
- "You can open and edit files within the application"
- "You are always aware of what is going on"

VI. Conclusion

Teachers need innovative tools for supporting new learning experiences. There are many issues to bear in mind when designing applications for supporting the curricula. So, this work is intended to present the main characteristics of blended learning approaches as well as the weaknesses and strengths found in a review of blended learning experiences and models of the last decade. A new web application has been designed and implemented taking into account the lessons learned.

Through the review the main characteristics that blended learning systems shared during the last decade have been recollected. We can summarize that blended learning is a wide area between a face-to-face and a fully on-line environment where students need to be treated as individuals and each institution has to understand the right model for delivering blended learning. Moreover, the fact that learning is more productive when doing than when listening how to do is taken into account. Finally, it is important to understand that technology is important, but pedagogy must lead.

With Drawer we make a contribution for dealing with the weaknesses present in most of the studied blended learning experiences: need of effective guidance, lack of communications, unsatisfactory use of the face-to-face session time and social interactions. The evaluation shows that Drawer improves the selected activity performed in blended learning scenarios compared to Moodle. The productivity and the usability have been taken as indicator for measuring the outcomes in one of the main tool in learning scenarios: assessments. Both indicators reflect a drastic improvement when using Drawer instead of Moodle.

As a future work, we want to introduce Drawer within the curricula of educational centers in order to receive feedback for the improvement of blended learning support.

References

[1] Valiathan, P. (2002). Blended learning models. Learning circuits, 11–14. Retrieved from http://old.astd.org/LC/2002/0802_valiathan.htm (Last access on May 21, 2013).

[2] Twigg, C. (2003). New models for online learning improving learning and reducing costs. Educause Review, (October).

[3] Aspden, L., & Helm, P. (2004). Making the Connection in a Blended Learning Environment. Educational Media International, 41(3), 245–252. doi:10.1080/09523980410001680851.

[4] Heinze, A., & Procter, C. (2004). Reflections on the use of blended learning. Retrieved from http://usir.salford.ac.uk/1658 (Last access on May 21, 2013).

[5] Graham, C. R. (2006). Blended Learning Systems: Definitions, Current Trends, and Future Directions. The Handbook of Blended Learning: Global Perspectives, Local Designs (pp. 3–21).

[6] Kulvietiene, R., & Sileikiene, I. (2006). The Blended Learning Delivery Design Model. Conference on Distance Learning and Web Engineering (pp. 1–5). Retrieved from http://labplan.ufsc.br/congressos/WSEAS/papers/517-192.pdf (Last access on May 21, 2013).

[7] Draffan, E. a., & Rainger, P. (2006). A model for the identification of challenges to blended learning. Alt-J, 14(1), 55–67. doi:10.1080/09687760500479787.

[8] Kim, K., Bonk, C., & Oh, E. (2008). The present and future state of blended learning in workplace learning settings in the United States. Performance Improvement, 47(8), 5–17. doi:10.1002/pfi.

[9] Wang, M. (2010). Online collaboration and offline interaction between students using asynchronous tools in blended learning. Australasian Journal of Educational Technology, 26(6), 830–846. Retrieved from http://ascilite.org.au/ajet/ajet26/wang.html (Last access on May 21, 2013).

[10] Dziuban, C., & Moskal, P. (2011). A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments. The Internet and Higher Education, 14(4), 236–241. doi:10.1016/j.iheduc.2011.05.003.

[11] Khan, A. I., Qayyum, N., Shaik, M. S., Ali, A. M., & Bebi, C. V. (2012). Study of Blended Learning Process in Education Context. International Journal of Modern Education and Computer Science, 4(9), 23–29. doi:10.5815/ijmecs.2012.09.03.

[12] Fleck, J. (2012). Blended learning and learning communities: opportunities and challenges. Journal of Management Development, 31(4), 398–411. doi:10.1108/02621711211219059.

[13] Moskal, P., Dziuban, C., & Hartman, J. (2012). Blended learning: A dangerous idea? The Internet and Higher Education, 1–9. doi:10.1016/j.iheduc.2012.12.001.

[14] Graham, C. R., Woodfield, W., & Harrison, J. B. (2012). A framework for institutional adoption and implementation of blended learning in higher education. The Internet and Higher Education, 1–11. doi:10.1016/j.iheduc.2012.09.003.

[15] Bohle Carbonell, K., Dailey-Hebert, A., & Gijselaers, W. (2012). Unleashing the creative potential of faculty to create blended learning. The Internet and Higher Education, 18, 29–37. doi:10.1016/j.iheduc.2012.10.004.

[16] Taylor, J. a., & Newton, D. (2012). Beyond blended learning: A case study of institutional change at an Australian regional university. The Internet and Higher Education, 18, 54–60. doi:10.1016/j.iheduc.2012.10.003.

[17] Owston, R., York, D., & Murtha, S. (2012). Student perceptions and achievement in a university blended learning strategic initiative. The Internet and Higher Education, 18, 38–46. doi:10.1016/j.iheduc.2012.12.003.

[18] Taplin, R. H., Kerr, R., & Brown, A. M. (2013). Who pays for blended learning? A cost–benefit analysis. The Internet and Higher Education, 18, 61–68. doi:10.1016/j.iheduc.2012.09.002.

[19] Lambropoulos, N., Faulkner, X., & Culwin, F. (2011). Supporting social awareness in collaborative e-learning. British Journal of Educational Technology, no–no. Doi:10.1111/j.1467-8535.2011.01184.x.

[20] Albertos, F., Penichet, V. M. R., & Gallud, J. A. (2012). Collaboration within Moodle: Sharing Documents in Real-time with Drag & Share. Proceedings of the Interaction Design in Educational Environments (IDEE 2012), ICEIS.

[21] Brooke, J. SUS - A quick and dirty usability scale. In Usability Evaluation in Industry (1996).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 735–738

Computer Science E-Courses for Students with Different Learning Styles

Olga Mironova, Irina Amitan,
Jüri Vilipõld, Merike Saar
Faculty of Information Technology, Department of Informatics, Chair of Software Engineering, Tallinn University of Technology, Akadeemia tee St. 15A, Tallinn 12618, Estonia
Email: {olga.mironova, irina.amitan, juri.vilipold, merike.saar}@ttu.ee}

Tiia Rüütmann
Faculty of Social Sciences, Department of Industrial Psychology, Estonian Centre for Engineering Pedagogy, Tallinn University of Technology, Akadeemia tee St. 3, Tallinn 12618, Estonia
Email: tiia.ruutmann@ttu.ee

***Abstract*—E-learning is a contemporary teaching tool that has become popular and widely used in engineering education in recent years. This article presents the outcomes of a study on considering students` different learning styles in teaching information and communication technology using e-learning. Students were divided into two study groups. The reference group studied according to a provided learning model which including both theoretical educational material and practical assignments. Students of the test group were divided according to their learning styles using the Felder-Silverman model. Different relevant learning models, which included the same theoretical material and practical assignments, were designed for students of the test group based on the learning styles. The results of the study proved that the learning materials which were designed taking into account students` different learning styles considerably improved the achievement of the learning outcomes. A detailed description and analysis of the study is presented in the article.**

I. Introduction

ENGINEERING education is a large system and it is almost impossible to predict its behaviour over far too distant future since the system parameters show a high rate of change. All knowledge is changing so fast that we cannot give students what they will need to know tomorrow. Instead, we should be helping them develop their learning skills so that they will be able to learn whatever they need to. If we can achieve that, we will have world-class engineers, people who are innovative and resourceful.

Learning styles are characteristic cognitive, affective, and psychological behaviours that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment. Students learn best when instruction and learning context match their learning style.

Understanding students' different learning styles is one of the midpoints of effective education. The aim of the research described in the article was to abolish mismatches between students' common learning styles and teaching styles in e-learning and make teaching in engineering more effective.

According to Felder and Brent [1], students learn in many ways – by seeing and hearing; reflecting and acting; reasoning logically and intuitively; memorising and visualising; drawing analogies and building mathematical models.

Classroom activities of teachers and students take place in mutual communication. Therefore, the guidance and the formative role of the teacher should be realized in the creation and review of theoretical material and the material in practical classes. However, most of the learning processes are individual learning activities and here self-regulation of the student is realised. The task of the teacher in this case is to provide students with a supportive learning environment: motivate, guide and support. It should be noted that learning should be based on individual personality traits. This ensures successful acquisition of knowledge.

II. Methodology

Since 2010, we have applied a flexible, adaptive approach to teaching computer science in Tallinn University of Technology. The main idea of this method was students division into groups according to their prior subject knowledge. The tasks were also of different level and it has given visible results – the level of knowledge has increased [6]. In teaching we have been focused our attention on activating an individual student's learning.

Students learn in different ways: some like to listen to and talk, while the others prefer to read texts or study by investigating the charts, diagrams and drawings. Any learning style can give good results if it is timely identified and a right approach is chosen and applied.

Teaching must transfer knowledge and support learning, but it must also be cooperative and directed toward students' reflection and development. Helping students in finding and forming their own style of learning – should customize the learning process aimed at creating the conditions for each student for the maximum development of his/her abilities, aptitudes, satisfaction of cognitive needs and interests.

Since the beginning of the fall semester 2012, we have conducted experiments in which we have tried to identify the most suitable learning activities for students, based on an individual test on learning styles. 300 students of economics, social and technical disciplines have been involved in the experiment. In the e-environment Moodle (https://moodle.e-ope.ee/) students were divided into two equal groups of 150 participants: a reference group and a test group.

978-83-60810-53-8/$25.00

Students of both groups were taught the informatics courses depending on their prior knowledge: a test was carried out dividing them into beginners, advanced, and experts users. For beginners – the test result was 0% – 60%; for advanced – the test result was 61% – 80%, for experts – it was 81% – 100%. The test contained a different number of computer science related tasks with different difficulty levels.

The system and its effectiveness have been described in the article about a flexible approach to learning [6].

In addition, for the students of the test group all course materials and the whole learning process was designed to match their learning style preferences identified in the test [7].

Felder divides students based on their perception of the material and work with it into the following groups [2]:

active (ACT) and reflective (REF)
sensing (SEN) and intuitive (INT)
visual (VIS) and verbal (VRB)
sequential (SEQ) and global (GLO)

Active learners acquire new knowledge best by doing, discussing and explaining it to others in a group. At the same time reflective learners first think about it alone.

Sensing learners like learning facts and solving problems by well-known methods. Intuitive learners prefer discovering new possibilities and relationships and they are more innovative.

Visual learners remember pictures, diagrams, charts and video best. Verbal learners prefer written and spoken explanations.

Sequential learners like step by step studying, where each step follows logically from the previous one. Global learners prefer to get information by large portions and randomly.

The preferences of students, based on tests carried out among the students of the test group, are shown in Table I. The total for each student is 400% as each student could account for four different forms of information acquisition.

Data from Table I is shown in the following diagrams.

Tests carried out have shown that the majority of students do not have any preferences in the selection of learning materials and that they use a combination of different learning styles– they are well balanced (Fig. 1).

Figure 2 shows the types of students who acquire material better if their learning style has been taken into account. So, this group of students learns better if they are given possibilities to participate in group work, discuss, solve real tasks based on facts, etc.

However, some students have very strong preferences in learning. As presented in Figure 3, according to the tests active learners, sensing learners and visual learners fall into this group [4].

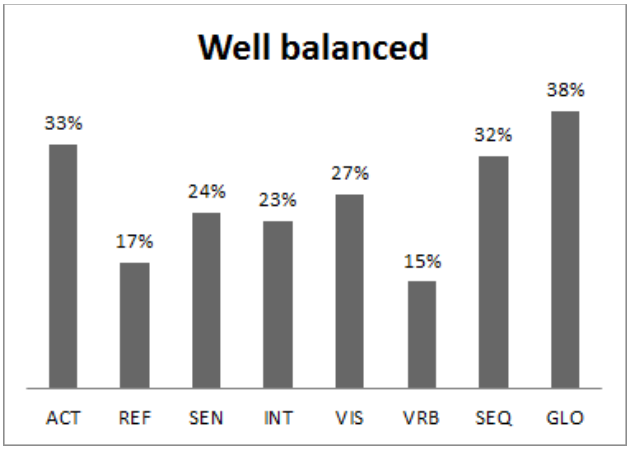


Fig. 1. Well balanced students

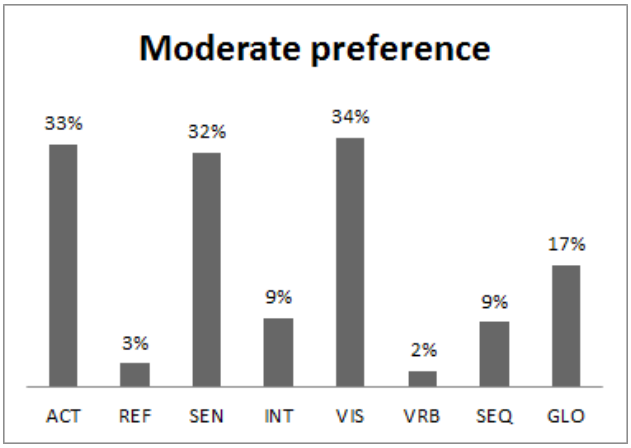


Fig. 2. Moderate preference

This way it was possible to find out the main preferences of students in the test group.

Based on the recommendations for the selection of educational material [3], [2], we designed and offered students assignments and theoretical materials according to their learning styles in the Moodle e-environment.

For example, to active learners we proposed group work assignments, to sensing learners – exercises, which were connected with solving real problems, and to visual learners – visual representation of course material, the same principles as have been used in the Khan Academy [5].

TABLE I.
THE PREFERENCES OF STUDENTS, OF THE TEST GROUP

	ACT	REF	SEN	INT	VIS	VRB
Well balanced	33%	17%	24%	23%	27%	15%
Moderate preference	33%	3%	32%	9%	34%	2%
Strong preference	13%	1%	10%	2%	22%	0%

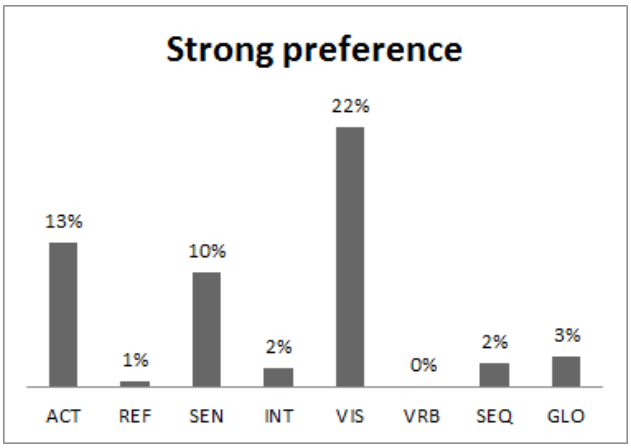


Fig. 3. Strong preference

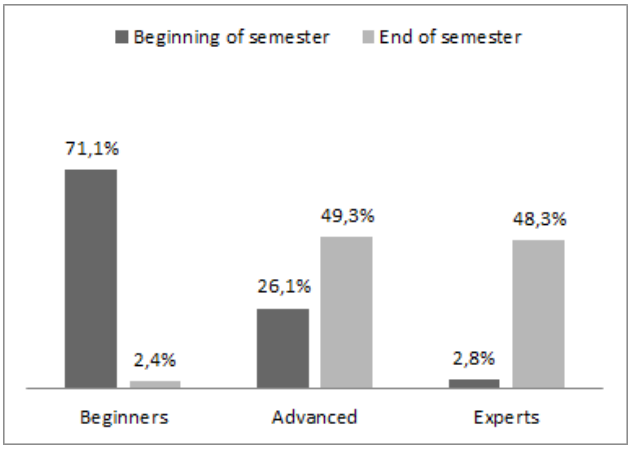


Fig. 4. Division of students into groups by test results in the reference group

All things considered, we managed to make the learning process more flexible by using e-environment opportunities: the students themselves chose the learning tempo, types of educational materials, and direction of individual and group work.

III. Results of the Experiment

The first results of our work showed a positive trend in the acquisition of knowledge. To divide students into groups by prior knowledge all of them were tested at the beginning of fall semester 2012. The same test was held at the end of fall semester. The test results confirm that students of the test group had mastered the learning material better than the students of the reference group (Fig 4 and 5).

Students of the test group coped better with their final exam due to the adopted learning material. Growth of knowledge has had a positive effect on their academic achievement (Fig 6).

Students' feedback in the test group also indicated that the material selected according to their learning styles motivated and helped them to learn.

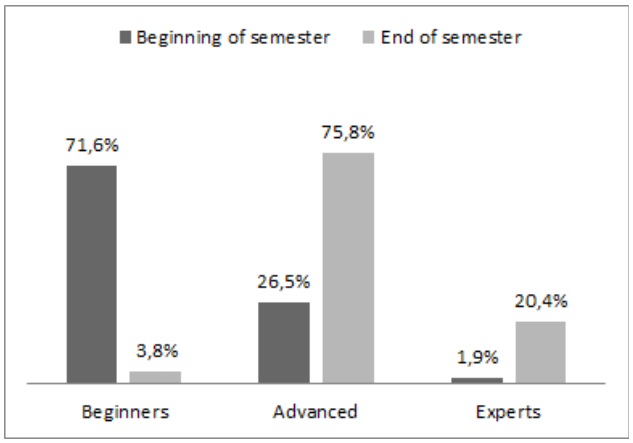


Fig. 5. Division of students into groups by test results in the test group

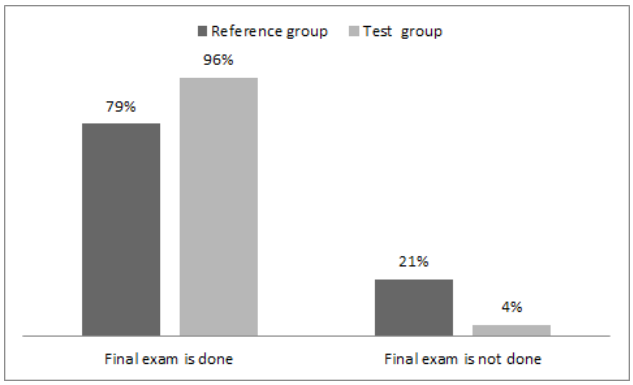


Fig. 6. Academic achievements. End of fall semester 2012

IV. Conclusions and Further Development

Students have different levels of motivation, different attitudes about teaching and learning, and different responses to specific classroom environments and e-learning. The more thoroughly teachers understand the differences, the better chance they have of meeting the diverse learning needs of all of their students. Teachers should attempt to improve the quality and efficiency of their teaching, which in turn requires understanding of the learning styles of students and designing instruction to meet these preferences.

Our selected flexible adaptive learning approach improved the quality of educational material and enhanced the educational effect of the use of innovative methods. The approach also provided us with additional opportunities to build individual educational paths for students, and in addition, apply the approach on students with different levels of readiness to learn.

Thus, we gave students the opportunity to choose their own way of learning the course. Students themselves felt the need for further studies, and did not feel the pressure from the teacher. They had the opportunity to work with educational materials in the manner and volume that was appropriate for them directly.

In conclusion, we would like to emphasize again that the content of the material adapted for each learning style should also cater for individualization of learning. It is important to remember that any learning style works well with the right approach.

Our chosen direction is a deeper study and analysis of students' data which could give us a better overview of why and how students learn. Additionally, there is the need for the curricula adaptation and teaching materials composition in accordance.

REFERENCES

[1] Felder, R.M. and Brent, R. Understanding Students Differences, Journal of Engineering Education, 2005, 94(1), pp. 57-72.

[2] Felder, R.M. and Silverman L.K. Learning and Teaching Styles in Engineering Education, Engr. Education, 1988, 78(7).

[3] Felder, R.M and Soloman, B.A (n. d.) Learning styles and strategies. Retrieved August 20, 2012, from http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/styles.htm

[4] Felder, R.M. and Spurlin, J. Applications, Reliability, and Validity of the Index of Learning Styles. Intl. Journal of Engineering Education, 2005, 21(1), pp. 103-112.

[5] Khan Academy. http://www.khanacademy.org

[6] Mironova, O., Amitan, I., and Vilipõld, J. Computational Thinking and Flexible Learning: Experience of Tallinn University of Technology. Lecture Notes in Information Technology; 2012, 23-24, pp. 183 – 188.

[7] Soloman, B. A. and Felder, R.M. (n. d.). Index of Learning Styles Questionnaire. http://www.engr.ncsu.edu/learningstyles/ilsweb.html

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 739–746

HEQAM: A Developed Higher Education Quality Assessment Model

Amin Y. Noaman[1], Abdul Hamid M. Ragab[1], Ayman G. Fayoumi[1], Ahmed M. Khedra[1] and Ayman. I. Madbouly[2]

[1] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
{anoaman, aragab, afayoumi, ahmedkhedra}@kau.edu.sa
[2] Deanship of Admission and Registration, King Abdulaziz University, Jeddah, Saudi Arabia
amadbouly@kau.edu.sa

***Abstract*–This paper presents a developed higher education quality assessment model (HEQAM) at King Abdulaziz University (KAU). This is because of; there is no universal unified quality standard model that can be used to assess the quality criteria of higher education. Besides, there are shortcomings in the coverage of some current educational quality standards. A Developed questionnaire to examine the quality criteria at KAU is investigated. The analytically hierarchy process is used to identify the priority and weights of the criteria and their alternatives. The model is constructed of three levels including eight main objectives and 53 alternatives. It included e-services criteria which is one of the recent university components, in addition to new sub-criteria for enhancing the model. It produces important recommendations to KAU higher authorities for achieving demanded quality services. Also, it helps KAU to achieve one of its strategic objectives to be a paperless virtual university.**

I. Introduction

UNIVERSITIES all over the world face big challenges to meet the growing number of students, supporting life-long learning for larger and larger parts of the population and of dealing with growing student heterogeneity. Beside these challenges universities are required to provide and maintain a high education quality learning environment based on a standard High Education Quality Criteria (HEQC).The high education service quality has gained tremendous attention from managers and academics due to its importance on business performance, cost reduction, and student satisfaction [1-4]. So, most of the universities are struggling to enhance the professional experience and skills of their personnel in order to utilize the new technologies in their teaching activities in an efficient way [5]. This is to gain a competitive advantage among other universities. Therefore, Saudi universities seek to examine their strategic positions by evaluating existing quality services, and adapting to students' perceptions to enhance their leadership position. On the other hand, higher education plays a significant role in advancing society toward sustainable development [6].

Having an acceptable level of quality services have to be the main concerns of any higher education university system, for guiding the country toward sustainable development. The kingdom of Saudi Arabia (KSA) government spends a lot of efforts to achieve a highly recognized education level by maintaining and improving the HEQC for all universities in the kingdom. Also, KAU has taken significant steps towards the improvement of education quality to facilitate the academic and managerial process, and to support policy making within the university. Due to the rapidly growing concerns about higher education quality in both international and local contexts, this paper proposes a developed model for evaluating higher education quality standards, and applying it at KAU as a case study. The next sections of this paper explain the related work, model construction, model evaluation, model results and discussion, then the conclusion and references.

II. Importance of the Study

Quality assessment of higher education institutions can contribute to the process of standardization of academic degrees. In fact, because of the changing landscape and increased call for accountability, higher education is now being challenged to re-conceptualize methods and processes used to indicate quality and excellence, including those used for assessing and evaluating quality of education programs. The quality of higher education services, especially in developing countries must be viewed as a strategic issue for social and technological development and economic growth [7]. Another issue that shows the importance of evaluating the quality of higher education programs and the need to have HEQC, is the fact that the world has become an open space were people circulates freely throughout all countries; this circumstance requires the establishment of quality standards so that a qualification obtained in the different institutions can be accepted all over the world, simply we can say that applying these HEQC will lead to and help in achieving the goal of accreditation of KAU education programs. Also, this is a major requirement to enhance the academic rank of Saudi universities among other worldwide universities.

III. Literature Review

Nowadays, service quality assessment is an issue that cannot be neglected by any university, even in the higher education in developing countries. In order to tackle this problem, it is necessary to invest in quality systems and tools for improvement. Universities are usually driven to engage in reforms by a variety of forces, which mostly come from globalization, supply and demand issues, competition, accountability, and technology. Their survival and development are determined by improving service quality those satisfying students' needs, since it is a vital significance to higher education services. Earlier researchers studied higher education quality services emphasized academic issues more than managerial issues [8,9], concentrated on effective course delivery

mechanisms and the quality of courses and teaching. Table 1 shows a brief of recent quality models that are used to evaluate higher education in some well-known universities. In this paper, a new service quality assessment model will be explained in section (4).

TABLE 1
HIGHER EDUCATION SERVICE QUALITY MODELS.

Authors	Year	University	Purpose of the used Model
M.S. Owlia & E.M. Aspinall [10]	1996	Birmingham, U.K.	Presents a new framework for dimension of quality in higher education.
R.F. Waugh [11]	2001	Australia	Proposes a model for university administration quality.
M. LALOVIC [29]	2002	Belgrade University	Presents an ABET assessment model using Six Sigma methodology to assessment in education.
S.Lagrosen, R. S. Hashemi, and M.Leitner [12]	2004	Austrian and Swedish students	Examine the dimensions that constitute quality in higher education and to compare these with the dimensions of quality that have been developed in general service quality research.
Z. Yang, L. Yan-ping and T. Jie [13]	2006	Chinese Higher Education	To design a model that is suitable to evaluate the service quality of Chinese higher education, using Servqual.
M. Tsinidou, V. Gerogiannis and P. Fitsilis [14]	2010	Higher education institutions in Greece	The quality determinants are identified for education services provider by higher education institutions in Greece, to measure their relative importance from the students' points of view.
A. R. Arokiasamy [15]	2012	Institutions in Malaysia	Configure the importance of maintaining service quality in higher education industry
This paper	2013	King Abdulaziz University, Jeddah, S.A.	A developer model for assessment of higher education quality Standards, case study KAU.

The assessment of the quality of human resources, physical, technological, financial and information resources at KAU could be appropriate, sufficient and accessible to realize its mission. Also KAU works effectively to plan, provide, evaluate, assure, and improve the academic quality and integrity of its academic programs, curricula, credits and degrees awarded. However, identifying all HEQC within KAU is an important issue. In this paper, a new effective quality model for evaluating HEQC at KAU is presented. To achieve the research objectives, a questionnaire that is used to examine the HEQC at KAU was developed. In addition to that, the e-services criteria are added to the quality model. There is no doubt that e-Services are one of the most recent required components for KAU. They will help in achieving the virtual university strategic goals, among which are the distance learning education. In addition, they are required to implement a successful paperless university [16-19]. They include: interactivity, mobility, flexibility, accessibility, portability, and social process. It is also aimed to achieve, highly demanded HEQC model that helps KAU to become a model to be followed as a paperless university, as well as achieving one of the KAU strategic objectives to be a virtual university.

IV. PROPOSED HIGHER EDUCATION QUALITY ASSESMENT MODEL

Fig.1 shows the proposed higher education quality model. It is based on the development of the model explained in [9]. This proposed model is constructed of three hierarchy levels, including eight main objectives (criteria), and 53 sub-objectives (i.e. alternatives). The eight main objectives include the following:

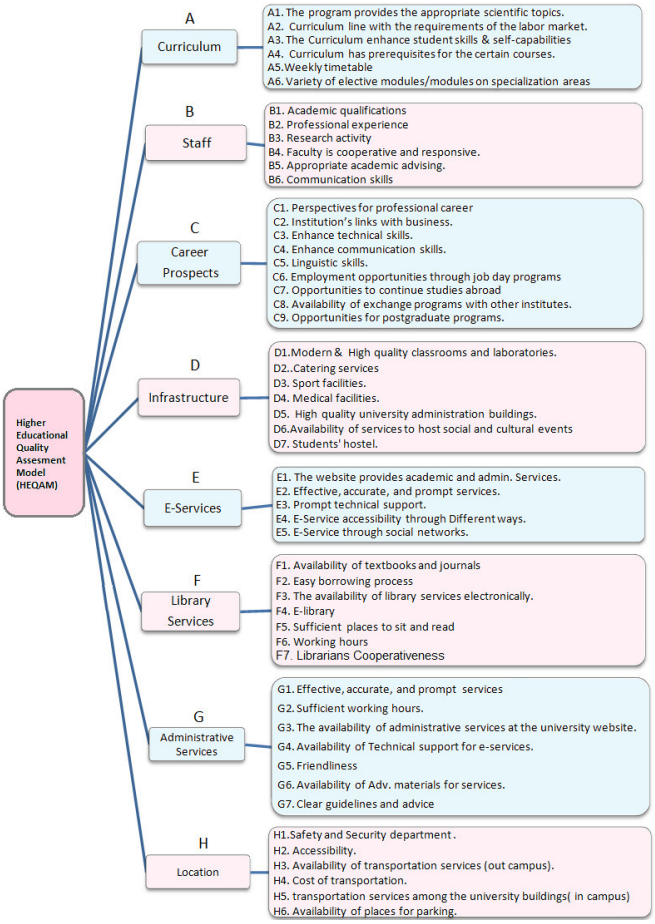


Fig.1 the Proposed HEQAM Model.

A. *Curriculum*

It is one of the main criteria that affect the higher education quality. It includes six sub-objectives (alternatives) named A1 to A6 and defined as shown in Fig.1. It plays an important rule for establishing the KAU university quality. The curriculum is an "organized program of study for a given degree, certificate award, incorporating all matters such as academic staff requirements, duration of academic program, admission requirements, content requirements and assessment process requirements" [20]. Also, all university curricula have to include [21]: enrolment requirements, objectives, scope, specific courses and content, duration, mode of assessment, standard references, and academic award. To achieve and maintain quality in curriculum development and delivery, university has to encourage academic excellence in research that enables departments to have professors, senior lecturers and several lecturers who participate in developing and reviewing curricula. The curriculum change imposed on higher education institutions through policies and strategies is required in order to develop graduations, enhance employability, widen access and improve retention. It sets out skills and employability curriculum framework for programs, including practical examples, and considers some of the challenges facing this holistic approach to a potentially fragmented area of policy development.

B. *Staff*

It includes six sub-objectives named B1 to B6 and defined as shown in Fig.1. The university that "holds essential educational facilities with affective staff of teaching and training will make students be more motivated, loyal and good performers"[22]. Good performance of teachers inside and outside the class is a significant feature for enhancing students' impartiality, motivation and satisfaction. Course instructors' teaching methodology is also a prime indicator considered by students, when they rate their teachers in their educational development and successful completion of their studies. Higher the intellectual ability of the instructor, the better will be the students' evaluation [23, 24] and consequently more will be the reliability on the teaching staff. The teachers who teach with punctuality, accuracy, reasonability and logical approach in a student friendly manner are more popular [25, 26]. Students level of satisfaction increases by working with those course instructors and lecturers who properly handle the assignments, projects, exams and facilitate students' logical reasoning and aptitude development.

C. *Career Prospects*

It includes nine sub-objectives named C1 to C9 and defined as shown in Fig.1. The quality of university education, allows students that get graduated an excellent career opportunities. Also taking into account the higher education to the needs of the labor market from diverse disciplines, provides job opportunities for graduates of eligible students.

D. *Infrastructure*

It includes seven sub-objectives named D1 to D7 and defined as shown in Fig.1. The infrastructure in higher education can include: facilities, researches, and faculties. In order to have a functional institution, all the aforementioned elements, have to be evaluated, improved and updated. University strategic planning has to include adequate infrastructure components into consideration, since good infrastructure enhances the quality of education and services provided. For examples, classrooms could be equipped with overhead projectors, internet connection, proper lighting and suitable cool system, to facilitate communications between instructors and students. In addition; for applied science; up-to-date laboratories and language labs are needed for experiments and projects of the fields, such facilities can increase learning quality and enhance the sense of research among students and faculty in the fields pursued.

It is necessary to have a healthy students and faculty body, by providing proper playgrounds, swimming pools and gym equipments. In addition, parking lots, which fulfill the needs of all university community, will ease the work conditions. Also, maintaining existing equipments and buying new ones are continuous tasks that require expertise and financial resources. There are needs for effective communication, cooperation, team-work among all the components inside university campus.

E. *E-Services*

It includes five sub-objectives named E1 to E5 and defined as shown in Fig.1. Using e-Services facilities, such as the integration of information and communication technologies, and internet in higher education, achieve imparting easily accessible, affordable and quality higher education leading to the uplift of Saudi Arabian universities. The benefits of e-services in education can provide, right from breaking time and distance barriers to facilitating collaboration and knowledge sharing among geographically distributed students. It increases the flexibility of delivery of education so that learners can access knowledge anytime and anywhere. It can influence the way students are taught and how they learn as now the processes are learner driven and not by teachers. This in turn would better prepare the learners for lifelong learning as well as to contribute to the industry. E-Services also play an important role for establishing the virtual university applying eLearning and also using necessary electronic resources capable for establishing the paperless university.

F. *Library Services*

It includes seven sub-objectives named F1 to F7 and defined as shown in Fig.1. The evolution of information technology has made students' needs for information services to change. This inevitably puts pressure on academic libraries, to work towards improving service quality and student satisfaction. This is necessary to face competition in global higher education industry whilst meeting the specific information needs of students. Students who constitute major users of academic libraries in universities often consider library's service quality based its ability to meet their expectations prior to enrolment. Thus, influencing their overall perceptions of the overall service quality of the institution necessitating a review of quality issues associated with services of academic libraries in universities. Adding electronic resources such as internet play important roles with research in the libraries. Journals and magazines library subscription also facilitate the students task of the faculty. Virtual libraries subscribing, also save time, money, and human resources.

G. *Administrative Services*

It includes seven sub-objectives named G1 to G7 and defined as shown in Fig.1. Administrative services managers plan, coordinate, and direct a broad range of services that allow universities to operate efficiently. A university may have several managers who oversee activities that meet the needs of multiple departments, such as mail, recordkeeping, security, building maintenance, and recycling.

The work of administrative services managers can make a difference in employees' productivity and satisfaction; for example; they might be responsible for making sure that the university has the supplies and services it needs. Administrative services managers also ensure that the university honors its contracts and follows government regulations and safety standards. Administrative services managers may examine energy consumption patterns, technology usage, and office equipment; for example; they may recommend plan for maintenance equipment or buying new ones.

H. Location

It includes six sub-objectives named H1 to H6 and defined as shown in Fig.1. University location security, safety and ease accessibility are important criteria from the student's point of view. They achieve a significant correlation between the quality of education and the distance of a college from the nearest town centre. Also, transportation services play an important role in the assessment of university location. They may include several alternatives among which are availability of transportation services in campus and out of campus, as well as cost of transportation.

V. Model Evaluation

Survey questionnaires are developed to collect information about current situation of higher education quality criteria at KAU. These questionnaires are adapted from a work explained in [9]. It is based on Servqual model aspects [8], although it does not use its defined dimensions. Two questionnaire are designed, one for students and the other for faculty members and expertise. The two questionnaires are developed, reviewed and updated with the assistance of KAU education expert consultants. Based on the results from these surveys, the main criteria for the main objectives and their related alternatives of the proposed model are identified. Then, the AHP method [27] is used as a tool for assessment of the weights of the model criteria and their priority. Table 2 shows the pairwise comparisons matrix among the main eight objectives of the higher educational quality model proposed, using the data collected from the developed questionnaires. Another additional eight pairwise comparison matrices are constructed to calculate the ranked weights for the sub-objectives, using AHP-Expert Choice [28].

VI. Results and Discussion

Based on the data collected from section (V) above, a group of eight main criteria with a total of 53 alternatives as shown in Table 2 are identified, in order to design the higher education quality model for enhancing service quality at KAU. Results in Table2 showed that the main eight criteria, including: Curriculum (A), Staff (B), Career Prospects (C), Infrastructure (D), e-Services (E), Library Services (F), Administrative Services (G), and Location (H) are ranked with 19.7%, 17.3%, 15.9%, 12.7%, 11.7%, 9.8%, 7.2% and 5.9% due to importance levels, respectively. The analyses of these criteria are explained in details next sections.

TABLE 2
PAIRWISE COMPARISON MATRIX.

	Curriculum	Staff	Career Prosp.	Infrastructure	E-Services	Lib.Services	Adm. Serv.	Location	Weights
Curriculum	1.00	1.00	1.00	2.00	2.00	2.00	3.00	3.00	19.7%
Staff	1.00	1.00	1.00	1.00	2.00	2.00	2.00	3.00	17.3%
Career Prospects	1.00	1.00	1.00	1.00	1.00	2.00	2.00	3.00	15.9%
Infrastructure	0.50	1.00	1.00	1.00	1.00	1.00	2.00	2.00	12.7%
E-Services	0.50	0.50	1.00	1.00	1.00	1.00	2.00	2.00	11.7%
Library Services	0.50	0.50	0.50	1.00	1.00	1.00	1.00	2.00	9.8%
Admin. Services	0.33	0.50	0.50	0.50	0.50	1.00	1.00	1.00	7.2%
Location	0.33	0.33	0.33	0.50	0.50	0.50	1.00	1.00	5.9%

A. Curriculum Quality

Six criteria are used to characterize the curriculum. Both faculty members and students were asked to give the importance rating these criteria. Results are shown in Fig.2; where appropriate scientific topics were the most important criteria. It can affect Curriculum with a 22.2% importance level. The second important criterion for this criterion is requirements of the labor market, with 20.1% importance level. Enhances student skills & self-capabilities is the third important criterion that can affect Curriculum in the model with 16.3% importance level. Prerequisites come in the fourth rank with 15.1% importance level. Weekly timetable and Elective modules have 13.4, and 12.8% importance level, respectively. Details of these ratings in relation to the model design are shown in column (A) in the Fig.2.

B. Staff Quality

Students and faculty member's questionnaire surveys reported that Staff within the university plays an important role in affecting the education quality. They have reported that Academic qualifications and Professional experience is the top of most importance level of this criterion. The applied AHP is used to assess quality determinants, to measure their weights to discover those that influence students' satisfaction most. Academic qualifications and Professional experience get 20.6% and 18.6% with respect to the staff criterion, respectively.

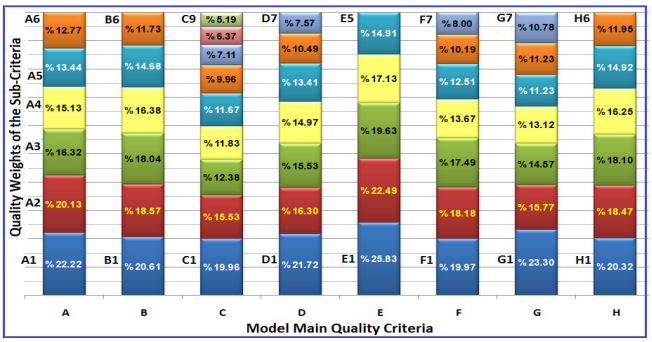


Fig. 2: Weight of alternatives to the HEQAM Model.

The third source of Staff quality criteria is the Research activity with an importance level of 18%. Cooperative, Academic advising, and Communication skills affect Staff quality criteria with 16.4%, 14.7%, and 11.7% importance level. Column (B) in the Fig.2 represents the Staff sub-criteria weights in percentage.

C. Career Prospects Quality

Another important factor that affects the higher education quality is Career Prospects. Students and faculty member's surveys reported the sub-criteria that affect this factor. Perspectives for professional career were the most frequently reported factor with an importance level of 20%. The other important factors that affect Career Prospects Quality are Institution's links with business, technical skills, communication skills, Linguistic skills, job day programs, studies abroad, exchange programs, and postgraduate programs.

Column (C) in the Fig.2 represents the Career Prospects sub-criteria weights in percentage.

D. *Infrastructure Quality*

This is one of the key criteria found and has been highlighted by both students and faculty members in the survey. Results showed that Modern & High quality classrooms and laboratories have 21.7% importance level with respect to Infrastructure Quality. The other related important level weights of the Infrastructure Quality sub-criteria is shown in column (D) of Fig.2.

E. *e-Services Quality*

In the survey faculty members and students rated the criteria of providing Academic and Admin-website Services as the most important criterion that may affect the E-Services with a 25.8% importance level. The other related important level weights for E-Services sub-criteria is shown in column (E) of Fig.2.

F. *Library Services Quality*

Results show that, the criteria of Availability of textbooks and journals were the most important criterion that may affect the Library Services Quality with 20% importance level. The other related important level weights of for these Library Services Quality sub-criteria is shown in the column (F) of Fig.2.

G. *Administrative Services Quality*

Results show that, the criteria of Effective, accurate, and prompt services were the most important criterion that can affect the Administrative Services Quality with a 23.3% importance level. The other related important level weights of for these sub-criteria is shown in the Fig.2, column (G).

H. *Location Quality*

Results show that the Safety and Security is the most important criterion that may affect the Location with a 20.3% importance level. The other related important level weights of these sub-criteria are shown in Fig.2, column (H).

VII. Model Quality Weights Comparison and Recommendations

All quality of model weights related to each criterion is shown in Fig.3. Comparison between different criteria's weights related to the criteria with the overall ranking for criteria's weights related to the HEQAM Model is shown in Table 3. The column of the total quality criteria (TQC) is computed by multiplying of weights related to criterion by the weight of the quality of the sub-criteria. For example, 19.7 X 22.2 = 4.4, 19.7 X 20.1 = 4, and 19.7 X 16.3 = 3.2, etc, hence the column of the TQC is computed as shown in Table 3. This table indicates the alternatives quality percentage alternatives quality weights arranged in ascending order. For example, the appropriate scientific topics (A1) have the first priority in the Curriculum, while the Academic qualifications (B1) have the first priority in the Staff quality. Table 4 shows the relation between the qualities of the sub-criteria alternatives and their weights related to the total quality of the model (TQM). The sub-criteria alternatives that occupied the first ten positions are:

(1) The appropriate scientific topics for a student's scientific path (A1).
(2) Curriculum line with the requirements of the labor market (A2).
(3) Academic qualifications (B1).
(4) Modern & High quality classrooms and laboratories (D1).
(5) Professional experience (B2).
(6) The Curriculum enhances student skills & self-capabilities (A3).
(7) Perspectives for professional career (C1).
(8) Curriculum has prerequisites for the certain courses (A4).
(9) The website provides academic and administrative services (E1).
(10) Research activity (B3).

These results have to be taken care of by the higher authorities at KAU. And be taken as recommendations to follow up in order to achieve high quality education.

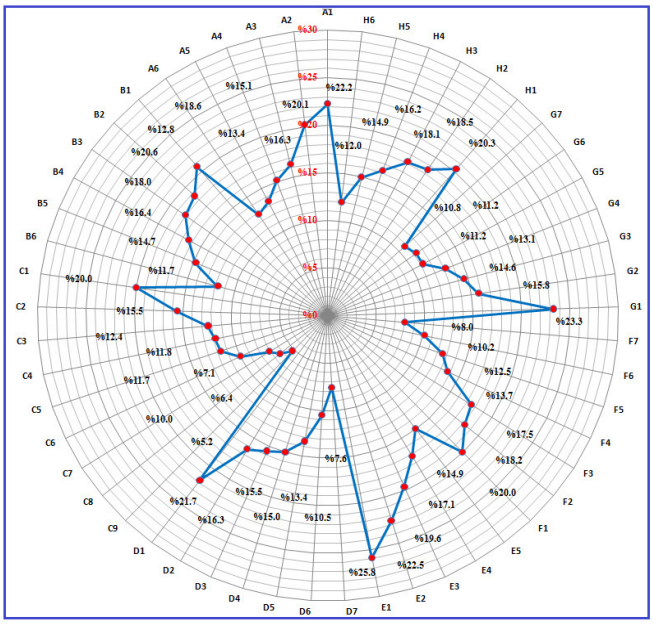


Fig.3: Ranking of all alternatives' weight to the HEQAM Model

TABLE 3
CRITERIA AND ALTERNATIVES OF HEQAM MODEL.

#	Main Quality Sub-Criteria	Alternatives	Weights related to Criterion	Weights related to TQC	Average Weights
A	Curriculum 19.7%	A1. The program provides the appropriate scientific topics	22.2%	4.4%	3.3%
		A2. Curriculum line with the requirements of the labor market.	20.1%	4.0%	
		A3. The Curriculum enhance student skills & self-capabilities	16.3%	3.2%	
		A4. Curriculum has prerequisites for the certain courses.	15.1%	3.0%	
		A5.Weekly timetable	13.4%	2.6%	
		A6. Variety of elective modules/modules on specialization areas	12.8%	2.5%	
B	Staff 17.3%	B1. Academic qualifications	20.6%	3.6%	2.9%
		B2. Professional experience	18.6%	3.2%	
		B3. Research activity	18.0%	3.1%	
		B4. Faculty is cooperative and responsive.	16.4%	2.8%	
		B5. Appropriate academic advising.	14.7%	2.5%	
		B6. Communication skills	11.7%	2.0%	
C	Career Prospects 15.9%	C1. Perspectives for professional career	20.0%	3.2%	1.8%
		C2. Institution's links with business.	15.5%	2.5%	
		C3. Enhance technical skills.	12.4%	2.0%	
		C4. Enhance communication skills.	11.8%	1.9%	
		C5. Linguistic skills.	11.7%	1.9%	
		C6. Employment opportunities through job day programs.	10.0%	1.6%	
		C7. Opportunities to continue studies abroad	7.1%	1.1%	
		C8. Availability of exchange programs with other institutes.	6.4%	1.0%	
		C9. Opportunities for postgraduate programs.	5.2%	0.8%	
D	Infrastructure 12.7%	D1.Modern & High quality classrooms and laboratories.	21.7%	2.8%	1.8%
		D2..Catering services	16.3%	2.1%	
		D3. Sport facilities.	15.5%	2.0%	
		D4. Medical facilities.	15.0%	1.9%	
		D5. High quality university administration buildings.	13.4%	1.7%	
		D6.Availability of services to host social and cultural events	10.5%	1.3%	
		D7. Students' hostel .	7.6%	1.0%	
E	E-Services 11.7%	E1. The website provides academic and admin. Services.	25.8%	3.0%	2.3%
		E2. Effective, accurate, and prompt services .	22.5%	2.6%	
		E3. Prompt technical support.	19.6%	2.3%	
		E4. E-Service accessibility through Different ways.	17.1%	2.0%	
		E5. E-Service through social networks.	14.9%	1.7%	
F	Library Services 9.8%	F1. Availability of textbooks and journals	20.0%	2.0%	1.4%
		F2. Easy borrowing process	18.2%	1.8%	
		F3. The availability of library services electronically.	17.5%	1.7%	
		F4. E-library	13.7%	1.3%	
		F5. Sufficient places to sit and read	12.5%	1.2%	
		F6. Working hours	10.2%	1.0%	
		F7. Librarian Cooperativeness	8.0%	0.8%	
G	Adminstrative Services 7.3%	G1. Effective, accurate, and prompt services	23.3%	1.7%	1.0%
		G2. Sufficient working hours.	15.8%	1.1%	
		G3. The availability of administrative services at the university website.	14.6%	1.0%	
		G4. Availability of Technical support for e-services.	13.1%	0.9%	
		G5. Friendliness	11.2%	0.8%	
		G6. Availability of Adv. materials for services.	11.2%	0.8%	
		G7. Clear guidelines and advice	10.8%	0.8%	
H	Location 5.9%	H1. Safety and Security department .	20.3%	1.2%	1.0%
		H2. Accessibility	18.5%	1.1%	
		H3. Availability of transportation services (out campus).	18.1%	1.1%	
		H4. Cost of transportation	16.2%	1.0%	
		H5. transportation services among the university buildings(in campus)	14.9%	0.9%	
		H6. Availability of places for parking.	12.0%	0.7%	

VIII. Conclusion

This paper proposed a higher education quality assessment model (HEQAM). It consists of eight sub-criteria, including 53 alternatives. The main criteria include Curriculum, Staff, Career Prospects, Infrastructure, E-Services, Library Services, Administrative Services, and Location Quality. The model is applied in KAU, for evaluating the education quality. The issue of main quality criteria and sub-criteria has been addressed to define determinates and their respective weight in the overall quality. The assessment of the university education quality from both students and expert's perspective are achieved using developed questionnaires. The work also provided recommendations on quality improvement of the institution based on its findings. The multi-criteria decision making AHP method was applied for qualitative and quantitative the model criteria. Results proved that the quality criteria that occupied the first five position included: the appropriate scientific topics for a student's scientific path (A1), Curriculum line with the requirements of the labor market (A2), Academic qualifications (B1), Modern & High quality classrooms and laboratories (D1), and Staff Professional experience (B2). The model quality weights obtained for the overall criteria have to be considered highly recommended factors to be followed for improving university education quality in the Kingdom of Saudi Arabia universities.

Acknowledgement

Authors would like to thank King Abdulaziz University in Jeddah, Saudi Arabia for offering the necessary facilities for accomplishing this research.

References

[1] T.Z. Chang, S.J. Chen," Market orientation, service quality and business profitability: a conceptual model and empirical evidence", Journal of Services Marketing, Vol. 12, No. 4, pp. 246-64, 1998.

[2] J. J. Cronin, S. A. Taylor, "Measuring service quality: A re-examination and extension", Journal of Marketing, Vol. 56, No. 3, pp. 55-68, 1992.

[3] C. Guru, "Tailoring e- service quality through CRM", Managing Service Quality, Vol.13, No. 6, pp. 20-531, 2003.

[4] G.S. Sureshchander, C. Rajendran, R.N. Anatharaman,"The relationship between service quality and customer satisfaction: a factor specific approach" , Journal of Services Marketing, Vol. 16, No. 4, pp. 363-79, 2002.

[5] N. Cavus, S. Kanbul "Designation of Web 2.0 tools expected by the students on technology-based learning environment", Procedia-Social and Behavioral Sciences, 2 (2), pp. 5824–5829, 2010.

[6] A. Craft," International Developments in Assuring Quality in Higher Education: Selected Papers from an International Conference", Montreal. Falmer Press, 1993.

[7] O. J. OLIVEIRA "Adaptation and application of the SERVQUAL scale in higher education", POMS 20th Annual Conference, Orlando, Florida U.S.A, May 1 to 4, 2009.

[8] O. J. OLIVEIRA "Adaptation and application of the SERVQUAL scale in higher education", POMS 20th Annual Conference, Orlando, Florida U.S.A, May 1 to 4, 2009.

[9] M. E. Malik, "The Impact of Service Quality on Students' Satisfaction in Higher Education Institutes of Punjab", Journal of Management Research, Vol. 2, No. 2: E10, 2010.

[10] M.S., Owlia, E.M. Aspinall, "A framework for the dimensions of quality in higher education", Quality Assurance in Education, Vol. 4 No. 2, pp. 12-20, 1996.

[11] R.F. Waugh, "Academic staff perception of administrative quality at universities". Journal of Educational Administration, Vol. 2 No. 2, pp. 172-88, 2001.

TABLE 4
RANKING OF ALL QUALITY OF THE SUB-CRITERIA ALTERNATIVES W.R.T. MODEL TOTAL QUALITIES (TQM).

#	Sub-Criteria Alternatives	Weights relate TQM
1	A1. The program provides the appropriate scientific topics	4.4%
2	A2. Curriculum line with the requirements of the labor market.	4.0%
3	B1. Academic qualifications	3.6%
4	D1.Modern & High quality classrooms and laboratories.	2.8%
5	B2. Professional experience	3.2%
6	A3. The Curriculum enhance student skills & self-capabilities	3.2%
7	C1. Perspectives for professional career	3.2%
8	A4. Curriculum has prerequisites for the certain courses.	3.0%
9	E1. The website provides academic and admin. Services.	3.0%
10	B3. Research activity	3.1%
11	B4. Faculty is cooperative and responsive.	2.8%
12	A5.Weekly timetable	2.6%
13	B5. Appropriate academic advising.	2.5%
14	E2. Effective, accurate, and prompt services .	2.6%
15	A6. Variety of elective modules/modules on specialization areas	2.5%
16	C2. Institution's links with business.	2.5%
17	E3. Prompt technical support.	2.3%
18	D2..Catering services	2.1%
19	E4. E-Service accessibility through Different ways.	2.0%
20	C3. Enhance technical skills.	2.0%
21	F1. Availability of textbooks and journals	2.0%
22	C5. Linguistic skills.	1.9%
23	D3. Sport facilities.	2.0%
24	D4. Medical facilities.	1.9%
25	F2. Easy borrowing process	1.8%
26	E5. E-Service through social networks.	1.7%
27	G1. Effective, accurate, and prompt services	1.7%
28	F3. The availability of library services electronically.	1.7%
29	D5. High quality university administration buildings.	1.7%
30	C4. Enhance communication skills.	1.9%
31	C6. Employment opportunities through job day programs.	1.6%
32	F4. E-library	1.3%
33	F5. Sufficient places to sit and read	1.2%
34	H1. Safety and Security Department	1.2%
35	B6. Communication skills	2.0%
36	H2. Accessibility	1.1%
37	C7. Opportunities to continue studies abroad	1.1%
38	H3.Availability of transportation services (out campus).	1.1%
39	C8. Availability of exchange programs with other institutes.	1.0%
40	G2. Sufficient working hours.	1.1%
41	F6. Working hours	1.0%
42	H4. Cost of transportation	1.0%
43	D6.Availability of services to host social and cultural events	1.3%
44	D7. Students' hostel .	1.0%
45	G4. Availability of Technical support for e-services.	0.9%
46	C9. Opportunities for postgraduate programs.	0.8%
47	H5. transportation services among the university buildings(in campus)	0.9%
48	G5. Friendliness	0.8%
49	G7. Clear guidelines and advice	0.8%
50	G6. Availability of Adv. materials for services.	0.8%
51	G3. The availability of administrative services at the university website.	1.0%
52	F7. Librarian Cooperativeness	0.8%
53	H6. Availability of places for parking.	0.7%

[12] S. Lagrosen, R. S. Hashemi, & M. Leitner, "Examination of the dimensions of quality in higher education", Quality Assurance in Education, Vol. 12 Iss: 2, pp.61 – 69, 2004.

[13] Z. Yang, L. Yan-ping and T. Jie," Study on Quality Indicators in Higher Education: An Application of The SERVQUAL Instrument" IEEE publications, Service Systems and Service Management International Conf., Vol. 2 Page(s): 1280 – 1286, 2006.
[14] M. Tsinidou, V. Gerogiannis, and P. Fitsilis, "Evaluation of the factors that determine quality in higher education: an empirical study", Quality Assurance in Education, Vol. 18 No. 3, 2010.
[15] A. R. Arokiasamy, "Literature Review: Service Quality in Higher Education Institutions in Malaysia", International Journal of Contemporary Business Studies Vol.3, No.4, pp. 227 – 244. May, 2012.
[16] S. Nyeck, M. Morales, R. Ladhari, & F. Pons, "10 years of service quality measurement: reviewing the use of the SERVQUAL instrument." from EBSCO host database. PP 101-107, July 8, 2007.
[17] Paperless University: http://oxforddictionaries.com/definition/american_english/paperless.
[18] S. M. Gilani, J.Ahmed, M. A. Abbas, "Electronic Document Management: A Paperless University Model", 2009.
[19] Virtual University: http://www.ascilite.org.au/conferences/brisbane99/papers/anderson.pdf.
[20] G. J. Cheserek," Quality Management in Curriculum Development and Delivery in African Universities: A Case Study of Moi University, Kenya, www.international-deans-course.org/uploads/media/Quality_Management_in_Curriculum_Development_and_Delivery_in_African_Universities_Cheserek_.pdf
[21] P. Wolf, A. Hill, F. Evers, Handbook for Curriculum Assessment, winter 2006, www.uoguelph.ca /tss/resources/pdfs/HbonCurriculumAssmt.pdf.
[22] C. C. Wei," Students Satisfaction towards the University: Does Service Quality Matters?" International Journal of Education, Vol. 3, No. 2: E15, 2011.
[23] A. R. Rodie, & S. S. Klein, Customer participation in services production and Delivery, Int. T. A. Swartz & D. Iacobucci (Eds.), Handbook of service marketing and management, pp. 111 – 126, Thousand Oaks, CA: Sage Publications, Inc.2000.
[24] H. K. Wachtel, "Student evaluation of college teaching effectiveness: A brief review". Assessment and Evaluation in Higher Education, 23(2), 191-212, 1998.
[25] J. G. Palli, R. and Mamilla," Students' Opinions of Service Quality in the Field of Higher Education", Creative Education, Vol.3, No.4, PP 430-438, 2012.
[26] J. Lu, " Measuring cost/benefits of e-business applications and customer satisfaction", Proceedings of the 2nd International Web Conference, 29–30 November, Perth, Australia, 139-47, 2001.
[27] T .L. Saaty, L.G. Vargas, Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, Kluwer's Academic Publishers, Boston, USA, 2001.
[28] A. Ishizaka and A. Labib," Analytic Hierarchy Process and Expert Choice: Benefits and Limitations", ORInsight, 22(4), p. 201–220, 2009.
[29] M. LALOVIC, An ABET Assessment Model using Six Sigma Methodology, A dissertation submitted to the Division of Research and Advanced Studies of the University of Cincinnati, Department of Mechanical, Industrial and Nuclear Engineering of the College of Engineering, Belgrade University,2002.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 747–750

Computer Modelling of Cognitive Processes

Nina Rizun
Alfred Nobel University,
Dnipropetrovs'k
ul. Naberezhna Lenina 18, 49000
Dnipropetrovs'k, Ukraine
Email: n_fedo@mail.ru

***Abstract*—Brief description of the author's results of development of cognitive processes (CP) computer modelling concept on the basis of improving the methodology and expanding the area of using computer-based testing technology in education is suggested. The fundamental heuristics for formalizing: the concept of degree of difficulty of test tasks (TT); the degree of confidence of individual's CP concept; the concept of stability modes of individual's CP and CP phases during the working time; the concept of the target level of the test session are presented.**

The heuristic algorithms of intellectual express and expanded analysis of the TT quality are developed. The algorithm of obtaining the cognitive individual's profile for formation and adequate interpretation of individual intellectual characteristics is offered. The concept of technical implementation of the CP computer model into the informational learning environment is formulated.

I. Introduction

BEGINNING of the XXI century represents a crucial stage in the development of education. In intellectually intense and fast-paced high-tech environment, with an excess of information, which has already exceeded the capacity of individual's perception of it, new approaches in teaching and educational technologies are necessary.

Due to the emergence of new trends in science up-to-date conceptual framework was born. Thus, the term "cognitive" (from the Latin word "cognition" – knowledge, perception), meaning "informative", "pertaining to knowledge," appeared in the sixties of the last century as a result of existence of a new paradigm in psychological research (cognitive psychology, cognitive science). In this paradigm special attention is paid to traditional cognitive processes (CP): perception, attention, memory, imagination and thinking etc. However, the cognitive approach is fundamentally different in a way that all of these processes are considered as components of the overall process of information exchange between the individuals during the learning.

Under the new conditions new learning technologies must be created – cognitive, i.e. ways, techniques, methods to ensure effective understanding by the individuals of information on the bases of unique indicators and characteristics of their CP. The task of great importance is to develop the modern methodologies of organizing individuals' learning and continuous monitoring and assessing of individuals' CP indicators, as well as of forming individual productive paths in the cognitive processes by means of introducing effective feedback systems via using the computer testing controlling methodology.

Recent scholarly research in the field of creating computer learning technologies can be divided into several categories:

Intelligent Tutoring Systems: information-reference systems; consultative type systems; intellectual training (expert-type tutoring) systems; accompanying type systems (e.g., ELM-ART-II, AST, ADI, ART-Web, ACE, KBS-Hyperbook, ILESA, DCG, SIETTE) [1].

E-Learning Management Systems elaborated the following SCORM standards, the specifications of the IMS Global Learning Consortium, and the Aviation Industry Computer-Based Training Committee (AICC) that regulate certain aspects of their development and use: the system's architecture and the system's interaction with outside systems; the ways of the learning system's interaction with learning resources; the presentation of courses' contents; the models of learning control; the testing algorithms and ways of presenting testing results (e.g., BaumanTraining, eLearning 3000, WebTutor) [2,3].

Diagnostic and Planning Expert Teaching Systems based on using the methodological tools of computer testing [4-7].

Despite the wide range of scholarly achievements and market offers in the field of computer learning technologies, all of them have a number of similar specifications:

a) The use of testing technology mostly for measuring students' learning individual achievements and the lack of methods for obtaining the specific characters of their individual intellectual activity and features of CP.

b) The necessity of development of the methods of intellectual diagnostics of test materials for increasing the quality of CP assessment.

c) The inflexible demands in most Ukrainian higher schools (HS) to the software required for their functioning and to the technical characteristics (in particular, the capacity characteristics) of computers, as well as to the speed and time of Internet resources use.

978-83-60810-53-8/$25.00

The purpose of this paper is brief description of the author's results of development of CP computer modelling concept on the basis of innovative approaches to improving the methodology and expanding the area of using the computer-based testing systems (CBTS) technology. The paper is also aimed at the presenting the key aspects of developing the use of author's modelling propositions as an instrument for receiving and adequate interpretation of the set of quantitative and qualitative identifiers of individual intellectual characteristics of individuals.

II. Cognitive Process Computer Modeling Concept

The CP computer modelling concept suggested by the author is based on implementing the following functional components: the database, the expert system of integrated diagnostics and cognitive process control (the latter comprising the logical conclusion mechanism, the working memory, the knowledge base, as well as the explanation subsystem and the dialogue subsystem), control systems, and CBTS.

A. The Database Structure

The database of the computer model is designed for storing:

– the structured learning content (LC), presented in the following systemically coordinated forms: brief textual notes of lectures for individuals' self-studying before the actual in-class teaching/learning starts Teach[1] (contain the basic notions, definitions, laws, examples, and algorithms of situational knowledge use from the course); slide-notes of the lecture materials for demonstrating and discussing directly in the in-class teaching/learning process – Teach[2]; laboratory assignments – Teach[3]; testing materials for assessing the degree of individuals' CP – Teach[4].

– reference and factual information about the syllabus, the number of individuals, the distribution of academic hours and learning units between in-class and out-of-class (independent) individuals' work;

B. Knowledge Base Structure

The core of the expert system of integrated diagnostics and cognitive process control is the knowledge base designed for storing: expert knowledge and the acquired analytically knowledge.

The foundations of expert knowledge are the author' research results in the area of computer testing methodology improvement and the development of intellectual instruments for supporting decision-making in what concerns the CP analysis. There are the following heuristics:

a) Reference time T_{ni} is the objective tool of complex quantitative formalization (scaling) of the degree of difficulty: the statement and visual representation of TT; the TT itself causes the timetable for task processing; the technology of entering the results of CP.

b) The degree of mismatching in factual and reference time for solving the TT – dynamic coefficient $D^f_{i=}t^f_i/t^n$, – demonstrates the objectivity of determined indicator of degree of difficulty of the TT – R_i.

c) The complex indicator of the degree of confidence of CP and the probability of guessing the correct solution is the correlation coefficient K_i between series of factual T_{fi} and reference time T_{ni} spent on correct result of cognitive processing of TT. The value of T_{ni} is determined after check testing of a group of experts [8].

d) The interpretation of the normalized K_i ranges may be the following: an individual with high level of the confidence solves the TT at steady pace ($K(t^*, t^f) \geq 0{,}5$); in the behavior of an individual, who has middle level of CP confidence, there are "gaps" in the problem domain's assimilation and uncoordinated pace of solving the tasks ($0{,}3 \leq K(t^*, t^f) < 0{,}5$); the individual, whose CP level is low, may try to guess the correct decision ($K(t^*, t^f) < 0{,}3$).

e) The concept of modes of stability of CP is a consequence of entering the concept of interpretation of the CBTS and individuals as a dynamic system. In this regard, the interpretation of equilibrium (EM) and periodic (PM) modes of the individual's CP may be the following [9]: EM corresponds to the situation, which is characterized by in-time constancy CP during the testing session; the PM is characterized by fluctuations in the CP individual's entropy as a result of changes in the level of complexity of TT.

f) In the author's interpretation the EM of CP is quantitatively identified by $K(t^*, t^f) \geq 0{,}5$ and PM – by $K(t^*, t^f) < 0{,}5$.

g) The concept of the CP phases is defined as a set of functional states of an individual during the test session: the primary reaction – short-term of reduction of the actual level of confidence and precision of CP; overcompensation and compensation – gradual improvement and stabilization of indicators of confidence and precision of CP; subcompensation and decompensation – reduction of actual level of confidence and accuracy of CP.

h) The concept of the informativeness level is considered on the basis of quantification of the effectiveness of test performance.

i) The concept of the target level $TL_i = f(U_j, P_j, Z_i)$ is considered on the basis of scaling the TT by: TT forms U_j (from lowest to highest: with only one correct answer; closed form of TT with multiple choice; matching TT; open form of TT with sequence-setting); boundary probability (P_j) of guessing the correct solution of TT; cognitive process difficulty levels Z_i (from lowest to highest: recognition and presentation; reproductive replay; productive replay).

j) The efficiency of CP testing is increased due to determining the consecutive order of giving TT in accordance with the decrease of their target level [10].

The Expert Block of the Knowledge Base includes the knowledge about subject area and the control knowledge:

a) The Algorithms of TT Quality Analysis:

– Heuristic Algorithm of the TT Quality Q_i Express Analysis. Presupposes the methodology of stage-by-stage guaranteed acquisition of a complete testing results' matrix from two incomplete matrices – the results of preliminary and final testing in the framework of one class period [10];

– Heuristic Algorithm of Expanded Analysis and of Improving the TT Quality Q_i. Includes possibility of establishing the fact (1) and assuming the ways of eliminating the "problematic" character of TT. This algorithm is based on entering the indicator of dynamic coefficient D^f_i in accordance with the following rule [11]:

$$\begin{aligned} &\text{if } Low_F \le Low_R \text{ and } High_F \ge High_R, \text{then } PROBLEM_v = -1 \\ &\text{if } High_F \le Low_R \text{ and } Low_F \ge High_R, \text{then } PROBLEM_v = +1 \\ &\text{else } PROBLEM_v = 0 \end{aligned} \quad ,(1)$$

where $D_{Ri} = \{Low_R, High_R\}$ is the boundary percentage of individuals, whose factual time of TT completion does not meet the reference one ($D^f_i \neq 1$).

b) The Algorithms of Adaptive Sequence of Distributing TT in order of decreasing the target level TL_i. This algorithm determines the necessity of transition λ to the next TL (2) in the conditions of surpassing the value of the boundary percentage G_{gi} of correct answers to TT as compared to the factual G_{fi} percentage. In the opposite case, this algorithm determines the necessity of terminating the testing procedure [12]:

$$\chi = \begin{cases} 0, \text{if } G^f_i < G^n_i \Rightarrow R_i = R_i - 1 \\ 1, \text{if } G^f_i \ge G^n_i \Rightarrow \text{Stop} \end{cases} \quad (2)$$

c) The Algorithm of Cognitive Individual's Profile (CP_PROFILE) obtaining:

– in the author's interpretation CP _PROFILE is a set of specific specifications describing: modes (MODES) and phase (PHAZES) stability of CP ; informativeness level (IL) of problem-oriented knowledge of individuals:

$$\text{CP\_PROFILE= <MODES> \& <PHAZES> \& <IL>} \quad (3)$$

– via using a statistical series of reference and actual time of CP of TT: processing transfer function charts for groups of individuals with a stable EM and PM of cognitive processes and individual "pictures" – structure of the time distribution (TD) during the test session and the intensity (high, middle or low frequency of correct answers) within the phase – of phases can be obtained (Table I). It helps to identify differences between individual's behaviour concerning the specificity of CP.

TABLE I.
SPECIFICATIONS OF THE INDIVIDUAL INTELLECTUAL PROFILE OF CERTIFIED INDIVIDUALS

MODES	Equilibrium mode of CP					Periodic modes of CP									
PHASES	High confidence level of CP					Middle confidence level of CP					Low confidence level of CP				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
TD (%)	5	87		8		20	60		20		10	70			20
Intensity	h	h	h	h	h	h	n	m	h	h	m	l	l	l	m
Explorations	Steady CP. Stable behavior within the fixed working time					Periodic state of CP. Insufficiently stable behavior within the fixed working time					Unstable state of CP. Insufficient level of knowledge				

– in accordance with algorithm of adaptive identification of the IL the formation of a quantitative assessment test perfomance (effectiveness) is carried out taking into account minimization of the impact of guessing on the objectivity of test results interpretation [8, 13] (Table II).

TABLE II.
THE ALGORITHM OF ADAPTIVE IDENTIFICATION OF IL

Levels of CP Confidence	Expert Conclusion	Algorithm of Identifying the IL
Low	Short time spent for solving difficult TT and long time spent for solving simple TT	$IL_k = \sum_{i=1}^{n} \left(S_i * \frac{t^f_i}{t^n_i} \right)$
Moderate	Only exceedingly prompt responses entail a "penalty"	$IL_k = \sum_{i=1}^{n} S_i * \frac{\sum_{i=1}^{n} t^f_i}{\sum_{i=1}^{n} t^n_i}$
Medium or High	Rational distribution of time for solving between difficult and simple TT	$IL_k = \sum_{i=1}^{n} S_i$

The Block of Knowledge Acquired by the Expert System in an analytical way comprises:

1. Knowledge about the current state of the CP: the individual's indications of CP, obtained on the bases of CP_PROFILE; the indications of learning content quality Q_k; the information about individual learning paths Din_k.
2. Actualized knowledge about the subject area – the dynamic boundaries.
3. Control knowledge which is a collection of algorithms for processing the database statistics and of control knowledge of the expert part of knowledge base that is launched by the mechanism of logical conclusion for acquiring new knowledge.

C. The Concept of Technical Implementation of the CBTS

The CBTS presupposes the implementation of a number of technological and methodological solutions. The first of them is the most effective integration into the informational learning environment of the majority of Ukrainian HS by means of using MS Office as an instrumental base for creating a unified up-to-date system of controlling the teaching/learning process.

The second one is placing the repository of the LC on a powerful (possibly distant) server in the Internet network with the aim of eliminating the limitations of technical characteristics in those computers that most HS in the country are equipped with.

The third is the use of such regimes as short-time deliveries of LC from the server; e-mailing by using wireless connections or the internal network; mobile (distant) regime

of system's work. These regimes are used for ensuring economy and efficiency of network resources [14].

D. Cognitive Model Control Algorithm

The developed methodology of CP computer modelling presupposes completion of the following principal control iteration stages [15]:

1. The Teacher, in advance of the next class, uses the control system for:

a) Formulating a request to the LC repository on the distant server as to compiling a set of learning units $Teach_z$.

b) Fine-tuning of the CBTS by indicating: the maximum number of points to be scored for every testing session (TS); the number of TL to be used in the TS; the modes of user's actions limitations: the possibilities for an individual to choose the order of TT and of returning the TT within test session.

c) Tuning the short-term connection with the Internet via a radio modem for transmitting the required LC $Teach_z$ to individuals.

2. Before coming to their class, individuals need: to get acquainted with the LC of the first form $Teach^1_z$; to take preliminary testing with using $Teach^4_z$; by means of tuning the short-term connection with the Internet – to transmit results of their testing to the control system.

3. On the basis of the Adaptive Methodology of Learning Process Individualization the following steps are implemented: multi-aspect diagnostics of the CP level with taking into account the quantitative and qualitative learners' intellectual characteristics; intellectual support of decision-making as to adaptive regulation of the structure and content of class work in accordance with the expert recommendations formulated in the knowledge base. Those recommendations allow:

a) Determining the most efficient conditions forms (group, individual autonomous, creative, calculative, research, or team learners' work) for individual cognitive processes.

b) Determining the learning elements requiring specific forms of cognitive processes (generalization, recapitulation, practical examples).

c) Automated tuning of $Teach^2_z$ or $Teach^3_z$ with recommendations as to mandatory consideration, possible consideration, or lack of necessity for consideration of qualitative results of students' acquisition of certain LC.

4. At the end of the class students have the final testing. As a result, information is collected and transmitted into the intellectual knowledge base.

III. Conclusion

Scientific novelty and practical value of the suggested CP computer modelling concepts, which allow to improve the CBTS-methodology and develop indicators of individual's intellectual activity and features of CP, are attested to and confirmed by the patents of Ukraine [8, 10-12, 14, 15]. The model has been introduced into the teaching practice of the Department of Economic Cybernetics and Mathematical Methods in Economics at Alfred Nobel University, Dnipropetrovs'k, Ukraine.

References

[1] Brusilovsky, P. Adaptive and intelligent technologies for Web-based education. In C. Rollinger and C. Peylo (Eds.). Konstliche Intelligenz, Special Issue on Intelligent Systems and Teleteaching, No. 4, pp. 19-25, 2004.

[2] Advantages and Disadvantage of Distance Learning. Education as a Road to Success. Ufa, 2010 (the original is in Russian).

[3] Chen, L., Wang, N., & Chen, H. (in press). E-learning in Chinese schools and universities. In J. Baggaley & T. Belawati (Eds.). Distance Education Technology in Asia (pp. 90-101). New Delhi: Sage. Retrieved from www.pandora-asia.org/downloads/Book-2/PANdora-book2_v6-Chap4.pdf as of December 12, 2009.

[4] Rybina, G.V. The development and use of integrated teaching expert systems in the teaching/learning process. In The Collection of Scholarly Papers from the Russian Scholarly-Methodological Conf*erence "Improvement of IT Specialists' in Applied Informatics* Training on the Basis of Innovative Technologies and E-*Learning"* (pp. 219-226). Moscow: MESI, 2007 (the original is in Russian).

[5] Ermolenko, O.B., Kovalyiov, V.I., Lysnoy, A.I., & Serkov, O.A. The Method of Designing the Adaptive Teaching/Learning System: Patent No. 3619U. Ukraine: 7G09B 7/07. Patent holder: Kharkiv Polytechnic Institute, No. 2004010029, 2004 (the original is in Ukrainian).

[6] Kudrjavcev, V.S., Waschik, K., Strogalov, A.S., Aliseytshik, P.A., & Peretruchin, V.V. The educational computer system of automatic machine type. In Problems of Theoretic Cybernetics (p. 111). Moscow: Publishing Center RSHU, 1996.

[7] Khmelyov, A.G. Neuronetwork technologies in the systems of automated checking of students' knowledge. In The Proceedings of the 16th All-Ukrainian Scholarly-Methodological *Conference "The Issues of Economic Cybernetics"* September 14-16, 2011 (pp. 120-125). Odessa, 2011 (the original is in Russian).

[8] Taranenko, I.K., & Rizun, N.O. The Meth*od of Measuring Learners'* Level of Knowledge in Computer Testing: Patent No. 51559. Ukraine: MPK G06F 7/00. Patent holders: Taranenko, I.K., & Rizun, N.O., No. u200913726, 2009 (the original is in Ukrainian).

[9] Rizun, N.O., & Taranenko, I.K. Receiving of the mathematical model of testee's intellectual activity with the use of statistical methods. System technologies. Regional Interuniversity Collection of Scientific Papers. – Issue 3(86). – Dnipropetrovs'k, 2013 (pp. 97-108). (the original is in Russian).

[10] Rizun, N.O., & Taranenko, I.K. The System of Teaching and Measuring the Quality of Test Materials: Patent No. 65657. Ukraine: MPK G06F 7/00. Patent holders: Rizun, N.O., & Taranenko, I.K., No. u201106558, 2011 (the original is in Ukrainian).

[11] Rizun, N.O. Heuristic algorithm for improving the technology of test task quality assessment. The Eastern-European Journal of Progressive Technologies. No. 3/11 (45), 40-49, 2010 (the original is in Russian).

[12] Kholod, B.I.,& Taranenko, I.K., & Rizun, N.O. The Method of Com*puter Testing of Students' Knowledge*: Patent No. 97149. Ukraine: G06F 7/04 (2006.01). Patent holders: Alfred Nobel University, Dnipropetrovs'k, No. a 2009 12950, 2012 (the original is in Ukrainian).

[13] Rizun N. Development of methods and models of inaccuracy minimization of the machine-to-human interaction in automated systems of professional readiness level diagnostics. Scholarly and Technical Journal "Nauchniy Visnyk NMU, No 2, 2013 (pp.17-25).

[14] Taranenko, I.K., & Rizun, N.O. The Mobile System of Teaching with Using Computer Testing: Patent No. 64481. Ukraine: MPK G06F 7/00. Patent holders: Taranenko, I.K., & Rizun, N.O., No. u201104361, 2011 (the original is in Ukrainian).

[15] Rizun, N.O., Taranenko, I.K., Tarnopolskyy, O.B., & Kholod, B.I. The System of Teaching/Learning with the Use of Computer Testing: Patent No. 64873. Ukraine: MPK G06F 7/00. Patent holders: Rizun, N.O., Taranenko, I.K., Tarnopolskyy, O.B., & Kholod, B.I., No. u201104040, 2011 (the original is in Ukrainian).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 751–758

Hands-On Exercises to Support Computer Architecture Students Using EDUCache Simulator

Sasko Ristov, Blagoj Atanasovski, Marjan Gusev, and Nenad Anchev
Ss. Cyril and Methodius University,
Faculty of Information Sciences and Computer Engineering,
Rugjer Boshkovikj 16, PO Box 393,
1000 Skopje, Macedonia
Email:sashko.ristov@finki.ukim.mk, blagoj.atanasovski@gmail.com, marjan.gushev@finki.ukim.mk,
nenad_ancev@hotmail.com

***Abstract*—EDUCache simulator [1] is developed as a learning tool for undergraduate students enrolled the Computer Architecture and Organization course. It gives the explanations and details of the processor and exploitation of its cache memory. This paper shows a set of laboratory exercises and several case studies with examples on how to use the EDUCache simulator in the learning process. These hands-on laboratory exercises can be also used in learning software performance engineering and to increase the student willingness to learn more hardware based courses in their further studying.**

***Index Terms*—Education; HPC; CPU Cache; Multiprocessor.**

I. Introduction

THE Computer Architecture and Organization course is devoted to help the students to understand how the computers work. This course is usually in the first study year and the teaching material is almost always totally new for the students. Computer architecture is acknowledged as a significant part of the body of knowledge and an important area in undergraduate computer science curricula [2], [3]. Learning the course requires huge efforts by the students, especially in case of computer science students. Instead of wanting to know how the hardware (computer) works, they just want to use it as a necessary tool to execute their software programs. While developing programs by using some high-level programming language, the students do not get a clear picture of how they are executed by the computer. This decreases the students' interest and deeper understanding in learning of the Computer Architecture and Organization course. Therefore, it makes the teaching even more difficult requiring a lot of effort from both instructors and students [4]. Teachers must not only cover a body of knowledge, but they must motivate students and make the course exciting by selecting appropriate topics, such as which processor should be learned [5].

Today's modern multi-processors consist of multilayer cache memory system [6] to speedup data access balancing the gap between CPU and main memory. This complicates the learning process even more since the students must learn the organization inside the multi-processor, and not only the architecture. We have developed EDUCache simulator [1] that visually presents cache hits and misses, cache line fulfillment, cache associativity problem [7], for both sequential and parallel algorithm execution. In this paper we present several hands-on exercises for EDUCache simulator that will improve the teaching and alleviate the students' learning process. Several predefined examples for special memory patterns that cover data locality and cache set associativity are also presented.

The rest of the paper is organized as follows. In Section II we discuss the related work about improving the teaching and learning of computer architecture and other hardware courses for computer science students. Section III briefly describes the Computer Architecture and Organization course. The EDUCache architecture, user interface and different working modes are described in Section IV. The newly proposed hands-on laboratory exercises and some predefined examples are presented in Section V. The final Section VI is devoted on conclusion and future work.

II. Related Work

This section presents the existing similar visual simulators that cover the area of computer architecture and organization. We also present the proposed methodologies and laboratory exercises in order to lighten the learning and teaching of the course.

A. Hands-on Exercises: Simulation or Real Hardware

Introducing appropriate hands-on exercises, homework assignments and projects besides the lectures will make the course more interesting and will provoke the students to dive more deeply to learn how the computer works. Liang [8] performed a nice survey of hands-on assignments and projects.

Two approaches exist in organizing the laboratory exercises, i.e., either with visual simulators or working on real hardware. Using appropriate visual simulators on hands-on exercises lightens the teaching process and can significantly improve the students' interest in hardware generally [9], [10]. The simulators or web online tools share the laboratory equipment, thus removing the obstacles of cost, time-inefficient use of facilities, inadequate technical support and limited access to design and laboratory resources [11].

Practical work on real hardware is also very important [12], [13], [14]. Wang [15] and Lee [16] proposed FPGA-based configurable processors to be used in laboratory exercises.

978-83-60810-53-8/$25.00

The students can develop, implement and monitoring both hardware and software of multi-core processor systems on real hardware.

B. Teaching Methodologies

Several methodologies are developed to teach the hardware courses in general for computer science students. Reinbrecht et al. [17] present a methodology that integrates functionally verified ASIC (Application-Specific Integrated Circuit) soft cores into a FPGA in order to allow the students to learn the fundamentals of hardware and its designs challenges, not only development, but also verification and physical implications.

Ackovska and Ristov [18] improved the hands-on laboratory exercises and introduced a new teaching methodology. They divided the exercises in tutorials and tasks. The former are published a week before the exercises as students can prepare for the exercise, while the latter are given to the students during the laboratory exercises. These changes improved students' grades without making the course exams easier. Hatfield and Jin [19] designed and developed many laboratory exercises to design and implementation of an operating model of a pipelined processor.

Da [20] elaborated the common methods of classroom teaching and experimental teaching and some efficient methods for classroom and laboratory teaching. He [21] addressed five problems related to computer architecture education in multi-core era and suggested a possible solution.

C. Visual Simulators

We have found many visual simulators that cover a particular fundamental part of computer architecture and organization. Nikolic et al. [9] evaluated many simulators and concluded that some simulators are designed for teaching, some for data profiling, and none of them covers all topics in computer architecture and organization. We [1] have presented a very comprehensive overview of several visual simulators:

- EduMIPS64 [22] for instruction pipelining, hazard detection and resolution, exception handling, interrupts, and memory hierarchies;
- Dinero IV [23] for memory hierarchy with various caches on single core systems;
- CMP$im [24] - based on the Pin binary instrumentation tool;
- HC-Sim [25] generates traces during runtime and simulates multiple cache configurations in one run;
- Herruzo's [26] simulator for understanding the cache look up process and writing elements in the cache memory.
- Misev's [27] visual simulator for ILP dynamic out-of-order executions;
- Valgrind [28] (with its module Cachegrind) profiler for cache behavior;

All these simulators were not primarily developed for teaching the cache memory although most of them are visual. They lack educational features since they are built to complete the simulation as fast as possible rather than to present the architecture and organization of cache memory system in a modern multi-processor. Our EDUCache simulator [1] offers step by step simulation allowing the students to pause the simulation and analyze the cache hits and misses in each cache level. Its power increases with appropriate hands-on exercises.

SimpleScalar [29] and SMPCache [30] are additional simulators selected by William Stallings as a simulation tool for the implementation of student projects [31].

SimpleScalar is used for program performance analysis, detailed microarchitectural modeling, and hardware-software co-verification. It supports non-blocking caches, speculative execution, and branch prediction.

SMPCache is a widely used simulator in more than 100 universities and research centers. It is a trace-driven simulator for the analysis and teaching of cache memory systems on symmetric multiprocessors, analyzing program locality; influence of the number of processors, cache coherence protocols, schemes for bus arbitration, mapping, replacement policies, cache size (blocks in cache), number of cache sets (for set associative caches), number of words by block (memory block size).

Although SMPCache is primary developed as a learning tool for the Computer Architecture and Organization course, it should be redeveloped since it has an option for one cache level and symmetric processor only, while our EDUCache simulator offers simulation of heterogeneous multiprocessor with three cache levels. Our hands-on laboratory exercises proposed in this paper supplements its value.

III. The Computer Architecture and Organization Course

This section briefly describes the Computer Architecture and Organization Course.

The course's main objective is to offer the students a clear understanding of the main computer architectures, performance of the computer parts and the whole computer system. It also covers the topics of today's modern multi-chip and multicore multiprocessors, as well as the digital logic circuits.

A. Course Organization

The teaching of the course is organized in three parts: theoretical lectures with 2 classes per week, theoretical exercises with 2 classes per week and practical exercises with 1 class per week in laboratory. Lectures and theoretical exercises are organized in larger groups of around 100 students, while practical exercises are carried out in computer laboratories in groups of up to 20 students, with each student working on its own workstation. Prerequisites for enrolling in the course are previously completed course in Discrete Mathematics.

Theoretical lectures cover the computer abstractions and technology, the computer language (MIPS), computer arithmetic, the processor, memory, storage, and multichip multicore multiprocessors [32].

Theoretical exercises are divided in two parts. The first midterm covers the topics: computer arithmetic, codes and performance parameters, while the second part deals with

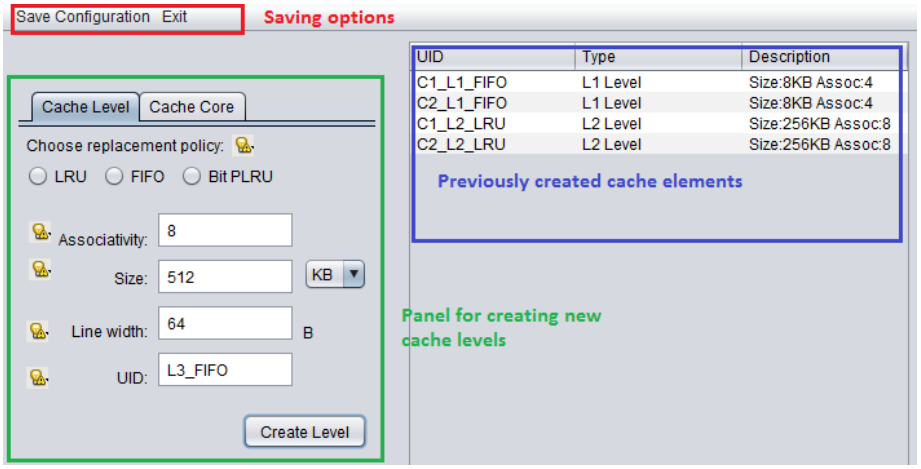


Fig. 1: Overview of design mode of EDUCache simulator - Creating L3 cache instance

digital logic. Hands-on laboratory exercises follow the topics of theoretical exercises.

The course can be passed in two ways, i.e., through midterms or final exam. The students must take the theoretical lecture part and exercise part (plus logic circuits) in either way.

B. Course Obstacles

The previous section briefly describes the course organization. We have analyzed the student results and determined that they had more problems with the topics of theoretical lectures compared to the exercises, and more precisely, the material of the second midterm, i.e., the processor, memory, I/O and parallelization. Our analysis show that although these subjects are covered during the theoretical lectures, neither theoretical nor practical exercises are provided for these topics, since the exercises cover to the design of logic circuits. Even more, IEEE Computer Society and ACM stated that more attention should be given to the multi-core processors architecture and organization, instead of the logic design level [33].

Therefore, we developed the EDUCache simulator that covers these topics. In this paper, we present the hands-on laboratory exercises that will make it even more appropriate in the teaching process, mainly focused on multiprocessor, cache and main memory.

IV. EDUCache Simulator

This section briefly describes the main features, interfaces and user interface of the EDUCache simulator. More details about the EDUCache simulator are presented in [1].

The EDUCache simulator is a platform independent simulator developed in JAVA whoose main simulation is described by a set of Java classes, each for a different CPU cache parameter. It allows the students to design a multi-layer cache system with different multi-core multi-cache hierarchy and to analyze sequential and parallel execution of user algorithm. Each chip can have one or several homogeneous cores. Each core has access to some cache of different cache level (generally L1 to L3). Particular cache can be owned by one, several or all cores of the chip. In general, L1 and L2 caches are private per core in modern multi-processors, while L3 cache is shared among several or all cores.

The particular cache level parameters can vary. Cache is determined by cache memory size, cache line size, cache associativity and replacement policy. EDUCache simulator allows the students to configure all these cache parameters and cache levels.

A. User Interface

The EDUCache simulator user interface is visual and user friendly. It uses the Multiple Document Inteface (MDI) paradigm. The EDUCache simulator works in two modes: *Design* and *Simulation*.

1) Design Mode: The students can configure various cache parameters and levels to create instances of cache levels and share them among chip cores.

The students create the cache levels with unique ID (UID). Figure 1 depicts an overview of EDUCache simulator user interface in design mode and an example on how a student can very easily create a particular cache level instance, select a replacement policy, set associativity, cache line size, cache size and select unique cache level UID.

After creating cache level instances, the students can create a core, selecting cache instances from the list of previously created ones (visible in the table in the right frame) for each cache level. Figure 2 depicts a user interface to create a core with unique Core UID, using previously created cache level instances.

Finally, the students can save the created configuration that represents a CPU chip. They configure the core instances, which they prefer to include on the chip and they are prompted where to save the configuration file.

After completing a multi-core chip with different caches in the design mode, the students can move to the simulation mode in order to simulate some memory accesses and analyze which of them will generate hit or miss in particular cache level of particular core.

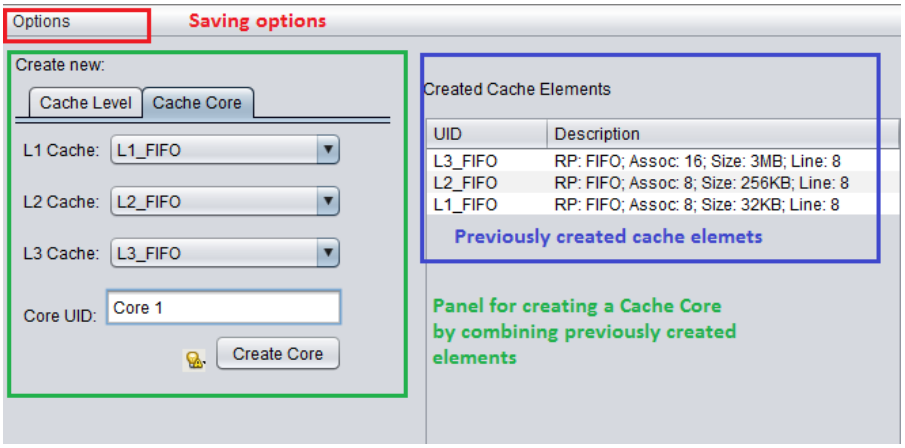


Fig. 2: Overview of design mode of EDUCache simulator - Creating a CPU core

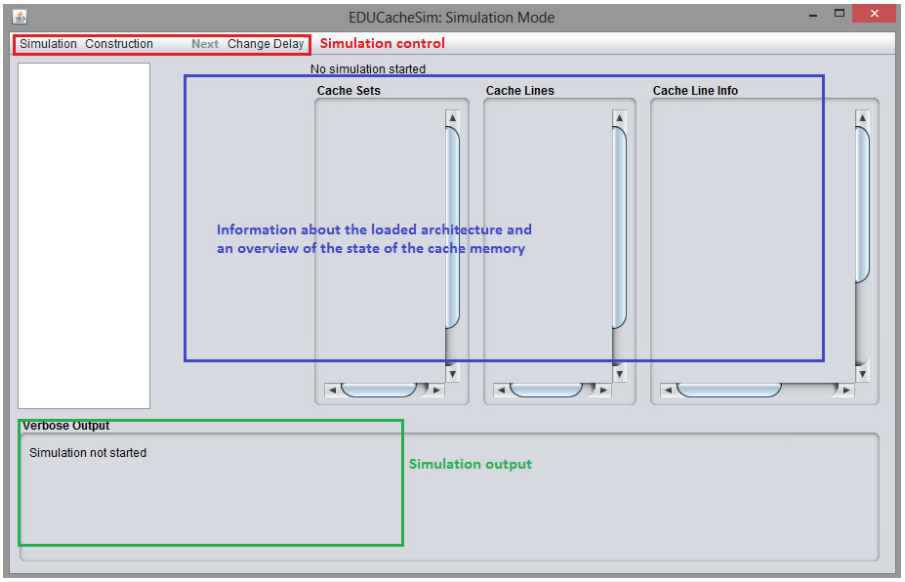


Fig. 3: Overview of Simulation mode - hit in L1 level of core C2, set #8, line #1, address 123416

2) Simulation Mode: After a configuration of the CPU chip with multiple cores per chip and multiple cache levels per each core, the students should load the memory addresses and run the simulation.

Figure 3 depicts the Simulation mode. Its main window consists of:

- *Simulation Control Menu Bar* - is the central control hub for the simulation process. It contains 2 menus, i.e., *Simulation* and *Construction*. The latter creates new or loads the existing configuration file for a core. The former loads a study case file, trace file, and operates the simulations (start, pause, stop, or step by step working mode);
- *Loaded Address Trace Frame* - shows the contents of the trace file, i.e., which core should read the address and the physical address that is loaded.
- *Verbose Output Frame* - shows the addresses that are read by cores, the search in L1 cache and selecting the set in which the address is supposed to map, the result, i.e. cache hit or miss, the cache line number if it is hit and the evicted line if the chosen set was full and read miss is generated; and
- *Visual Representation Frame* - is the main feature of simulation mode which gives a visual representation to the lookup process. It represents different levels of the cache level architecture: Core Pane, Cache Sets Pane, Cache Lines Pane, and Cache Line Info Pane.

V. Hands-on Exercises for EDUCache Simulator

This section presents how the EduCache simulator can be used as a tool in the laboratory exercises to introduce the students with the basic concepts of processor and its cache memory.

A. General Terms

All hands-on laboratory exercises follow the same concept. Each hands-on exercise starts with an explanation of the goal and objectives. Next, a brief coverage of the required topics is presented. This will remind the student of the topics that need to be learned or revised in order to be able to prepare for the exercise and complete it. The exercises should be given to the students a week before the exercise [18].

The objectives are step by step guides on what the student is supposed to do, e.g., configure the simulator, create a certain architecture, execute a simulation or analyze the results from an executed simulation. The guidelines are posed as simple instructions or as learning objectives. Questions are placed between the guidelines alerting the students which areas require more attention. In the end, the students must answer all questions, create the configuration file of the simulator and the simulation result file.

B. The Hands-on Exercises

This section presents the hands-on laboratory exercises for the EDUCache simulator. We present several exercises, some of which can be gathered or divided according to the available time for the hands-on laboratory exercises.

1) Exercise 1: Intro to EDUCache Environment: The first laboratory exercise is designed to introduce the EDUCache simulator to the students. The exercise goes over the different types of files that the simulator uses. The basic commands require the students to go through a simulation and to analyze the results. Learning objectives include: EDUCache design and simulating modes, loading a cache configuration file, and basic cache memory elements. The exercise concludes with running a simulation on a loaded trace file, creating a new trace file and running the simulation again, finishing with an analysis of the statistics that the simulator presents after the simulation is finished.

Although this exercise does not require a lot of students' effort, it should be graded. Otherwise, the students may not pay enough attention on learning the elementary controls of the simulator, which will cause them to have trouble with later exercises.

2) Exercise 2: Different Cache Parameters: The second laboratory exercise aims to present the basic parameters of cache memory to the students. That is, the size, associativity and the principles of multiple cache levels. The learning objective of this exercise is for the students to understand how these parameters impact on specific program execution. The principles of time and data locality are covered. The simulator is used to create multiple configurations with different parameters regarding to the cache size. Simulation is realized on a single memory trace. The students must observe into the results of the simulation and compare to find how the different sizes effect the program execution.

The final part of the exercise is to determine the smallest size for a cache level that gives the same performance as an infinite cache size. The grading should include optional questions for extra credit to inspire the students to show interest in the exercises since this exercise contains a significant number of tasks (and objectives) that the students must complete.

3) Exercise 3: Overview of Cache Set Associativity and Replacement Policies: The third laboratory exercise goes over the concepts of cache set associativity and replacement policies as one of the more complex cache memory parameters. The exercise shows the impact of these parameters on a specific program execution. The students must create configurations and use them to execute and analyze multiple simulations.

A set of address traces is given and the students' task is to observe and conclude the optimal replacement policy for each address trace. The exercise also offers the possibility to create experimental configurations which do not usually appear in real systems, such as certain cache levels with certain replacement policy. Another set of objectives takes a look at the influence of the set associativity.

TABLE I: Example 1 that generates cache misses

Parameter	L1	L2	L3
Size	32B	64B	128B
Associativity	2	2	4
Replacement	FIFO	FIFO	FIFO
Cache line	8B	8B	8B

TABLE II: Results of the simulation of the Example 1

Parameter	L1	L2	L3
Total reads	50	50	50
Cache hits	0	0	0
Cache misses	50	50	50

C. The Demo Examples

In this section we present several demo examples for characteristic memory access patterns in order to lighten the learning and understanding of the processor and its cache memory architecture and organization.

1) Example 1: Cache Miss due to "Loosely" Data: This example demonstrates the continuous cache misses for the loosely data. Table I presents the example of cache parameters. The trace file forces the access of the elements with 8B offset since we want to force a cache miss for each memory read (each read accesses the element of different cache line). Total 50 reads are realized and the results of the simulation are presented in Table II.

2) Example 2: Each Second Access is Cache Hit due to Data Locality: This example accesses pairs of elements such that each pair is placed in the unique cache line. The cache parameters are presented in Table III. Total 10 reads are realized. The results of the simulation are presented in Table IV. That is, we forced 5 pairs of miss and hit in L1 cache.

3) Example 3: Always Cache Hit due to Data Locality: This example demonstrates how the set associative cache memory generates cache hits for "tightly" data (data locality) of a single cache line. The cache parameters are the same

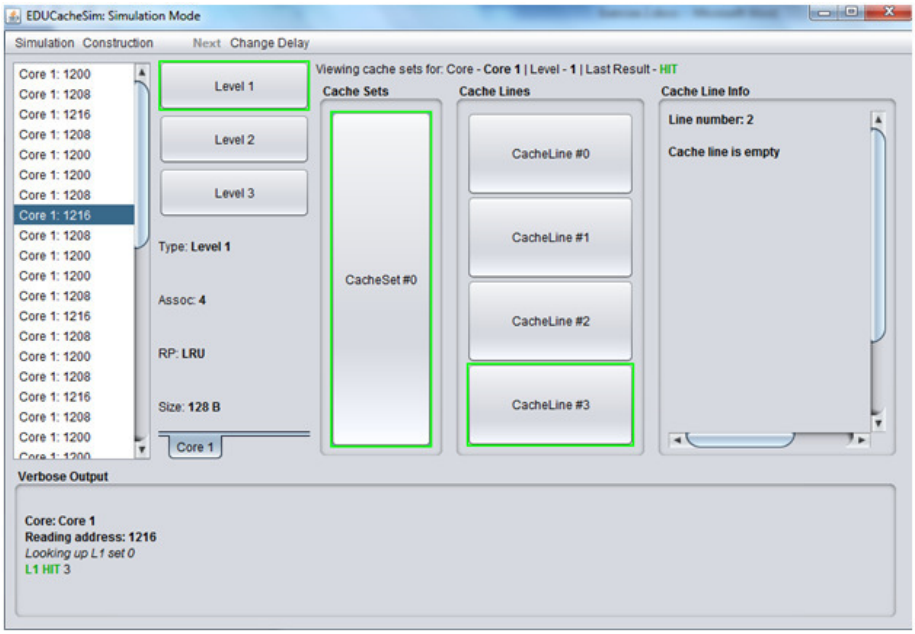


Fig. 4: Simulation of Example 3

TABLE III: Example 2 that generates cache hits for the second access

Parameter	L1	L2	L3
Size	128B	192B	256B
Associativity	4	4	8
Replacement	LRU	LRU	LRU
Cache line	32B	32B	32B

TABLE IV: Results of the simulation of the Example 2

Parameter	L1	L2	L3
Total reads	10	5	5
Cache hits	5	0	0
Cache misses	5	5	5

as the Example 2 presented in Table III. Since we want to generate a cache hit for each memory access, all addresses in the memory trace are in the range of a single cache line. Total 42 reads are realized. The results of the simulation are presented in Table V. That is, 1 cache miss is generated by the first access, and 41 cache hits by all others.

Figure 4 depicts a hit occurring on L1 cache always in the same cache line, as the rightmost pane shows the other cache

TABLE V: Results of the simulation of the Example 3

Parameter	L1	L2	L3
Total reads	42	1	1
Cache hits	41	0	0
Cache misses	1	1	1

TABLE VI: Configuration for Example 4

Parameter	L1	L2	L3
Size	16B	32B	64B
Associativity	2	2	4
Replacement	FIFO	FIFO	FIFO
Cache line	8B	8B	8B

lines are empty because all the required elements have been loaded into a single cache line Line #3.

4) Example 4: Cache Associativity Problem: This example demonstrates how the set associative cache memory can generate continuous cache misses if the data access are always in the same cache set, i.e., cache associativity problem [7]. The cache parameters are presented in Table VI. Total of 15 reads are realized, but only 3 different addresses are accessed. The results of the simulation are presented in Table VII.

This example results in constant L1 cache misses because of constant eviction of cache lines in the same set. Because we want to generate a cache miss by looking up the same cache set (in our case $CacheSet\#0$) for each memory read, all addresses in the memory trace must satisfy $Block\_address = X \cdot Number\_of\_cache\_sets$.

For our configuration, the cache line is 4 bytes, which yields that if the address in main memory is N bytes long, the block address will be the first $N - 2$ bits [6].

Figure 5 depicts a step in the simulation of this exercise. Reading the element stored in address 0 generates a cache miss on the Level 1 cache, because the previous read replaced it from the cache set.

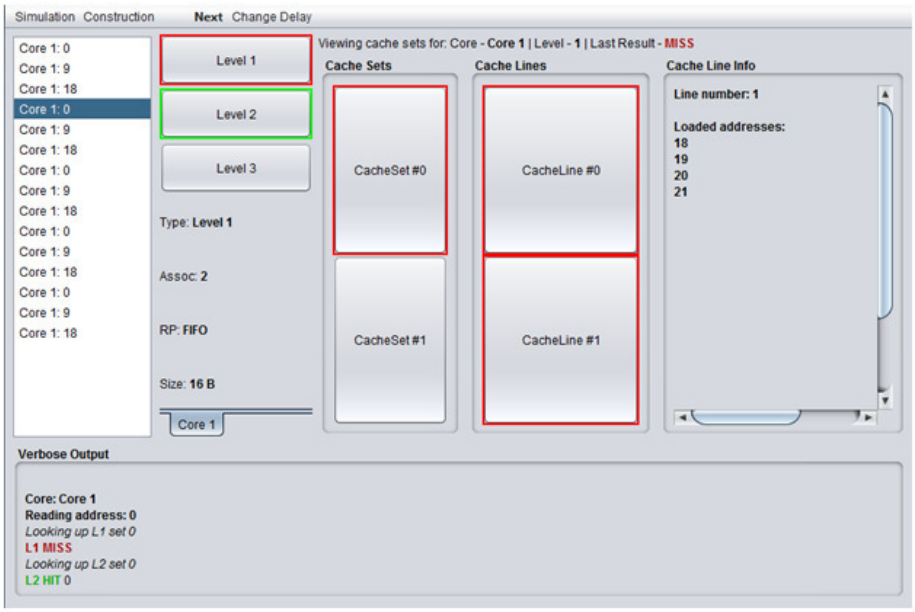


Fig. 5: Simulation of Example 4

TABLE VII: Results of the simulation of the Example 4

Parameter	L1	L2	L3
Total reads	15	15	3
Cache hits	0	12	0
Cache misses	15	3	3

Reading the element will generate cache hit in the Level 2. The rightmost panel shows the loaded addresses in $CacheLine\#0$ in $CacheSet\#0$.

VI. Conclusion and Future Work

EDUCache visual simulator offers the students a tool to design their own CPU core with multi level cache memories. It simulates cache misses and hits in particular cache set and memory location for sequential and parallel execution of an algorithm. The students can interactively learn about the cache hierarchy, architecture and organization of private cache level per core or shared cache level among all or a group of cores, the cache capacity and associativity problem, cache line, cache replacement policy, data locality etc.

This paper presents several hands-on laboratory exercises to support the students for the Computer Architecture and Organization course, i.e., using the EDUCache simulator they will better understand the architecture and organization of the modern processor and its cache memory. Several predefined examples are also presented to lighten the learning process and increase the students' willingness for the Computer Architecture and Organization course. This will help the students to develop their algorithms to achieve maximum performance using the same hardware resources.

We will introduce the EDUCache simulator and the hands-on exercises to this semester in courses Computer Architecture and Organization and Parallel and Distributed Processing, and survey the students about the impact to their willingness for learning the processor and its cache memory. Additional analysis will be realized after finishing the course this year to determine the results of the exams for the topics that cover the EDUCache simulator and the proposed hands-on exercises and examples.

References

[1] B. Atanasovski, S. Ristov, M. Gusev, and N. Anchev, "EDUCache simulator for teaching computer architecture and organization," in *Global Engineering Education Conference (EDUCON), 2013 IEEE*, March 2013, pp. 1015–1022.

[2] R. Shackelford, A. McGettrick, R. Sloan, H. Topi, G. Davies, R. Kamali, J. Cross, J. Impagliazzo, R. LeBlanc, and B. Lunt, "Computing curricula 2005: The overview report," *SIGCSE Bull.*, vol. 38, no. 1, pp. 456–457, Mar. 2006.

[3] M. Stojcev, I. Milentijevic, D. Kehagias, R. Drechsler, and M. Gusev, "Computer architecture core of knowledge for computer science studies," *Cyprus Computer Society J.*, vol. 5, no. 4, pp. 39–42, 2003.

[4] M. Stolikj, S. Ristov, and N. Ackovska, "Challenging students software skills to learn hardware based courses," in *Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd International Conference on*, june 2011, pp. 339 –344.

[5] A. Clements, "Arms for the poor: Selecting a processor for teaching computer architecture," in *Frontiers in Education Conference (FIE), 2010 IEEE*, 2010, pp. T3E–1–T3E–6.

[6] J. L. Hennessy and D. A. Patterson, "Computer Architecture, Fifth Edition: A Quantitative Approach," MA, USA, 2012.

[7] M. Gusev and S. Ristov, "Performance gains and drawbacks using set associative cache," *Journal of Next Generation Information Technology (JNIT)*, vol. 3, no. 3, pp. 87–98, 31 Aug 2012.

[8] X. Liang, "A survey of hands-on assignments and projects in undergraduate computer architecture courses," in *Advances in Computer and Information Sciences and Engineering*, T. Sobh, Ed. Springer Netherlands, 2008, pp. 566–570.

[9] B. Nikolic, Z. Radivojevic, J. Djordjevic, and V. Milutinovic, "A survey and evaluation of simulators suitable for teaching courses in computer architecture and organization," *Education, IEEE Transactions on*, vol. 52, no. 4, pp. 449 –458, nov. 2009.

[10] S. Ristov, M. Stolikj, and N. Ackovska, "Awakening curiosity - hardware education for computer science students," in *MIPRO, 2011 Proceedings of the 34th International Convention, IEEE Conference Publications*, 2011, pp. 1275 –1280.

[11] D. Pop, D. G. Zutin, M. E. Auer, K. Henke, and H.-D. Wuttke, "An online lab to support a master program in remote engineering," in *Proceedings of the 2011 Frontiers in Education Conference*, ser. FIE '11. USA: IEEE Computer Society, 2011, pp. GOLC2–1–1–GOLC2–6.

[12] I. Kastelan, D. Majstorovic, M. Nikolic, J. Eremic, and M. Katona, "Laboratory exercises for embedded engineering learning platform," in *MIPRO, 2012 Proc. of the 35th Int. Conv.*, 2012, pp. 1113–1117.

[13] J. Qian, R. Wang, S. Shi, Y. Zhu, and Z. Xie, "Simplifying and integrating experiments of hardware curriculums," in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, vol. 9, 2010, pp. 610–614.

[14] D. Kehagias and M. Grivas, "Software-oriented approaches for teaching computer architecture to computer science students," *Journal of Communication and Computer*, vol. 6, no. 12, pp. 1–9, Dec. 2009.

[15] X. Wang, "Multi-core system education through a hands-on project on fpgas," in *Frontiers in Education Conference (FIE), 2011*, 2011, pp. F2G–1–F2G–6.

[16] J. H. Lee, S. E. Lee, H.-C. Yu, and T. Suh, "Pipelined cpu design with fpga in teaching computer architecture," *Education, IEEE Transactions on*, vol. 55, no. 3, pp. 341–348, 2012.

[17] C. Reinbrecht, J. Da Silva, and E. Fabris, "Applying in education an FPGA-based methodology to prototype ASIC soft cores and test ICs," in *Programmable Logic (SPL), 2012 VIII Southern Conference on*, 2012, pp. 1–5.

[18] N. Ackovska and S. Ristov, "Hands-on improvements for efficient teaching computer science students about hardware," in *Global Engineering Education Conference (EDUCON), 2013 IEEE*, March 2013, pp. 295–302.

[19] B. Hatfield and L. Jin, "Improving learning effectiveness with hands-on design labs and course projects for the operating model of a pipelined processor," in *Frontiers in Education Conference (FIE), 2010 IEEE*, 2010, pp. F1E–1–F1E–6.

[20] L. Da, "Computer hardware curriculums, curriculum contents and teaching methods," in *Computer Science Education, 2009. ICCSE '09. 4th International Conference on*, 2009, pp. 1506–1511.

[21] L. He, "Computer architecture education in multicore era: Is the time to change," in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, vol. 9, 2010, pp. 724–728.

[22] D. Patti, A. Spadaccini, M. Palesi, F. Fazzino, and V. Catania, "Supporting undergraduate computer architecture students using a visual mips64 cpu simulator," *Education, IEEE Transactions on*, vol. 55, no. 3, pp. 406 –411, aug. 2012.

[23] J. Edler and M. D. Hill, "Dinero iv trace-driven uniprocessor cache simulator," 2012. [Online]. Available: http://pages.cs.wisc.edu/~markhill/DineroIV/

[24] A. Jaleel, R. S. Cohn, C.-K. Luk, and B. Jacob, "Cmpsim: A pin-based on-the-fly multi-core cache simulator," in *The Fourth Annual Workshop MoBS, co-located with ISCA '08*, 2008.

[25] Y.-T. Chen, J. Cong, and G. Reinman, "Hc-sim: a fast and exact l1 cache simulator with scratchpad memory co-simulation support," in *Proc. of the 7-th IEEE/ACM/IFIP Int. conf. on HW/SW codesign and system synthesis (CODES+ISSS '11)*. USA: ACM, 2011, pp. 295–304.

[26] E. Herruzo, J. Benavides, R. Quislant, E. Zapata, and O. Plata, "Simulating a reconfigurable cache system for teaching purposes," in *Microelectronic Systems Education (MSE '07). IEEE International Conference on*, 2007, pp. 37 –38.

[27] A. Misev and M. Gusev, "Visual simulator for ILP dynamic OOO processor," in *WCAE '04, Proceedings of the workshop on Computer architecture education: in conduction with the 31st International Symposium on Computer Architecture*, E. F. Gehringer, Ed. ACM, USA, 2004, pp. 87 –92.

[28] Valgrind, "System for debugging and profiling linux programs," [retrieved: May, 2013]. [Online]. Available: http://valgrind.org/

[29] SimpleScalar LLC, "Simplescalar tool set," [retrieved: May, 2013]. [Online]. Available: http://www.simplescalar.com/

[30] University of Extremadura, "Smpcache - simulator for cache memory systems on symmetric multiprocessors," [retrieved: May, 2013]. [Online]. Available: http://arco.unex.es/smpcache/

[31] W. Stallings, *Computer Organization and Architecture: Designing for Performance*, 6th ed. Prentice Hall, 2003.

[32] D. A. Patterson and J. L. Hennessy, "Computer organization and design, forth edition: The hardware/software interface," MA, USA, 2009.

[33] ACM/IEEE-CS Joint Interim Review Task Force, "Computer science curriculum 2008: An interim revision of cs 2001, report from the interim review task force," 2008. [Online]. Available: http://www.acm.org/education/curricula/ComputerScience2008.pdf

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 759–765

Concept of competence management system for Polish National Qualification Framework in the Computer Science area

Przemysław Różewski
West Pomeranian University of Technology ul. Żołnierska 49, 71-210 Szczecin, Poland
Email: prozewski@wi.zut.edu.pl

Bartłomiej Małachowski
West Pomeranian University of Technology ul. Żołnierska 49, 71-21Szczecin, Poland
Email: bmalachowski@wi.zut.edu.pl

Piotr Dańczura
West Pomeranian University of Technology ul. Żołnierska 49, 71-210 Szczecin, Poland
Email: piotrdanczura@gmail.com

***Abstract*—This article regards analysing the literature of processing competence in education, as well as competence management systems (CMS) and their role in developing competencies for students of higher education cycle. The Bologna Process and its results are described later in the text, explaining the need for National Qualification Frameworks and the benefits that they can produce when implemented correctly. We focuses on creating the basis for competence management system for Polish National Qualification Framework in Computer Science area, how it should work and how it should be implemented.**

I. Introduction

TECHNOLOGICAL market for jobs related with IT is constantly changing. The reasons for this are constant changes in technology and innovative products that affect the workings of certain services and sites. This results in constant changes in competences required by the IT market [23]. New competencies show up, they have new names and contents. Competencies that are already in existence, change their contents due to technological advancement.

The proposed system got two main roles: to be the Personal Competence Manager (PCM) for each student and to be the Organisational Competence Manager (OCM) for the given faculty. The system is based on competence development lifecycle [18] which includes elements like: creation of a reference competence description, the assessment of existing competences at individual or/and group level, the gap analysis, the definition of competence development programmes, continuous performance monitoring and assessment. Based on the PCM, users can choose their own competence development plans and follow them to built the desired competences. [20]. Based on OCM, the faculty can implement the following scenarios:

1. Knowledge analysis of a student that is trying to begin his second education cycle.
2. Market analysis to estimate how much student's competencies differ from those required by the market.
3. Reporting the level of students' summary competences, for example for accreditation committee.
4. Substantive evaluation of a given curriculum.

The definition of competence can be found in many scientific works [4], [8], [16]. Let us focus on those related with accepted standards. ISO 9000:2005 defines competence as the "demonstrated ability to apply knowledge and skills". ISO 19011 defines competence as "demonstrated personal attributes and demonstrated ability to apply knowledge and skills". ISO/IEC 17021:2011 defines competence as "ability to apply knowledge and skills to achieve intended results".IEEE Standard 1484.20.1-2007 [12] describes competency as "any aspect of competence, such as knowledge, skill, attitude, ability, or learning objective". In addition, there is a running discussion about difference between competence and competency term [8]. The IEEE Standard 1484.20.1-2007 interpreted the competency in the broadest sense to include learningobjectives (those things that are sought) as well as competencies (those things that are achieved).

The discussed competence management system (CMS) is being prepared for faculties related with computer science. Main goal of this system is to help to map the competences of a learner - which will be stored in some kind of a learner profile -with the competences that result from the competence development program[13]. In every higher education curriculum there are many types of different competencies [22]. While designing the concept of our Competence Management System we focused on key competences which we interpreted as Core Qualifications.

IT market is assessed for competencies in many research projects. The [15] is a good example of this practice where the standardisation of ICT job profiles was done and ICT job profiles model was defined. It is based on ontologies principle to describe "Knowledge Objects", "Skills", "Competences" and relations between them.

II. Problem Statement

Analysing main characteristics of a typical curriculum gives us the following mechanisms that develop competencies. Each curriculum consists of courses (subjects) realized in every next semester. On figure 1., an exemplary implementation of a given curriculum is shown, each circle represents a course. Some of them are extended for periods longer than 1 semester (for example 1.1, 2.1 and 2.3, 3.3. 4.3). In the course of studies, elective subjects start to appear from which student got to choose one (for example 4.6, 4.7, 4.8). Another matter are internships that student got to do (4.5). The curriculum ends with an implementation of specialisation courses (5.2-5.6, 6.2, 6.3), which affect the

[1] This work was supported by Project „Platforma Informatyczna TEWI", nr POIG.02.03.00-00-028/09,

overall profile of the student. It must be mentioned that the completion of each course means passing the course with a positive grade. Analysing the above educational path we can distinguish the following mechanisms that develop competencies:

- Competency is developed after completing the given course (e.g. after completing 1.2 or 3.2).
- Competency is developed after completing a certain set of related courses (e.g. 2.3, 3.3, 4.3) which can represent one bigger course divided into many stages (semesters).
- Competency is developed after finishing the internship (4.5).
- Competency is developed after completing a certain group of courses related to a certain technological or scientific aspect (3.1, 3.2, 4.1).
- Competency is developed after completing a specialisation (5.2-5.6, 6.2, 6.3).
- Competency is developed after completing the whole curriculum.

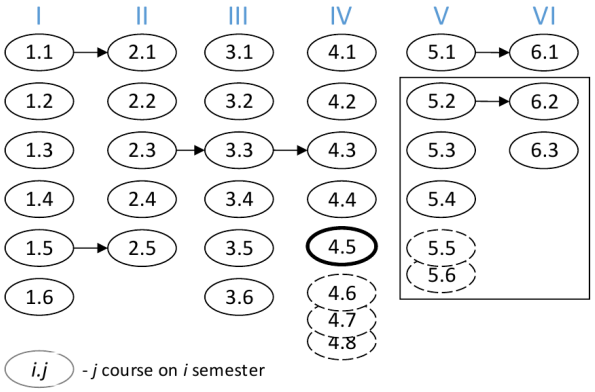


Fig. 1 Schematic representation of university curriculum

III. COMPETENCE MODELING IN IT EDUCATION

For every learning system Bloom's taxonomy is the basis for competencies description. In case of engineering sciences it can be expanded to two dimensions represented by cognitive and knowledge dimensions (Table I). Basic skills and features must be adjusted to the realities of IT's technical field. It seems that expanding skills' base with competencies related with mathematics and system analysis is a must. Typical mathematic-based competencies are[19]: thinking and reasoning, communication, argumentation, representation, modeling, problem posing and solving, symbolic and technical language.

Basing on literature we can define the target set of computer science student's competencies as a total set of knowledge, technology, skills and attitudes which function as action characteristics of an organizational member who can do his or her tasks outstandingly and efficiently in the computing environment [24]. Generally speaking student's computing competency consists of four components [24]:

- The computing mindset (driven from self-concepts and traits).
- The knowledge of computing technology (based on knowledge).
- The capability of computing application (determined from cognitive and behaviour skills).
- The potential of computing capability (driven from personal motives).

Each student upon completion of a given curriculum should posses a set of Core Competencies which are the basis for typical problem solving in the field of IT. The core competence in the literature on education defines a set of learning outcomes (skills or competencies) which each individual should acquire during or demonstrate at the end of a period of learning. It is one of a number of associated con-

TABLE .I
BLOOM TAXONOMY (based on [11])

Cognitive dimension		**Knowledge dimension**	
Remember	Exhibit the memory of previous-learned materials by recognizing or recalling facts, terms, basic concepts and answers.	*Factual Knowledge*	Knowledge about terminology and specification details.
Understand	Understanding of facts and ideas by interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining main ideas.	*Conceptual Knowledge*	Knowledge about generation, classification, and structural modelling of certain concept.
Apply	Using the available knowledge to execute and implement solutions in different ways.	*Procedural Knowledge*	Knowledge about workflows, algorithms, methods, procedures, and events.
Analyze	Differentiating, organizing, and attributing knowledge by manipulating information using certain criteria.	*Meta-Cognitive Knowledge*	Knowledge about strategies and decisional conditions.
Evaluate	Checking and Judgments about information, validity of proposed ideas, or quality of work by certain criteria.		
Create	Generating, planning, and producing information or knowledge together and proposing new solutions.		

cepts, including core skills, core competency, generic skills and key qualifications [10]. According to [21] the core competence applied to education as a whole could be defined as facilitating the empowerment of people, through learning how to acquire information (data), turn it into knowledge and apply that knowledge to solve problems. The example of core competence for network building can be found in [9].

IV. National Qualifications Framework: an overview

A. Bologna Process and Lifelong Learning

Proposed in 1999 by Education Ministers from 29 European Countries, Bologna Process was started to create the European Higher Education Area (EHEA) [16]. After a series of ministerial meetings (Prague 2001, Berlin 2003, Bergen 2003, London 2007, Leuven 2009) and by the year 2013 there are now 47 participating countries in the Bologna Process. The main purpose of it was to create the Qualifications Framework of the EHEA which were greatly influenced by the UK's National Qualifications Framework (NQF) [2]and its later version, the Qualifications and Credit Framework (QCF) [7]. During this process the European Qualifications Framework which acts as a medium to translate national qualifications across European countries, was created. This way, workers and students in European Union gain more mobility between countries allowing them to study or work abroad without the difficulties of complicated analysis of their current competencies, knowledge and skills. Many other countries took prime example of UK and also implemented 8-level NQFs into their education systems. Those national qualifications can be easily translated to EQF and people who moved from one European country to another would not have to repeat what they already learned. The Bologna Process moved on to create the "Bologna qualifications framework".

Another accomplishment of the Bologna Process was creating the idea of Lifelong Learning. Pursuing knowledge for either personal or professional reasons for the individual's entire life rather than only learning "in the classroom". Lifelong Learning focuses on teaching outside schools and universities, using methods like home schooling, education for adults (for individuals that want to develop themselves), continued education (usually extended courses offered by higher education institutions), working with knowledge (using obtained knowledge in professional work) and personal learning (individuals learning using for example online sources of distance education).

B. How NQF works

EQF just like British NQF/QCF is divided into 8 levels, each of them describing what the learner knows, understand and is able to do (where 8th level is the most advanced and 1st level is the most basic). We can already see the similarity to British NQF/QCF which also had 8 levels and the last one was the most advanced[European Communities 2008]. The Bologna qualifications framework states that there are 3 cycles/levels of qualifications framework:

- Cycle 1 - usually correlated to qualifications of bachelor's degree
- Cycle 2 - usually correlated to qualifications of master's degree
- Cycle 3 - usually correlated to qualifications of doctoral degree

Those levels correspond naturally with NQF and thus EQF last 3 levels (6th, 7th and 8th) as they describe the same levels of education cycle. In Bergen 2005 all higher education ministers agreed on EQF levels 6-8 descriptors (as in Table II) to be also the descriptors for the three education cycles of qualification frameworks within European Higher Education Area

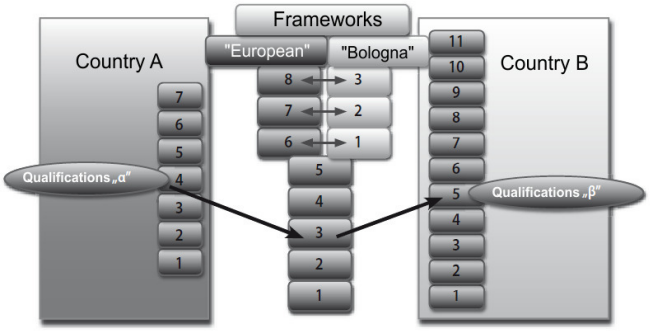


Fig. 2. EQF to NQF translation(based on [3])

Not every NQF easily translates to EQF on a 1:1 level. Different countries have adopted NQF system adjusted to their needs and modified it greatly. Figure 3. is an example of NQFs of France and England compared to EQF.

For example, the last three levels of EQF directly correspond with English and French NQFs but fifth level is different for both countries. This situation is similar with other countries' NQFs. That is one of the reasons Bologna Process focused on the last free cycles of education: bachelor's, master's and doctoral degree (the names for each degree can differ in various countries).

Introduction to European Credit Transfer and Accumulation System (ECTS) was necessary to make the 'translations' between different countries possible. Thus, each cycle referred to certain amount of ECTS credits/points:

- Level 1 - typically 180-240 ECTS credits
- Level 2 - typically 90-120 ECTS credits
- Level 3 - no ECTS

Usually 60 ECTS credits corresponds with 1 academic year which is equivalent to around 1500-1800 hours of study (during classes).

TABLE .II
DESCRIPTORS DEFINING LEVELS IN THE EUROPEAN QUALIFICATIONS FRAMEWORK (based on[8])

Level	Knowledge	Skill	Competence
1	Basic general knowledge	basic skills required to carry out simple tasks	work or study under direct supervision in a structured context
2	Basic factual knowledge of a field of work or study	basic cognitive and practical skills required to use relevant information in order to carry out tasks and to solve routine problems using simple rules and tools	work or study under supervision with some autonomy
3	Knowledge of facts, principles, processes and general concepts, in a field of work or study	a range of cognitive and practical skills required to accomplish tasks and solve problems by selecting and applying basic methods, tools, materials and information	take responsibility for completion of tasks in work or study; adapt own behaviour to circumstances in solving problems
4	Factual and theoretical knowledge in broad contexts within a field of work or study	a range of cognitive and practical skills required to generate solutions to specific problems in a field of work or study	exercise self-management within the guidelines of work or study contexts that are usually predictable, but are subject to change; supervise the routine work of others, taking some responsibility for the evaluation and improvement of work or study activities
5	Comprehensive, specialised, factual and theoretical knowledge within a field of work or study and an awareness of the boundaries of that knowledge	a comprehensive range of cognitive and practical skills required to develop creative solutions to abstract problems	exercise management and supervision in contexts of work or study activities where there is unpredictable change; review and develop performance of self and others
6* Cycle 1	Advanced knowledge of a field of work or study, involving a critical understanding of theories and principles	advanced skills, demonstrating mastery and innovation, required to solve complex and unpredictable problems in a specialised field of work or study	manage complex technical or professional activities or projects, taking responsibility for decision-making in unpredictable work or study contexts; take responsibility for managing professional development of individuals and groups
7* Cycle 2	Highly specialised knowledge, some of which is at the forefront of knowledge in a field of work or study, as the basis for original thinking and/or research Critical awareness of knowledge issues in a field and at the interface between different fields	specialised problem-solving skills required in research and/or innovation in order to develop new knowledge and procedures and to integrate knowledge from different fields	manage and transform work or study contexts that are complex, unpredictable and require new strategic approaches; take responsibility for contributing to professional knowledge and practice and/or for reviewing the strategic performance of teams
8* Cycle 3	Knowledge at the most advanced frontier of a field of work or study and at the interface between fields	the most advanced and specialised skills and techniques, including synthesis and evaluation, required to solve critical problems in research and/or innovation and to extend and redefine existing knowledge or professional practice	demonstrate substantial authority, innovation, autonomy, scholarly and professional integrity and sustained commitment to the development of new ideas or processes at the forefront of work or study contexts including research

C. Learning outcomes

Obtaining given knowledge, skill and competency means that student has reached the 'learning outcomes' planned to achieve after completing his education [1]. Whether we refer to only one lesson, learning module or the whole education cycle it does not matter, learning outcomes can correspond to all of these.

Learning outcomes can be divided into three groups: generic (outcomes that are general for a certain cycle of education, e.g. for outcomes for each bachelor's degree studies), field (outcomes that are specific for a certain type of studies, e.g. technical university), specific (outcomes that are specific for a certain learning module curriculum, major etc.). The general learning outcomes defined by NQF Work Group are also defined and incorporated into the first group. They can be defined by both Ministry of Education and the university as well, though they are distinguished which are which.

V. COMPETENCE (QUALIFICATION) MANAGEMENT SYSTEM

Student while completing the courses on studies achieve learning outcomes defined for the course, thus achieving learning outcomes for his/her major, and those learning outcomes correspond with learning outcomes by the Ministry of Higher Education. Achieving learning outcomes for a specific course means that the student posses a set of knowledge, skills and competencies. Each of those three can corre-

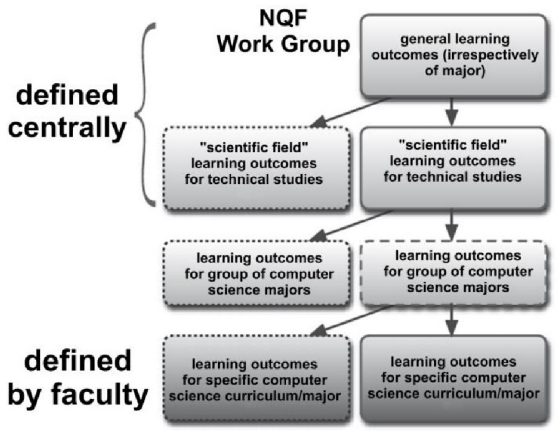


Fig.3 Defining learning outcomes in Poland (based on [3])

spond to a different learning outcome. The problem lies in finding an easy method to name and manage the key competencies we want to use in our management system, it is also essential for processing those sets in the system. That is when Core Qualifications (to correspond with the NQF name) term come in. Core Qualifications are combinations of knowledge, skills and competence, all of them contributing in the European Qualifications Framework. Using the same set of learning outcomes (knowledge, skill competence) as in the EQF makes it possible to easily process and compare Polish NQF with EQF but to make it even more adaptable Core Qualification (QF) must be implemented. In the future this system will be able compute the CQs of students from different Universities based exactly on learning outcomes and this knowledge, skill, competence sets.

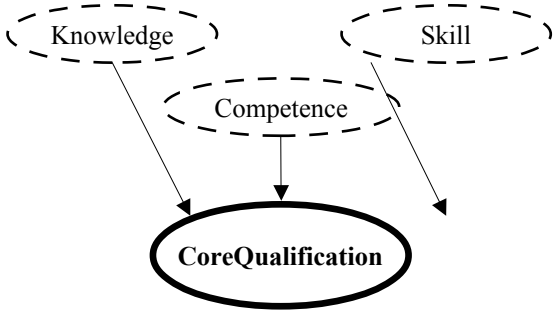


Fig. 4 Core Qualification components

Figure 6. presents the class diagram of NQF based on Computer Science Faculty at West Pomeranian University of Technology in Szczecin. The courses are created accordingly to EQF standards designed by the Bologna Process. The diagram clearly shows 3 distinctive learning outcomes groups like mentioned before. Although the 'specific learning outcomes' in the system are shown to more correspond with the specialisation instead of major, they mostly refer to the same learning outcomes designed for the major (or rather scientific field). The difference between this diagram relations and the learning outcomes mentioned before comes from the University-specific preferences. Sets of Knowledge, Skills and Competence visible from learning outcomes for specialisation and course will be the basis to create the Core Qualification.

The proposed system is using the Competence Object Library (COL) [17] for competence modelling. The COL based on the TENCompetence Domain Model (TCDM) for competence structure modelling and the competence set theory for fuzzy competence set expansion cost analysis [14]. The COL on one hand enables to model different kind of competence content, on the other hand the system can perform quantitative analysis of competence. The COL defines following classes [17]:

- Competency : any form of knowledge, skill, attitude, ability or learning objective that can be described in a context of learning, education, training or any specific business context.
- Competence: effective performance of a person within a context at a specific level of proficiency.
- Context: circumstances and conditions surrounding actions performed by a person.
- Category: indicates the relative level in a taxonomic hierarchy.
- Proficiency Level: indicates the level at which the activity of a person is considered.
- Relation: arbitrary association of competencies within a context and at specific proficiency level.
- Element of Competence: entity derived from competence that can form a set.
- Competence Set: collection of elements of competence. The system has to support the function CompetenceSet.CompareSet() with quantity outcome.
- Competence Profile: collection of competence sets. There is a relented function Competence-Profile.CompareProfile() for different profiles comparison.
- Required Competence Profile: requirements in terms of competence to be fulfilled by a person.
- Acquired Competence Profile : description of competencies possessed by a person.
- Learning outcomes: activity, job, skill, attitude, ability or learning objective for which competence requirements can be specified.
- Person (Student): competent actor performing activities.

VI. CONCLUSION

In this article, we focused on showing the complexity of the issue of competence management in university. The main area of study was computer science, because it is possible to use multiple reference models showing areas of expertise in the field (e.g. 2012 ACM Computing Classification System). Such knowledge-base enables in the future to decide on the level of student's competence and knowledge using the automatic ontology processing methods.

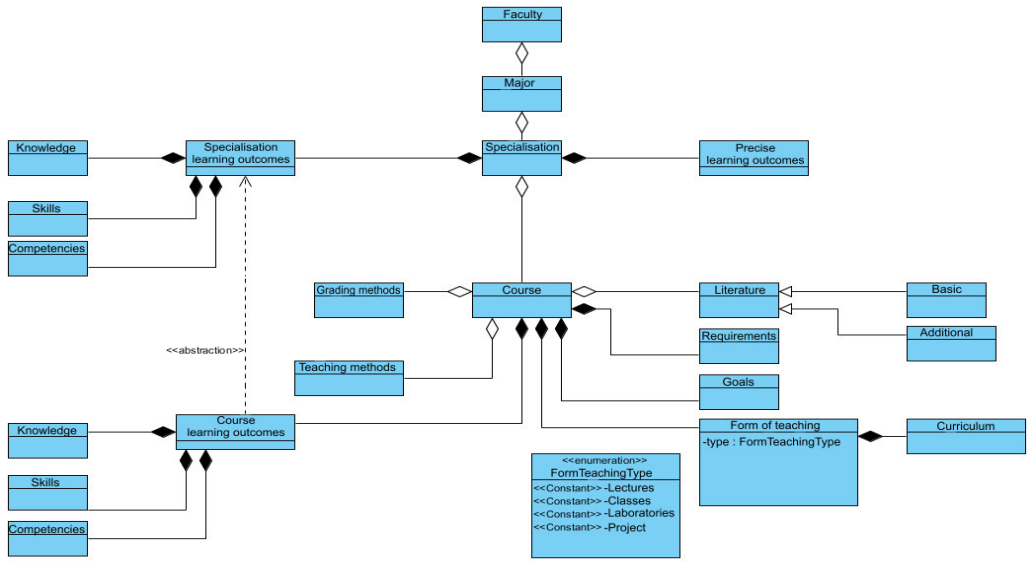


Fig. 5 Class diagram of the NQF at West Pomeranian University of Technology

The Bologna Process is a social and organizational program that includes a set of activities. Among them the qualification framework is one of the most important. The competence management system increasing the possibility of changing qualifications of the student depending on the market requirements.

REFERENCES

[1] S. Armstrong, "Natural Learning in Higher Education", Encyclopedia of the Sciences of Learning, Springer 2012.

[2] Bologna Working Group on Qualifications Frameworks, "A Framework for Qualifications of the European Higher Education Area", Ministry of Science, Technology and Innovation, Copenhagen 2005.

[3] E. Chmielecka, Z. Marciniak, A. Kraśniewski, "Autonomia Programowa Uczelni - Ramy kwalifikacji dla szkolnictwa wyższego", Warszawa, 2010.

[4] F. Draganidis, G. Mentzas, "Competency based management: a review of systems and approaches", Information Management & Computer Security, vol. 14, no. 1, pp. 51-64, 2006.

[5] European Qualifications Framework Website, European Commission - Education and Culture, http://ec.europa.eu/eqf/, access: May 2013

[6] European Commission - Education and Culture, "The European Qualifications Framework for Lifelong Learning (EQF)", European Communities, 2008

[7] European Association for Quality Assurance in Higher Education, "Standards and Guidelines for Quality Assurance in the European Higher Education Area", Helsinki, 2005.

[8] S. Grant, R. Young, "Concepts and Standardization in Areas Relating to Competence" International Journal of IT Standards and Standardization Research, vol. 8, no. 2, pp. 29-44, 2010.

[9] K. Hajlaoui, X. Boucher, M. Beigbeder, J. J. Girardot, "Competence Ontology for Network Building," in Proc. IFIP Advances in Information and Communication Technology, Vol. 307, pp. 282-289, 2009.

[10] G. Holmes, N. Hooper "Core competence and education," Higher Education, vol. 40, no. 3, pp. 247-258, October 2000.

[11] S.-M. Huang, H.-Y. Hsueh, J.-S. Hua, "Discovery of Educational Objective on E-Learning Resource: A Competency Approach", in Proc. Advances in Web Based Learning – ICWL 2007, LNCS, Springer, Vol. 4823, 2008, pp 618-629.

[12] IEEE Standard for Learning Technology - Data Model for Reusable Competency Definitions, IEEE Standard 1484.20.1, 2007.

[13] M. Kalza, J. van Bruggena, E. Rusmana, B. Giesbersa, and R. Kopera, "Positioning of learners in learning networks with content, metadata and ontologies, " Interactive Learning Environments, vol. 15, no. 2, pp. 191-200, 2007.

[14] E. Kushtina, O. Zaikin, P. Różewski, B. Małachowski, "Cost estimation algorithm and decision-making model for curriculum modification in educational organization," European Journal of Operational Research, vol. 197, no. 2, pp. 752-763, 2008.
[15] B. Pernici, P. Locatelli, C. Marinoni, "The eCCO System: An eCompetence Management Tool Based on Semantic Networks," in Proc. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, LNCS, Springer, Vol. 4278, 2006, pp. 1088-1099.
[16] P. Różewski, E. Kusztina, R. Tadeusiewicz, O. Zaikin, "Intelligent Open Learning Systems: Concepts, models and algorithms". Intelligent Systems Reference Library, Vol. 22, Springer-Verlag Berlin Heidelberg (2011)
[17] P. Różewski, B. Małachowski , "Competence Management In Knowledge-Based Organisation: Case Study Based On Higher Education Organisation," in: D. Karagiannis and Z. Jin (Eds.): KSEM 2009, LNAI 5914, pp. 358—369, 2009.
[18] D. G. Sampson, "Competence-related Metadata for Educational Resources that Support Lifelong Competence Development Programmes," Educational Technology & Society, vol. 12, no. 4, pp. 149-159, 2009.
[19] C. Sáenz, "The role of contextual, conceptual and procedural knowledge in activating mathematical competencies (PISA)," Educational Studies in Mathematics, vol. 71, no. 2, pp. 123-143, June 2009.
[20] K. Stefanov, N. Nikolova, M. Ilieva, E. P. Stefanova, "Turning university professors into competent learners," International Journal of Emerging Technologies in Learning (iJET), vol. 3, pp. 46-52, 2008.
[21] T.A. Stewart, Intellectual Capital. Brealey, London , 1999.
[22] H. Vogten, H. Martens, R. Lemmers, "The TENCompetence Infrastructure: A Learning Network Implementation," in Learning Network Services for Professional Development, R. Koper (ed.), Berlin Heidelberg: Springer-Verlag, 2009, pp. 329-357.
[23] Ch. Wai-MuiYu, Ch. Velde, "The Changing Context of Business Education: Competency Requirements for the New Paradigm," in International Perspectives on Competence in the Workplace, C.R. Velde (ed.), Berlin Heidelberg: Springer-Verlag, 2009, pp. 57-85.
[24] Ch. Y. Yoon, "A structural model of end-user computing competency and user performance," Knowledge Based Systems, vol. 21, no. 5, pp. 415-420, 2008.

Frontiers in Network Applications, Network Systems and Web Services

SYMPOSIUM SoFAST-WS focuses on modern challenges and solutions in network systems, applications and service computing. The Symposium builds upon the success of Frontiers in Network Applications and Network Systems (FINANS'2012) and 4th International Symposium on Web Services (WSS' 2012) held in 2012 in Wroclaw, Poland. These two events are now integrated into one event to fully exploit the synergy of topics and cooperation of research groups.

The topics discussed during the symposium include different aspects of network systems, applications and service computing. The primary objective of the symposium is to bring together researchers and practitioners analyzing, developing and administering network systems, with particular emphasis on Internet systems. Authors are invited to submit their papers in English, presenting the results of original research or innovative practical applications in the field.

Topics

Topics include (but are not limited to):

- Architecture, scalability and security of Open API solutions,
- Technical and social aspects of Open API and open data,
- Service delivery platforms - architecture and applications,
- Telecommunication operators API exposition in Telco 2.0 model,
- The applications of intelligent techniques in network systems,
- Mobile applications,
- Network-based computing systems,
- Network and mobile GIS platforms and applications,
- Computer forensic,
- Network security,
- Anomaly and intrusion detection,
- Traffic classification algorithms and techniques,
- Network traffic engineering,
- High-speed network traffic processing,
- Heterogeneous cellular networks,
- Wireless communications,
- Security issues in Cloud Computing,
- Network aspects of Cloud Computing,
- Control of networks,
- Standards for Web services,
- Semantic Web services,
- Context-aware Web services,
- Composition approaches for Web services,
- Security of Web services,
- Software agents for Web services composition,
- Supporting SWS Deployment,
- Architectures for SWS Deployment,
- Applications of SWS to E-business and E-government,
- Supporting Enterprise Application Integration with SWS,
- SWS Conversational Protocols and Choreography,
- Ontologies and Languages for Service Description,
- Ontologies and Languages for Process Modeling,
- Foundations of Reasoning about Services and/or Processes,
- Composition of Semantic Web Services,
- Innovative network applications, systems and services.

Event Chairs

Furtak, Janusz, Military University of Technology, Poland

Grzenda, Maciej, Orange Labs Poland and Warsaw University of Technology, Poland

Legierski, Jarosław, Orange Labs Poland, Poland

Luckner, Marcin, Warsaw University of Technology, Poland

Szmit, Maciej, Orange Labs Poland, Poland

Program Committee

Afonso, Joao, Foundation for National Scientific Computing, Portugal

Baghdadi, Youcef, Sultan Qaboos University, Oman

Benslimane, Sidi Mohammed, University of Sidi Bel-Abbès, Algeria

Chainbi, Walid, ENISO, Tunisia

Chojnacki, Andrzej, Military University of Technology, Poland

Dabrowski, Andrzej, Warsaw University of Technology, Poland

Davies, John, Glyndwr University, United Kingdom

Fernández, Alberto, Universidad Rey Juan Carlos, Spain

Frankowski, Jacek, Orange Labs, Poland

Fuchs, Lothar, Institute for technical and scientific hydrology, Germany

Furtak, Janusz, Military University of Technology, Poland

Gaaloul, Walid, Institut Mines Télécom, France

García-Domínguez, Antonio, University of Cádiz, Spain

García-Osorio, César, University of Burgos, Spain

Grabowski, Sebastian, Orange Labs, Poland

Kiedrowicz, Maciej, Military University of Technology, Poland

Korbel, Piotr, Technical University of Lodz, Poland

Kowalczyk, Emil, Orange Labs, Poland

Kowalski, Andrzej, Orange Labs, Poland

López Nores, Martín, University of Vigo, Spain

Macukow, Bohdan, Warsaw University of Technology, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 769–774

Genetic Algorithm with Different Feature Selection Techniques for Anomaly Detectors Generation

Amira Sayed A. Aziz[1,*], Ahmad Taher Azar[2,*], Mostafa A. Salama[3,*]
Aboul Ella Hassanien[4,*] and Sanaa El-Ola Hanafy[4]
[1]Universite Francaise d'Egypte (UFE), Cairo, Egypt
[2]Faculty of Computers and Information, Benha University, Egypt
[3]British University in Egypt (BUE), Cairo, Egypt
[4]Faculty of Computers and Information, Cairo University, Egypt
[*]Scientific Research Group in Egypt (SRGE)
http://www.egyptscience.net

***Abstract*—Intrusion detection systems have been around for quite some time, to protect systems from inside ad outside threats. Researchers and scientists are concerned on how to enhance the intrusion detection performance, to be able to deal with real-time attacks and detect them fast from quick response. One way to improve performance is to use minimal number of features to define a model in a way that it can be used to accurately discriminate normal from anomalous behaviour. Many feature selection techniques are out there to reduce feature sets or extract new features out of them. In this paper, we propose an anomaly detectors generation approach using genetic algorithm in conjunction with several features selection techniques, including principle components analysis, sequential floating, and correlation-based feature selection. A Genetic algorithm was applied with deterministic crowding niching technique, to generate a set of detectors from a single run. The results show that sequential-floating techniques with the genetic algorithm have the best results, compared to others tested, especially the sequential floating forward selection with detection accuracy 92.86% on the train set and 85.38% on the test set.**

***Index Terms*—Anomaly detectors generation, genetic algorithms, feature selection**

I. Introduction

WITH the expanding and increasing use of networks, and accumulating number of internet users, network throughput has become massive and threats are more diverse and sophisticated. Network and information security are of high importance, and research is continuous in these fields to keep up with the development of attacks. Intrusion Detection is a major research area that aims to identify suspicious activities in a monitored system, from authorized and unauthorized users. An Intrusion Detection System (IDS) could be host-based or network-based. In Network IDSs (NIDS), network administrators are not able to keep up with the increase in network attacks number and complexity, for both known and unknown attacks. So, there is an urgent and pressing need for replacing them by automated systems for constant monitoring and quick responses [1]. Machine Learning techniques are used to create rules for an IDS by enhancing the domain knowledge. They improve performance by helping to automate the knowledge acquisition process by using training data to find and exploit regularities. They learn how to estimate knowledge from the training data sets. Because one of the biggest challenges of IDS is the massive amount of data collected from the system, learning algorithms are used to discover models that are appropriate to classify normal and anomalous behaviors [2][3].

Different machine learning (ML) techniques are used in the development of anomaly intrusion detection. They could be pattern classifiers, single classifiers, hybrid classifiers, or ensemble classifiers. Such techniques include Support Vector Machine (SVM), Artificial Neural Networks (ANN), Self-Organizing Maps (SOM), Decision Trees (DT), Genetic Algorithms (GA), and many others. In this paper, a genetic algorithm is used to generate detectors for an anomaly intrusion detection system. Some feature selection techniques are applied for feature reduction, before applying the GA.

The paper is organized as follows: Section II gives a background for different components of the system. Section III introduces the proposed anomaly detectors generation algorithm. Section IV describes the experiment steps and the involved data sets. Section V shows the experiment result. Conclusions and future work are discussed in Section VI.

II. Background

A. Anomaly Intrusion Detection

Intrusion Detection Systems can be classified - based on methodology - into misuse-based and anomaly-based. Misuse-based IDS builds a database of attacks signatures and use them to detect anomalies. It's accurate and definitive concerning detecting unknown that do not match the stored patterns. This is where the anomaly-based IDS acts better. Anomaly IDS builds a model that represents the normal behaviour of a system and assume that deviations from such model are attacks or suspicious activity. So, it detect unknown attacks but it may have a high false alarm rate, based on its adjustments. A model in anomaly IDS can be statistical, knowledge-based, or machine learning. A learning technique could be embedded too, to update and adjust the normal model from time to time,

978-83-60810-53-8/$25.00

to represent the actual system behaviour that may change from time to time [4][5].

B. Feature Selection

For each problem with some sample, there is a maximum number of features where performance degrades instead of improves – which is called the curse of dimensionality. An accurate mapping of lower-dimensional space of features is needed so no information is lost by discarding important and basic features. Two issues one should pay attention to while doing this: (1) How dimensionality can affect classification accuracy and (2) How dimensionality affects a classifier complexity. A feature is good when it is relevant but not redundant to the other relevant features. There are two techniques to follow for this: feature extraction and feature selection. Feature extraction algorithms tend to create a new subset of features by combining existing features. Feature selection (FS) algorithms tend to limit the features to only those which would improve a task performance. The FS [6][7][8] is an essential machine learning technique that is important and efficient in building classification systems. When used to reduce features, it results in lower computation costs and better classification performance. Feature selection algorithms are composed of three components: search algorithm, evaluation function, and performance function. The search algorithm could be: exponential – which is expensive to use as they have exponential complexity in number of features, sequential where it adds and subtracts features, so they have polynomial complexity; or randomized – where it require biases to yield small subsets, and they usually achieve high accuracies. An objective function is a function to evaluate the candidate features for feature selection.

Based on evaluation criteria, FS techniques can be divided into filter methods and wrapper methods. Filters evaluate feature subsets by their information content, using distance measures, correlation measures...etc. Wrappers use a classifier for features subset evaluation by their predictive accuracy. Filter techniques discards feature upon their evaluation based on data general characteristics or using some kind of statistical analysis, without any learning mechanism involved. Wrapper techniques use a learning algorithm to find the features subset with the best performance. They are more expensive computation-based, and slower due to the repeating process, but they give more accurate results than filter techniques. This might be a drawback for high dimensional data but it could be defeated by using a fast learning algorithm.

1) Correlation-based Feature Selection: Correlation Feature Selection (CFS) is a heuristic approach that evaluates the worthiness of a features subset. So, based on correlation concept, a feature is considered good if it is highly correlated to the class but not to the other features [9]. So, a suitable measure of correlation between features needs to be defined in which it represents the important and highly effective features. A function that evaluates the best individual feature is:

$$M = \frac{k * r_{fc}}{\sqrt{k + k * (k-1) * r_{ff}}} \tag{1}$$

Where: M is the heuristic merit of a features subset S containing K features, $r_{fc}$is the average feature-class correlation, and $r_{ff}$is the average feature-feature inter-correlation.

The numerator indicates how predictive a group of features are; and the denominator indicates how much redundancy there is among those features.

2) Sequential Floating Selection: Sequential-floating selection is a flexible extension of sequential forward and backward selection (SFS and SBS respectively), with backtracking capabilities [10][11]. Sequential selection methods find optimal group of features by applying step-optimal method where in each step the best/worst feature is always added/discarded. No additional steps are taken to evaluate the selected features in each iteration to refine the subset. An improvement to the methodology is to apply the plus l-take away r method, where additional forward/backward steps are applied after each iteration to correct the selection decision in order to find the optimal final subset of features. If $l > r$ then is Sequential-floating forward selection, if $l < r$ then it is sequential-floating backward selection [12]. Sequential-Floating Forward Selection (SFFS) basically starts with an empty set, then at each iteration it adds sequentially the next best feature. Then, it tests if it maximizes the objective function when combined with the features already selected – and this is how SFS works. In addition to that, after each forward step, SFFS performs a backward step that discards the worst feature of the subset after a new feature is added. The backward steps are performed as long as the objective function is increasing.

In a similar manner but in an opposite direction, Sequential-Floating Backward Selection (SFBS) starts with the full set of features, then it sequentially removes the feature that least reduces the objective function value. This is how the SBS works. So, SFBS performs forward steps after each backward step, as long as the objective function increases.

3) Principle Components Analysis: Principal Components Analysis (PCA) is a way to find and highlight similarities and differences between data by identifying the existing patterns [13][14]. PCA is based in the idea that most information about classes are within the directions with the largest variations. It works in terms of standardized linear projection that maximizes the variance in the projected space. It is a powerful tool in the case of high-dimensional data. It calculates the eigenvectors of the covariance matrix to find the independent axes of the data. The main problem with PCA is that it does not take into consideration the class label of the feature vector, hence it does not consider class separability.

4) Information Gain: Information Gain (IG) help us to determine which feature is most useful for classification, using its entropy value. Entropy indicates the information content of a feature or how much information it is giving us. The higher the entropy, the more the information content. IG value is calculated as:

$$IG(T,a) = H(T) - H(T|a) \tag{2}$$

where H is the information entropy, T is a training example, and a ia a variable value. This equation calculates the IG

that a training example T obtains from an observation that a random variable A takes some value a. IG in machine learning is used to define a sequence of attributes to investigate, which leads to the building of a decision tree. A decision tree can be constructed top-down using the IG by first beginning at the root node, and use the attribute with the highest IG as an ancestor node. Then child nodes are added for each possible value of that attribute. All examples are attached to suitable child nodes where the examples values are identical to the node's attached value. If the all examples attached to a child node can be labelled with a unique class label, then this node is marked as a leaf and that classification is added to the node. These steps are repeated until all classifications are added to the child nodes [15][16][17].

5) Rough Sets: Rough Set Theory (RST) [18][19] can be used to discover dependencies in data for features reduction, using the data alone without additional information required. RST results in the most informative feature set, which would be the most predictive of the class label. The concepts are used to define necessity of features, which is calculated by functions of lower and upper approximations. The measures of necessity are employed as heuristics to guide the feature selection process. An information table is defined for the objects and the attributes as a tuple:

$$T = (U, A) \tag{3}$$

where U is a finite non-empty set of the primitive objects, and A is a finite non-empty set of the attributes. Each feature is associated with a set of its values V. The attribute set would be divided into 2 subsets C and D, which are the condition and decision subsets, respectively. If P is a subset of A, the in-discernibility relation is defined as:

$$IND(P) = \{(x,y) \in UxU : \forall a \in P, a(x) = a(y)\} \tag{4}$$

where $a(x)$ is a feature value a of object x. x and y are said to be in-discernible with respect to P, if $(x,y) \in IND(P)$

C. Genetic Algorithms

As mentioned before, ML techniques are used to create rules for the intrusion detection systems, and genetic algorithms is a common algorithm that is been used for such purpose. Genetic Algorithms (GA) are search algorithms inspired by evolution and natural selection, and they can be used to solve different and diverse types of problems. The algorithm starts with a group of individuals (chromosomes) called a population. Each chromosome is composed of a sequence of genes that would be bits, characters, or numbers. Reproduction is achieved using crossover (2 parents are used to produce 1 or children) and mutation (alteration of a gene or more). Each chromosome is evaluated using a fitness function, which defines which chromosomes are highly-fitted in the environment. The process is iterated for multiple times for a number of generations until optimal solution is reached. The reached solution could be a single individual or a group of individuals obtained by repeating the GA process for many runs [20][21].

D. Negative Selection Approach

Negative Selection Approach (NSA) as an artificial immune system (AIS) technique that is based on the self/non-self discrimination. It first builds a database of normal profiles, and then trains the detectors on that profile to be able to detect anomalous behaviour (that is not normal), when they are later released in the system. The detectors do this discrimination process by being able to recognize normal patterns, then any pattern that is not recognized as normal is considered anomalous. For its similarity to the anomaly detection concept, NSA has been widely applied in different anomaly intrusion detection and fault detection systems in different areas. NSA is very common and easy to implement, and it gives very good results specially when it is combined with classification techniques. Detectors generated and used are simply rules, that define low and high limits, or specific values, of the features used for the intrusion detection [22][23].

III. The Proposed Anomaly Detectors Generation Approach

In this paper, a genetic algorithms is applied with deterministic crowding niching technique, to generate a set of detectors in a single run. The detectors are generated following the NSA concept – as mentioned before – using the self samples from the training set. The original algorithm was originally implemented in [24] using only real-valued features. Then it was applied in [25] using a predefined set of features of different data types. Equal-width binning is applied on the features values as a preprocessing step to first create homogeneity between features, and second to discretize the values so that there would be no extreme values in the experiment. The number of bins for each feature was defined using the following formula.

$$k = max(1, 2 * \log l) \tag{5}$$

where l is the number of observed values. Then each value is replaced with the enclosing bin (the bin that includes this value). Algorithm 1 shows the detailed steps of the detectors generation approach.

A set S of normal connections (self particles) is defined in the beginning, to start generating the detectors that will be able to match normal connections in the future. The self space S is filled with randomly selected individuals from the normal connections in the training set. An individual is composed of a set of values representing the selected features for the IDS. Going through the GA process, crossover and mutation are applied where crossover provides exploitation and mutation provides exploration. The final result is a set of detectors (individuals) with features values that most represent normal connections.

The fitness of an individual is measured by calculating the matching percentage between an individual and the normal samples, as follows:

$$fitness(x) = \frac{a}{A} \tag{6}$$

Algorithm 1 GADG

Fill S Space with normal individuals from training set.
Run equal-width binning algorithm on continuous features.
Initialize population by selecting random individuals from the space S.
for The specified number of generations **do**
  for The size of the population **do**
    Select two individuals (with uniform probability) as $parent_1$ and $parent_2$.
    Apply crossover to produce a new individual (*child*).
    Apply mutation to child.
    Calculate the distance between *child* and $parent_1$ as d_1, and the distance between *child* and $parent_2$ as d_2.
    Calculate the fitness of *child*, $parent_1$, and $parent_2$ as f, f_1, and f_2 respectively.
    if $(d_1 < d_2)$ and $(f > f_1)$ **then**
      replace $parent_1$ with *child*
    else
      if $(d_2 <= d_1)$ and $(f > f_2)$ **then**
        Replace $parent_2$ with *child*.
      end if
    end if
  end for
end for
Extract the best (highly-fitted) individuals as your final solution.

where a is the number of samples matching the individual by 100% , and A is the total number of normal samples. where a is the number of samples matching the individual, and A is the total number of normal samples.

Both the Euclidean and Minkowski distance measures were tested in the GA – each in a separate trial – to calculate the distance between a child and a parent. The Euclidean distance is also used in the discrimination function to detect anomalies. The Euclidean distance between two individuals is calculated as follow:

$$d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \ldots (x_n - y_n)^2} \quad (7)$$

The Minkowski distance, which uses the p-norm dimension as the power value, between two individuals is calculated as:

$$d(X,Y) = (\sum_{i=0}^{n} (|x_i - y_i|^p))^{1/p} \quad (8)$$

IV. Experimental results and discussion

A. Data Set

The NSL-KDD [26] data set was used in the experiment, as it is more refined and less biased than the original KDD Cup99 data set [27]. In addition to that, it contains much less number of records, so the whole training and test sets can be used in the experiments. Table I shows the distributions of normal and attacks records in the NSL-KDD data set.

TABLE I
DISTRIBUTIONS OF NSL-KDD RECORDS

	Total Records	Normal	DoS	Probe	U2R	R2L
Train_20%	25192	13449	9234	2289	11	209
		53.39%	36.65%	9.09%	0.04%	0.83%
Train_All	125973	67343	45927	11656	52	995
		53.46%	36.456%	9.25%	0.04%	0.79%
Test+	22544	9711	7458	2421	200	2754
		43.08%	33.08%	10.74%	0.89%	12.22%

B. Feature Selection

Each one of the feature selection – that were mentioned earlier – was applied to the NSL-KDD data set, and each came up with a set of feature, shown in table II. In addition to the FS techniques applied in this paper, other two sets of features selected in [3][28] using Information Gain (IG) and Rough Set (RS) for degree of dependency respectively, were used.

TABLE II
SELECTED FEATURES

FS Technique	Selected features
CFS	3, 4, 8, 12, 29, 33, 39
SFBS	2, 3, 6, 29, 34, 35, 36, 37, 38, 39, 40, 41
SFFS	1, 2, 3, 4, 6, 7, 10, 12, 17, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 40, 41
PCA	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 17, 18, 22, 27, 28, 29, 31, 32, 35, 37
IG	1, 6, 12, 15, 16, 17, 18, 19, 31, 32, 37
RS	3, 6, 12, 23, 25, 26, 29, 30, 33, 34, 35, 36, 37, 38, 39

The suggested approach was executed for each set of features, then the generated detectors were tested against the Train Set and the Test Set to investigate the detection accuracy and which set of features gave the best results. The values used for the algorithm parameters are shown in table III.

TABLE III
PARAMETERS VALUES

Population size	200, 400, 600
Number of generations	200, 500, 1000, 2000
Mutation rate	2/L, where L is the number of features
Crossover rate	1.0
p	0.5

A mutation rate is a measure of the likeness that random elements of your chromosome (individual) will be mutated, which is dependent on the number of features in our experiment. A crossover rate defines the percentage of chromosomes used in reproduction, in this case all selected chromosomes are used. The p norm set for the Minkowski distance measurement is selected as a small value between 0.0 and 1.0 to detect similarity more than difference.

V. Results and Discussion

The algorithm was applied twice for each features group, with the two previously mentioned distance measurements.

Then, the resulted groups of detectors were tested against the whole Train Set and the Test Set. The performance measurements used are accuracy, sensitivity, and specificity, and they are calculated as:

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \tag{9}$$

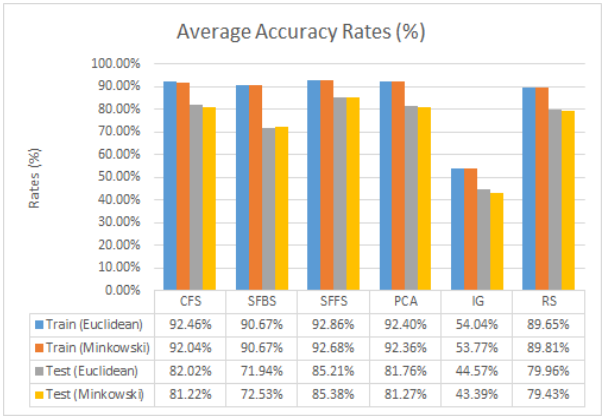
$$Sensitivity = \frac{TP}{TP+FN} \tag{10}$$

$$Specificity = \frac{TN}{TN+FP} \tag{11}$$

Where:

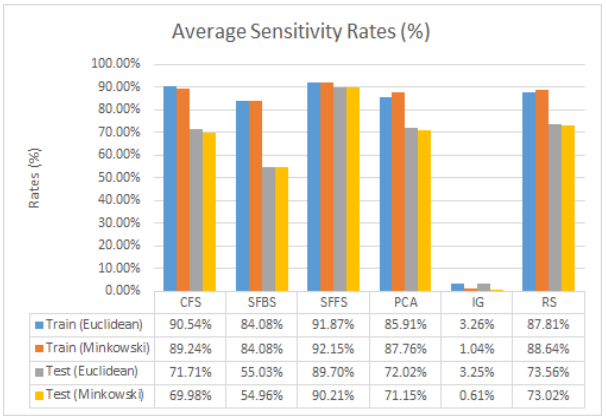
- TP is the True Positives, when an attack is detected successfully and raises an alarm.
- TN is the True Negatives, when a normal connection does not raise an alarm.
- FP is the False Positives, when a normal connection is wrongfully detected as an attack and raises an alarm (false alarm).
- FN is the False Negatives, when an attack is not detected and does not raise an alarm.

Accuracy mean how much accurate the system is to define anomalous and normal activities. Sensitivity expresses the ability of an IDS to correctly classify a connection as an attack. Specificity expresses the ability of an IDS to correctly classify a connection as normal. The average and maximum rates are shown in figures 10, 11, and 12 respectively.



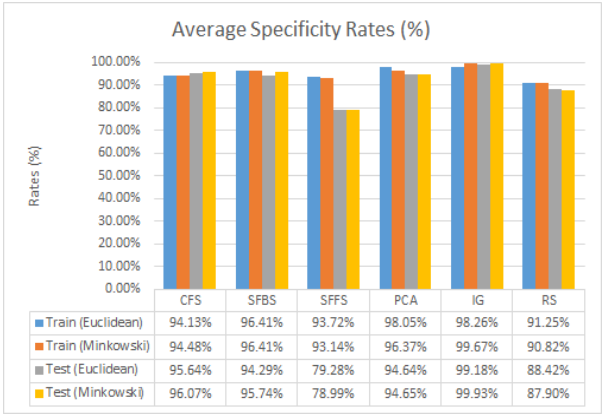
	CFS	SFBS	SFFS	PCA	IG	RS
Train (Euclidean)	92.46%	90.67%	92.86%	92.40%	54.04%	89.65%
Train (Minkowski)	92.04%	90.67%	92.68%	92.36%	53.77%	89.81%
Test (Euclidean)	82.02%	71.94%	85.21%	81.76%	44.57%	79.96%
Test (Minkowski)	81.22%	72.53%	85.38%	81.27%	43.39%	79.43%

Fig. 1. Average Accuracy for Train and Test Sets



	CFS	SFBS	SFFS	PCA	IG	RS
Train (Euclidean)	90.54%	84.08%	91.87%	85.91%	3.26%	87.81%
Train (Minkowski)	89.24%	84.08%	92.15%	87.76%	1.04%	88.64%
Test (Euclidean)	71.71%	55.03%	89.70%	72.02%	3.25%	73.56%
Test (Minkowski)	69.98%	54.96%	90.21%	71.15%	0.61%	73.02%

Fig. 2. Average Sensitivity for Train and Test Sets



	CFS	SFBS	SFFS	PCA	IG	RS
Train (Euclidean)	94.13%	96.41%	93.72%	98.05%	98.26%	91.25%
Train (Minkowski)	94.48%	96.41%	93.14%	96.37%	99.67%	90.82%
Test (Euclidean)	95.64%	94.29%	79.28%	94.64%	99.18%	88.42%
Test (Minkowski)	96.07%	95.74%	78.99%	94.65%	99.93%	87.90%

Fig. 3. Average Specificity for Train and Test Sets

We can realize from the charts that CFS, SFFS, and PCA give the best results in general – figure 10. They give close results for the train set, but SFFS has the best results for the test set. Looking into the details, the SFFS selected features result in the best sensitivity rates (figure 11), which means the detectors can successfully detect anomalies in higher rates (91.87% and 92.15% for the train set, 89.7% and 90.21% for the test set). Although SFFS gives lower specificity rates (figure 12) than other algorithms, but as an overall no other algorithm resulted in such high true positive rates. We can realize too that although IG resulted in the highest specificity rates, it almost totally fails to classify anomalous activities in the data.

VI. Conclusion

In this paper, an anomaly-based, NSA inspired, network intrusion detection system was implemented, where GA was used to generate anomaly detectors for the system. In order to decide which features subset should be used, multiple feature selection approaches were used, and the results were compared to see which algorithm gave the best results. As shown above, although the SFFS gave the biggest features subset, it also gave the best accuracy and the best sensitivity rates - which is

most obvious on the Test set which includes unknown attacks. Checking the features selected by each algorithm, one can realize that:

- Feature 3 — service — is very important in the detection process, it was not selected among the effective features using the IG technique and it shows poor performance in anomalies detection.
- Feature 1 — duration – is also very important, it was selected by the techniques that resulted in better accuracy and detection rates.
- Features that are concerned with the connection rates counts — 22, 29, 30, 31,...,41 — affects the performance too in a better way, they help improve detection accuracy.

In the future, a classification technique should be used to classify the detected anomalies and refine the results more.

REFERENCES

[1] J Bartlett, *Machine Learning for Network Intrusion Detection*, 2009.

[2] Y Singh, P K Bhatia, O Sangwan, *A review of studies on machine learning techniques*, International Journal of Computer Science & Security, Vol. 1(1) , 2007, pp. 70-84.

[3] HG Kayacik, AN Zincir-Heywood, MI Heywood, *Selecting features for intrusion detection: a feature relevance analysis on KDD 99 intrusion detection data sets*, Proceedings of the Third Annual Conference on Privacy, Security and Trust, October 2005.

[4] P Garcia-Teodorro, J Diaz-Verdejo, G Marcia-Fernandez, E Vazquez, *Anomaly-based network intrusion detection: Techniques, systems and challenges*, Computers and Security, Elsevier, 2009, Vol. 28(1-2), pp. 18-28.

[5] A Murali, M Roa, *A survey on intrusion detection approaches*, First International Conference on Information and Communication Technologies, ICICT 2005, IEEE.

[6] P Langley, *Selection of Relevant Features in Machine Learning*, Defense Technical Information Center, 1994, pp. 140-144.

[7] J Hua, WD Tembe, ER Dougherty, *Performance of feature-selection methods in the classification of high-dimension data*, Pattern Recognition, 2009, Vol. 42(3), pp. 409-424.

[8] H Liu, H Motoda, L Yu, *Feature selection with selective sampling*, Machine Learning-International Workshop Then Conference, 2002, pp. 395-402.

[9] L Yu and H Liu, *Feature Selection for High-Dimensional Data – A Fast Correlation-Based Filter Solution*, In Machine Learning-International Workshop Then Conference, 2003, Vol. 20(2), p. 856.

[10] D W Aha and R L Bankert, *A Comparative Evaluation of Sequential Feature Selection Algorithms*, Learning from Data, 1996, pp. 199-206, Springer New York.

[11] R Gutierrez-Osuna, *Pattern analysis for machine olfaction: a review*, Sensors Journal, IEEE Vol. 2(3), 2002, pp. 189-202.

[12] DW Aha and RL Bankert, *A comparative evaluation of sequential feature selection algorithms*, In Learning from Data, 1996, pp. 199-206, Springer New York.

[13] S Aksoy, *Feature Reduction and Selection*, Department of Computer Engineering Bilkent University, 2008, CS 551.

[14] F Song, Z Guo, D Mei, *Feature Selection Using Principal Component Analysis*, System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 International Conference on, Vol. 1, pp. 27-30

[15] M Hazewinkel, *Information* Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4, 2001.

[16] DJC MacKay, *Information theory, inference and learning algorithms*, Cambridge university press, 2003.

[17] D Roobaert, G Karakoulas, NV Chawla. *Information gain, correlation and support vector machines*. Feature Extraction, 2006, pp. 463-470, Springer Berlin Heidelberg.

[18] M Zhang, and JT Yao, *A rough sets based approach to feature selection*, In Fuzzy Information, Processing NAFIPS'04. IEEE Annual Meeting of the, Vol. 1, 2004, pp. 434-439, IEEE.

[19] R Jensen and Q Shen, *Rough set based feature selection: A review*, Rough Computing: Theories, Technologies and Applications, 2007.

[20] W Li, *Using Genetic Algorithm for Network Intrusion Detection*, Proceedings of the United States Department of Energy Cyber Security Grou, Training Conference, 2004, Vol. 8, pp. 24-27.

[21] C Sinclair, L Pierce, S Matzner, *An Application of Machine Learning to Network Intrusion Detection*, In Computer Security Applications Conference, ACSAC'99, Proceedings, 15th Annual, pp. 371-377, IEEE.

[22] U Aickelin, J Greensmith, J Twycross , *Immune system approaches to intrusion detection - a review*, In Artificial Immune Systems, Springer Berlin Heidelberg, 2004, pp. 316-329.

[23] D Dasgupta, *Advances in artificial immune systems*, IEEE Computational Intelligence Magazine, 2006, Vol. 1(4), pp. 409–49.

[24] A S A Aziz, M A Salama, A E Hassanien, SE O Hanafi, Artificial Immune System Inspired Intrusion Detection System Using Genetic Algorithm. Informatica 36 (2012) 347-357

[25] A S A Aziz, A T Azar, A E Hassanien, SE O Hanafi,*Continuous Features Discretizaion for Anomaly Intrusion Detectors Generation*, 2012 Online Conference on Soft Computing in Industrial Applications Anywhere on Earth, 2012.

[26] NSL-KDD Intrusion Detection data set, Available on: http://iscx.ca/NSL-KDD/, March 2009.

[27] KDD Cup99 Intrusion Detection data set, Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, October 2007.

[28] A A Olusola, A S Oladele, D O Abosede, *Analysis of KDD99 Intrusion Detection Dataset for Selection of Relevance Features*, In Proceedings of the World Congress on Engineering and Computer Science, 2010, Vol. 1, pp. 20-22.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 775–780

How to Develop a Biometric System with Claimed Assurance

Andrzej Bialas
Institute of Innovative Technologies EMAG,
ul. Leopolda 31, 40-189 Katowice, Poland
Email: a.bialas@emag.pl

***Abstract*—The article concerns the process of developing biometric devices with a view to submit them for certification in compliance with ISO/IEC 15408 Common Criteria. The author points at the assurance paradigm which shows that the source of assurance is a rigorous process of the product development along with methodical and independent evaluation in an accredited laboratory. The state of the art of certified biometric devices was discussed. There was some focus put on the issue of insufficient support that the developers get in this respect. Basic processes related to the Common Criteria methodology were described (IT security development, IT product development, IT product evaluation). These processes were illustrated by the elements of security specifications of certified biometric devices. The author proposes that development patterns can be used to prepare evidence material, while specialized devices supporting development processes – to deal with basic difficulties encountered by the developers of biometric devices.**

I. Introduction

TODAY'S IT applications, especially those used in the large businesses, banking, e-government and e-health sectors require dependable identification and authentication. One of the possible group of solutions used in these applications is provided by biometrics.

Biometric authentication concerns the automatic identification of humans by their intrinsic physiological characteristics (finger images, hand/facial geometry, vascular patterns, iris, retina, etc.) or behavioural characteristics (hand writing, keystroke dynamics, etc.).

Biometrics can be used for:

- identification of a person's identity; the captured biometric sample is compared with enrolled templates contained in the database to find the matching one;
- verification of a person's identity; the captured biometric sample of the person claiming the given identity is compared with the enrolled template associated with the claimed identity and stored in the database.

Both processes, identification and verification, should be supported by the enrolment process, responsible for capturing biometric samples and storing them in a secure way. Providing mechanisms to associate an identity with a person, biometric devices are often used when quick, secure and positive authentication is needed.

Biometric devices implement the best matching technologies for the given application domains. These devices encompass hardware and software parts. The implementation of these parts is important, as it is always critical for the entire security system in which these devices work.

IT users require trustworthy biometric devices because these devices usually secure their critical applications in high risk environments. The Common Criteria (CC) [1] methodology can be used to develop trustworthy biometric devices.

The developers of biometric devices should be familiar with the Common Criteria methodology because they should be able to perform different CC-related security analyses and tests in order to prepare biometric IT products for evaluation, to elaborate evaluation evidences and to assist the evaluation process. Most of IT developers, not only biometric technology developers, have difficulties to successfully perform these tasks. For this reason some Common Criteria supporting documents and guidances (e.g. [2]) have been elaborated and consulting services are offered. One of the Common Criteria-based methodologies supporting the IT security developers in their works will be presented in this paper. It was elaborated during the CCMODE (Common Criteria compliant, Modular, Open IT security Development Environment) R&D project [3], co-financed by the EU within the European Fund of Regional Development. The objective of this project was to elaborate a CC-compliant methodology and tools to develop and manage development environments of IT security-enhanced products and systems for the purposes of their future certification. The CCMODE project resulted in the following products: knowledge, patterns (including documentation, procedures, evidences, specification means, etc.), methodology and tools which can be used by different organizations to create and manage IT development environments [4]–[5]. The contribution of this paper is to provide developers of biometric devices with the new patterns-based and software-supported assurance methodology to make this development process easier. The paper shows how the general purpose patterns and tools elaborated in the CCMODE project can be adopted for biometric devices. The paper also discusses the state of the art of the certified biometric devices pointing out sources of knowledge useful for developers.

The paper presents a short primer for the CC methodology, a range of the CC-related support offered for biometric technology developers, a review of the development process of biometric devices in the CCMODE development environment, and conclusions.

978-83-60810-53-8/$25.00

II. Common Criteria methodology – a primer

The ISO/IEC 15408 standard Common Criteria [1] assumes that the reliability of security measures depends on how much accuracy and rigour is put into the development, testing, verification, documenting etc. of IT products. The more rigorous is this process, the more precise are the used good engineering practices, the better is the organization of the development /production /maintenance environment – the more reliable, trustworthy is the IT product. In the nomenclature of the standard, the commonly understood reliability was replaced by a more precise term – assurance. The assurance can be measured by means of Evaluation Assurance Levels (EAL) in the range from EAL1 (minimal value) to EAL7 (maximal value). The applied degree of rigour affects the cost of the product development, manufacturing and maintenance, therefore when the EAL is declared, the developer has to compromise between the product costs and the assurance level. In practice, among already evaluated 1,200 IT products, the biggest number are those on levels EAL3 and EAL4 [6]. An IT product in the CC nomenclature is called TOE – Target of Evaluation.

The Common Criteria methodology comprises three basic processes:

- IT security development based on different types of security analyses; a special document is worked out, called Security Target (ST), which is a set of security requirements – functional requirements describing how security measures should work and assurance requirements describing how reliable the developed products are;
- TOE development, including its documentation; this documentation, being an extension to the above mentioned ST, is evidence material prepared for the sake of the third process – security evaluation;
- IT security evaluation, carried out in an independent, accredited laboratory [6].

The standard has a wide application range as it is difficult to find an IT product without any security measures of its functions. Rigorous regulations related to the product development, along with independent evaluation, are the source of assurance for such a product.

Biometric products are security-related products requiring assurance.

III. Common Criteria Support for the Biometric Devices Developers

The developers of biometric devices can use the BSI guide [2] which is about the preparation of evidence material. The guide has a general character (concerns any IT products) and does not give any patterns to prepare the material. Therefore the developers have to use consulting services in this respect. There are few software tools which support the development of evidence material. One of them was described in [7]. The tool allows to generate a Security Target pattern which is one of over a dozen documents needed in the whole process. Some valuable practical hints about the evidence preparation and the certification itself are available in [8].

The developers of biometric devices can get some assistance from the so called Protection Profiles. These are evaluated sets of requirements for a certain class of IT products. For biometric devices only two Protection Profiles have been developed so far.

The [9] profile presents a biometric verification system in terms of [1] and defines functional and assurance requirements for such a system. Two other biometric systems, i.e. enrollment- and identification systems, are not considered in this profile. The profile focuses on the stand-alone version of the biometric device. Moreover, it does not discuss the biometric modality and related hardware. For this reason, the [9] focuses only on a software solution. This PP has EAL2 claimed. Testing is not considered (thresholds). This profile is of basic significance for the developers of biometric devices. The second PP [10] provides fingerprint spoof detection.

Up until now only three biometric devices have successfully passed the certification process [6]. The Security Target [11] presents the functionality of the Palm Secure biometric verification system, based on the structure of the veins in the palm as a unique characteristic of a human body. The Security Target [12] specifies a system that provides fingerprint spoof detection as part of a biometric system for fingerprint recognition. The ST [13] (EAL2+) specifies a distributed (server-based) authentication system based on biometric data.

IV. Development Process of Biometric Devices in the CCMODE Development Environment

In order to obtain a certificate for an IT product, including a biometric product, it is necessary to carry out the three basic processes mentioned in section 2.

4.1 IT security development process

This process encompasses activities aiming at the elaboration of the TOE security functions (TSF) meeting security functional requirements (SFR), to be implemented at the claimed EAL during the next process – TOE development. The IT security development process includes (key parts):

1. Preparation of the ST introduction.

The developer should assign the TOE type (i.e. biometric device) and provide a concise but precise description of the TOE, which can be an entire biometric device or its part only. The TOE can encompass software, hardware or both. It should be described in the ST introduction what the TOE is and what the TOE operational environment is, including the required non-TOE hardware/software/firmware in this environment. In the TOE description physical and logical scope of the TOE should be specified. The ST introduction should present the TOE usage and its major security features.

2. Conformance claims.

They specify conformance with the used CC standard version (e.g. v.3.1), with protection profiles (if applied) and with assurance packages expressing the EAL level.

3. Security problem definition (SPD).

The security problem can be expressed as the assets protection against threats (this method is recommended to apply

more reliable technical measures) or as OSP (Organizational Security Policy) rules to be fulfilled to avoid incidents (organizational measures are less appreciated than the technical ones). A good practice is to start with the identification of the TOE protected assets (they can be inside or outside the biometric TOE) and external entities interacting with the TOE (sometimes called subjects). The biometric TOE protects usually the users' assets placed outside the TOE (e.g. on servers), called primary assets. To protect these assets, it is vital to protect the TOE internal assets, e.g.: biometric reference and life records, claimed identity, configuration data, etc., (sometimes called secondary assets). The external entities can be authorized or not, can be humans or processes. Usually, the main "actors" are: administrator, user, attacker. Specifying threats, OSPs or both, some assumptions for the operational environment concerning connectivity-, personal- or organizational aspects can be added. Examples of threats are: "Using the identity of another user, an attacker may perform a brute force attack to be positively verified by the TOE.", "An attacker modifies biometric references or other security-relevant system configuration data.". An example of OSP is: "The TOE shall meet recognized national and/or international criteria for its security relevant error rates like: False Accept Rate (FAR) and False Rejection Rate (FRR)." More examples are included in [9]. The elementary items of the SPD (as well as SO, TSF) are specified by mnemonic names called generics.

4. Solution of this problem by setting the security objectives (SO) – for the TOE and its operational environment.

The security objectives are concise statements of the intended solution to the given SPD problem (i.e. threat, OSP, assumption solutions). The security problem can be solved partially by the TOE (specifying the TOE security objectives countering threats or enforcing OSPs) and partially by its environment (specifying the security objectives for the operational environment countering threats, enforcing OSPs or satisfying assumptions). The first case expresses the elementary TOE responsibility for security, e.g.: "The TOE shall ensure that all users can be held accountable for their security relevant actions." [9]. The second one expresses the elementary TOE operational environment responsibility for security, e.g.: "The TOE operating equipment and adequate infrastructure shall be available (e.g.: operating system, database, LAN, public telephone, and guardian)." [9]. The developer should provide a rationale that security objectives really solve the problem and are necessary. Security objectives represent an elementary security measure.

5. Working out the security requirements.

The security functional requirements specification (SFRs) is elaborated on the basis of TOE security objectives, while the security assurance requirements specification (SARs) is derived mainly from the declared EAL (please note: EALs are predefined packages of SARs). The SFRs are expressed with the use of the functional components from Part 2 of the standard [1], while the SARs are expressed by the assurance components from Part 3. Both kinds of components are grouped in families and the families – in classes representing ordered security issues. The components can be considered the semiformal specification language of Common Criteria. The informally expressed TOE security objectives are translated to the SFR components and they will be implemented in the TOE security functions. For example, the "FAU_GEN.1 Audit data generation." component presents requirements how the audit records should be created and what they should contain. The security objectives for the TOE operational environment are not translated to the components and will be expressed in technical and operational documentation of the biometric system. The set of SARs implied by the claimed EAL can be modified by adding extra components or replacing components existing in the EAL by more rigorous ones (this is expressed by EAL+). The SARs will determine the range and details of the TOE development and the TOE evaluation processes. The security requirements elaboration is finalized by their rationale. An example of SAR is "ADV_TDS.3 Basic modular design." describing the TOE decomposition into subsystems and modules.

6. Preparation of the TOE summary specification (TSS).

The TSS contains the TOE security functions (TSF) derived from the SFRs, functions which should be implemented in the considered IT product or system during the next step – the TOE development process. The best practice is to group the SFRs around the specific security functionality and assign them to the defined TSF which implements this group. The TOE summary specification provides potential consumers of the TOE with a description how the TOE satisfies all the SFRs (presenting details concerning the SFRs implementation). Examples of TSFs expressed by generics (short mnemonics) [12] are: "TSF_FFD Detecting if a finger presented on the sensor is a fake or not.", the "TSF_AUDIT Producing an audit record for every use of the security functions of the TOE.".

The IT security development process provides a set of TOE security functions, which should be implemented, at the claimed EAL. This process can be facilitated by the use of the ST pattern elaborated in the CCMODE project.

Fig. 1 presents the CCMODE Tools – documents generator window with the security target pattern. On the left side the pattern structure is shown, while the right side presents some fields to be filled in by the IT product related data. Some fields are automatically filled in by data from the project knowledge base. On the bottom part some users functions and knowledge access points are available.

Using patterns the developer focuses on the TOE security issues only, not on composing the evidence documentation compliant with Common Criteria. During this work he/she is guided by the advanced help system.

4.2 TOE development process

The TOE development process encompasses the elaboration of the evidences documentation implied by SAR components of the claimed EAL (please note Table 1 placed on page 31 in the third part of the standard [1]).

The evidence material can have different forms:

- documentation, for example: configuration management plan, manuals for the maintenance personnel or for the administrator, security policy of an institution that develops the product, configuration list, procedure of the system installation, delivery procedure, testing documentation, plan of penetration tests, and many other documents of that type which, with respect to their contents, always resulting from proper SAR requirements;
- documented results of independent research or observations conducted by the evaluators, e.g. a report concerning the analysis of the TOE vulnerability and TOE development environment, report from independent testing of the TOE, report from the inspection of the TOE development environment, or a ranking list of risk cases identified in the development environment;
- behaviour or activities of people who play certain roles in the TOE life cycle, for example the roles resulting from a certain procedure (accepting the product of system before it is delivered to the client, etc.); an example of such evidence can be a protocol, note or the so called records, i.e. traces of different operations (activity reports, logs – either electronic or not) recorded in the management system.
- security target or protection profile.

This process of the TOE development includes:

1. Preparation of the ADV (Development) assurance class evidences (architecture, interfaces, design, implementation).
2. Preparation of the ALC (Life cycle support) assurance class evidences (configuration management, product delivery, development process security, used tools).
3. Working out the test documentation (ATE class), including tests specification, their depth and coverage.
4. Working out the TOE guidance documents (AGD class), i.e. manuals and procedures.
5. Vulnerability analysis support (AVA class).

The result of this process are evaluation evidences for the given IT product and the EAL claimed for it.

4.3 IT security evaluation process

The IT security evaluation is performed by an independent security lab accredited according to the existing national evaluation scheme. The basic tool is the security evaluation methodology CEM [14]. The certificates are published in the Common Criteria portal [6].

IT security development and TOE development processes can be conducted in a traditional way – from the basics with the help of consultants, or they can be carried out on the basis of patterns and supporting tools. The developed evidence material prepared with the use of tools is more coherent – thanks to that there are fewer problems during the evaluation.

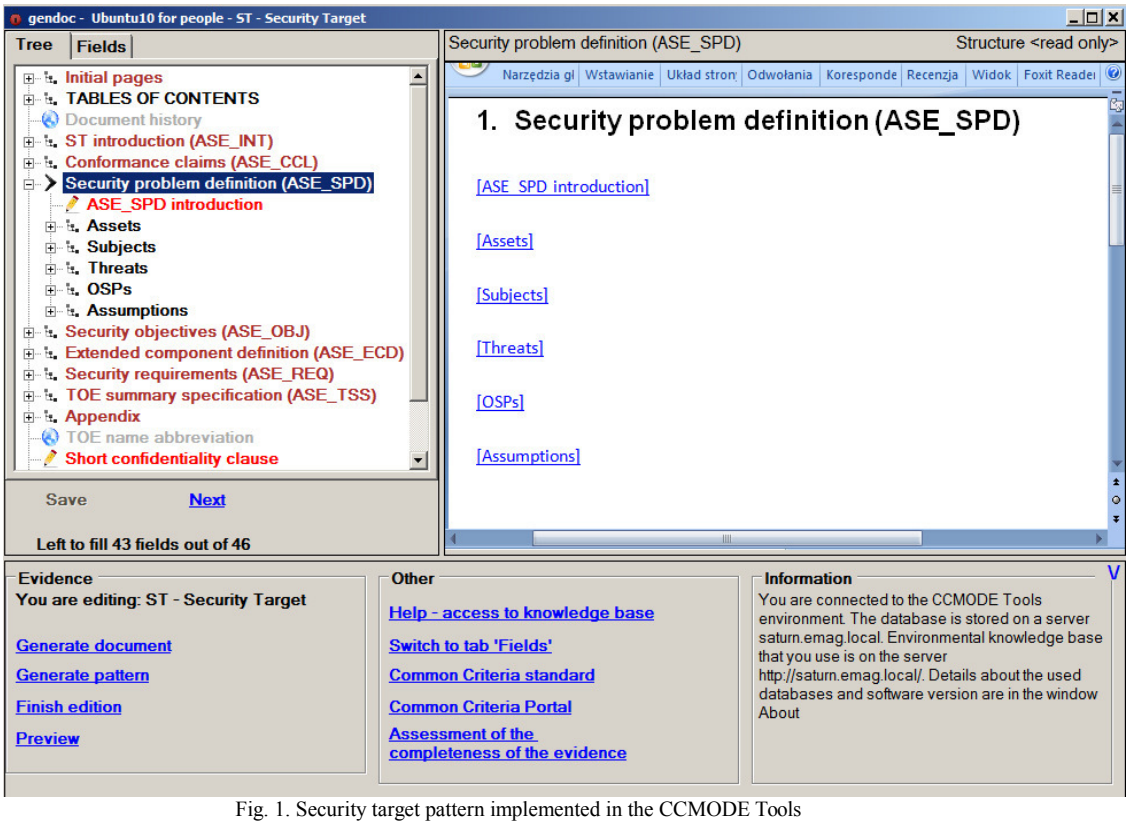


Fig. 1. Security target pattern implemented in the CCMODE Tools

V. Computer aided development of IT products

IT developers, including biometric devices developers, are focused on their products, use technology-specific language and have difficulties to express the results of their work in the Common Criteria specific terms. This standard specifies a set of detailed requirements (SARs) and the developers have troubles how to produce evidences meeting these SARs. They expect assistance by experts or patterns of evidences. Thanks to these patterns they focus only on the product-related issues, not on composing the evidences. Within the CCMODE project such patterns were elaborated as Microsoft Word templates for all assurance components [3]–[4], [15]–[16]. Additional advantages were achieved thanks to the software support of the Common Criteria related processes. The CCMODE Tools for developers [5] was elaborated as a result of the CCMODE project. The Tools encompass:

- project manager module, responsible for initialization of projects and their management in the life-cycle models,
- configuration management module, responsible for the configuration management according to the CC requirements on different EALs,
- Microsoft Word-based GENDOC module designed to work out evidences,
- Sparx System Enterprise Architect (EA)-based module for security analyses and the ST/PP elaboration,
- Subversion (SV)-based module responsible for versioning the project artifacts (including evidences),
- Redmine-based module for TOE design bug tracking and the ALC_FLR implementation,
- Testlink-based module for test development and management (ATE),
- project self-assessment module (CEM compatible),
- auditing module allowing to assess the conformance with different standards,
- knowledge base module for the project management,
- standard-related knowledge.

CCMODE Tools support traditional CC-related projects as well as the site certification concept [17].

VI. Conclusions

The paper presents general guidance for the biometric technology developers with respect to the Common Criteria standard requirements. This standard allows to develop biometric devices with the claimed measurable assurance. The assurance is based on the rigorous methodical development and independent evaluation by an accredited body. The paper provides biometric technology developers with concise information about three basic Common Criteria processes.

The paper draws the readers' attention to certain barriers in the dissemination of certified products, including biometric products. The major barriers are the lack of knowledge and skills among the developers in the use of the Common Criteria standard, high costs of the products development, lack of supporting tools and patterns that would facilitate the use of the CC methodology. The barriers result in the fact that in some IT domains the number of certified products is low. This concerns biometric technologies too.

In the huge number of certified IT products (more than 1,200) only 3 are biometric devices and only 2 protection profiles of biometric products were elaborated and evaluated. The developers point at difficulties in the preparation of evidence material [8]. To help IT developers in this activity, a set of evidence patterns was elaborated and all CC-related development and evaluation processes were computer supported (CCMODE Tools). It is extremely important to have access to knowledge which enables to carry out projects. Therefore the set of tools is supported by an extensive knowledge base.

The CCMODE project focused on the computer support of the CC-related projects management, CC-related security analyses, and pattern-based development of the evaluation evidences. More information about using this tool is placed in [5], [18]. The developers of biometric products who are free from going deep into the nuances of the Common Criteria standard and do not have to prepare the structure and layout of their evidence material from the basics, would be certain to say that their work is easier.

Computer support of the security development process according to the Common Criteria standard is the value provided by the CCMODE project. This is particularly due to the following:

- central management of the project with respect to: roles, development tools (UML, SDK, calibration tools, personalization tools, CAE/CAD, etc.), life cycle models,
- providing CCMODE Tools with the tools to manage the versions and configuration of the product, documentation, faults, tests, security measures of the development environment, and with the tools to conduct analyses, make security models, and carry out audits for compliance and security evaluation,
- providing the developers with proper-structure patterns supported by precise guidelines from the data base about what kind of information should be put in particular fields; these fields are partially filled in automatically with data from the project knowledge base.

These activities are undertaken to facilitiate the developers' work, lower the cost and shorten the time of new products development. This is particularly important in niche-market domains of the standard application, where there are not many products developed. Biometric technology is such a domain.

CCMODE Tools and the accompanying patterns were validated on the basis of several projects concerning software systems and intelligent sensors [4]–[5], [19]–[22]. The paper is an encouragement to take up validation in the field of biometrics. This work should start with the extension of the data base with a subset of generics describing assets, subjects, threats, OSPs, assumptions, security objectives, and

functions that would allow to make and analyze security models of biometric devices.

References

[1] Common Criteria for IT security evaluation, part 1-3. v. 3.1. 2009.

[2] Guidelines for Developer Documentation according to Common Criteria Version 3.1, Bundesamt für Sicherheit in der Informationstechnik, 2007.

[3] CCMODE (Common Criteria compliant, Modular, Open IT security Development Environment) Project. http://www.commoncriteria.pl/ (Accessed 18 May 2013).

[4] Białas A. (Ed.), "Zastosowanie wzorców projektowych w konstruowaniu zabezpieczeń informatycznych zgodnych ze standardem Common Criteria", Wydawnictwo Instytutu Technik Innowacyjnych EMAG, UE POIG 1.3.1, Katowice 2011 r. (in Polish).

[5] Białas A. (Ed.), "Komputerowe wspomaganie procesu rozwoju produktów informatycznych o podwyższonych wymaganiach bezpieczeństwa", Wydawnictwo Instytutu Technik Innowacyjnych EMAG, UE POIG 1.3.1, Katowice 2012 r. (in Polish).

[6] Common Criteria Portal, http://www.commoncriteriaportal.org/ (Accessed 18 May 2013).

[7] Daisuke Horie, Kenichi Yajima, Noor Azimah, Yuichi Goto, and Jingde Cheng, "GEST: A Generator of ISO/IEC 15408 Security Target Templates", In R. Lee, G. Hu, H. Miao (Eds.): Computer and Information Science 2009, SCI 208, Springer-Verlag Berlin Heidelberg 2009, http://link.springer.com/chapter/10.1007%2F978-3-642-01209-9_14#page-1 (Accessed 18 May 2013), pp. 149–158.

[8] Higaki W.H.: "Successful Common Criteria Evaluation. A Practical Guide for Vendors". Copyright 2010 by Wesley Hisao Higaki, Lexington, KY 2011.

[9] Biometric Verification Mechanisms Protection Profile, BVMPP v1.3, Bundesamt für Sicherheit in der Informationstechnik, Bonn 2008.

[10] Fingerprint Spoof Detection Protection Profile based on Organisational Security Policies, FSDPP_OSP v1.7, Bundesamt für Sicherheit in der Informationstechnik, Bonn 2009.

[11] Security Target for PalmSecure Fujitsu Limited, BSI-DSZ-CC-0511, 2008.

[12] MorphoSmart Optic 301 Public Security Target, Safran Morpho, 2013.

[13] AuthenTest Server, Authenware, 2010 (in Spanish).

[14] CEM v3.1, Common Methodology for Information Technology Security Evaluation – Evaluation Methodology, 2009.

[15] Bialas, A., "Patterns Improving the Common Criteria Compliant IT Security Development Process". In: Zamojski W., Kacprzyk J., Mazurkiewicz J., Sugier J., Walkowiak T. (Eds.): Dependable Computer Systems; Advances in Intelligent and Soft Computing, Vol. 97, 2011, Springer-Verlag: Berlin Heidelberg, pp. 1-16.

[16] Bialas A., "Patterns-based development of IT security evaluation evidences", The 11th Int. Common Criteria Conference, Antalya, 21-23 September 2010 (published in an electronic version), http://www.11iccc.org.tr/presentations.asp (Accessed 10 Feb 2013).

[17] Rogowski D., Nowak P.: "Pattern based support for Site Certification". W. Zamojski et. al. (Eds.): Complex Systems and Dependability, Advances in Intelligent and Soft Computing (AISC) 170, pp. 179-193. Springer-Verlag Berlin Heidelberg 2012.

[18] Rogowski D., "Software Implementation of Common Criteria Related Design Patterns" Proceedings of the 2013 Federated Conference onComputer Science and Information Systems (FedCSIS), pp. 1147–1152, ISBN 978-1-4673-4471-5 (Web), IEEE Catalog Number: CFP1385N-ART (Web).

[19] Białas A., Security-related design patterns for intelligent sensors requiring measurable assurance, *Electrical Review (Przegląd Elektrotechniczny)*, ISSN 0033-2097, vol. 85 (R.85), Number 7/2009, pp. 92-99, Sigma-NOT, Warsaw (2009)

[20] Białas A., Ontological approach to the motion sensor security development, *Electrical Review (Przegląd Elektrotechniczny)*, ISSN 0033-2097, vol. 85 (R.85), Number 11/2009, pp. 36-44, Sigma-NOT, Warsaw (2009)

[21] Bialas, A. Common Criteria Related Security Design Patterns—Validation on the Intelligent Sensor Example Designed for Mine Environment. *Sensors* 2010, 10, 4456-4496, http://www.mdpi.com/1424-8220/10/5/4456

[22] Bialas, A. Common Criteria Related Security Design Patterns for Intelligent Sensors—Knowledge Engineering-Based Implementation. *Sensors* 2011, 11, 8085-8114, http://www.mdpi.com/1424-8220/11/8/8085/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 781–786

Real-Time Carpooling and Ride-Sharing: Position Paper on Design Concepts, Distribution and Cloud Computing Strategies

Dejan Dimitrijević
Faculty of Technical Sciences
Trg Dositeja Obradovića 6, 21000
Novi Sad, Serbia
Email: dimitrijevic@uns.ac.rs

Nemanja Nedić
Faculty of Technical Sciences
Trg Dositeja Obradovića 6, 21000
Novi Sad, Serbia
Email: nedicn@uns.ac.rs

Vladimir Dimitrieski
Faculty of Technical Sciences
Trg Dositeja Obradovića 6, 21000
Novi Sad, Serbia
Email: dimitrieski@uns.ac.rs

Abstract—Many carpool and ride-sharing solutions have been proposed and even developed in the previous decades, but rarely have they been able to attain a global user base, at least not up until recently. That was mostly because many of them were not initially designed as scalable, leaving their users with a sub-par user experiences as their user base grew, and often their mobile or desktop client reach was not ubiquitous enough, leaving them available only to a small portion of mobile client devices and/or desktop browsers. This paper describes the design concepts, distribution and cloud computing strategies the authors feel any future global carpool and ride-sharing solution could follow, making it very scalable and ubiquitous enough to successfully reach and serve a global user base.

I. Introduction

THE carpooling, thus also the ride-sharing industry, has only recently started becoming globally interesting. However, carpooling formally appeared in the US in the mid-1970s, after the 1973 oil crisis [1]. At that time the rising costs of using a personal vehicle for transportation of only one passenger made it prudent to drive more than one passenger, usually co-workers commuting daily to and from the same workplace, splitting transportation costs. However, the reduction of oil and gas costs in the 1980s and the breakdown of a typical 9AM to 5PM workday in the 1990s led to a spiral down trend in carpooling popularity. Federal government in the US tried to counter such a trend by giving incentives to carpooling drivers, growing the number of no-toll carpool lanes–the so called, High Occupancy Vehicle (HOV) lanes–across many highways. Those lanes were also allowing for relief from ever growing traffic jams and gridlocks, as the number of vehicles on the roads was ever increasing, which in 2000 exceeded 740 million globally [2] and was projected to be over 2 billion motorized vehicles by 2030 [3]. The sheer number of vehicles alone will create many well-documented problems for urban areas, such as increased traffic, increased pollution, parking congestion, and the need for expensive infrastructure maintenance. To reduce those and also personal transportation costs why not make a global real-time carpool and ride-sharing solution?

This work was not supported by any organization

A. Problems

As said, the expenses, both environmental and fiscal, of single occupancy vehicles could be reduced by utilizing the empty seats in personal transportation vehicles. Carpooling and ride-sharing target those empty seats: taking additional vehicles off the road reducing traffic and pollution, whilst providing opportunities for social interaction. However, historically carpool scheduling often limited users to consistent schedules and fixed rider groups–carpooling to the same place at the same time with a set person or a group of people. To make that problem worse, the leading problem concerns, given in a 2009 survey about why people don't carpool, were difficulty to organize carpools and inconvenience of organization [4]. We feel both of those can be addressed by employing some novel web technologies and modern day available data stores which hold social and location based individual user's data. Besides having to solve the aforementioned problems for making a carpooling and ride-sharing solution that users will want to use, to make it usable on a global scale the ubiquity problem should also be addressed. By ubiquity we mean the problem of having to make it available across both various mobile and desktop platforms, current and future ones, so our proposed solution also utilizes few other rather novel web technologies.

This paper attempts to propose concepts, distribution and cloud strategies that we feel will bring best value for any future global carpool and ride-sharing solution. The rest of the paper is organized as follows: Section II overviews some related work. Section III gives overall design concepts and our objective. Section IV elaborates on our proof-of-concept prototype system implementation choices, with subsections focusing on several specifics. Section V discusses our future work, plans and intentions and finally Section VI concludes this paper.

II. Related Work

As noted, this section deals with existing carpool related work. Subsection A reviews carpool and ride-sharing related solutions currently available and subsection B surveys some of the literature and papers on the subject.

978-83-60810-53-8/$25.00

A. Current carpool and ride-sharing solutions

Entering carpool and ride sharing search terms in some of the largest mobile app store and internet search engines returns a great deal of mobile apps and internet websites offering either classic or dynamic carpooling and ride-sharing. Classic carpool mobile app or website indicates that its users effectively schedule and advertise their plans for a trip well in advance, effectively via a searchable electronic bulletin board, seeking other users travelling in the same direction at the same time either in part or fully. Although some of those apps and websites, such as carpooling.com [6] and its mobile client apps have their uses and large user bases, the static routing problems they help solve makes those uses fairly limited.

The inconvenience of having to search though large carpools or even smaller but fixed choice driver groups, hoping to amongst them find a pre-scheduled and advertised trip adequately consistent with one owns schedule, makes such apps or websites non-practical for relatively short and near-immediate on-the-go carpool and ride sharing trip plans. It is for that reason that even [6] and its large network of European subsidiary websites, added advanced time constrained search features to “find a lift”, which to a certain extent alleviate some of the inconveniences for their on-the-go passenger users. However, the added hourly time-constrained advanced searching still inconveniences their vehicle driving users to be mindful of their advertised pre-trip given schedules, even though that may not always be objectively possible, due to unforeseen events such as: road accidents, gridlocks, etc.

Thus, a new form of dynamic carpool and ride-sharing mobile apps and websites is emerging, indicated by their use of real-time passenger requests along with real-time vehicle driving users’ location data, foregoing the need for well in advance pre-scheduled and advertised trips. Amongst some of the most known and pioneering mobile apps and websites offering dynamic carpooling and ride-sharing are Lyft [21] and SideCar [22], screenshots of their mobile apps are given in Fig. 1.

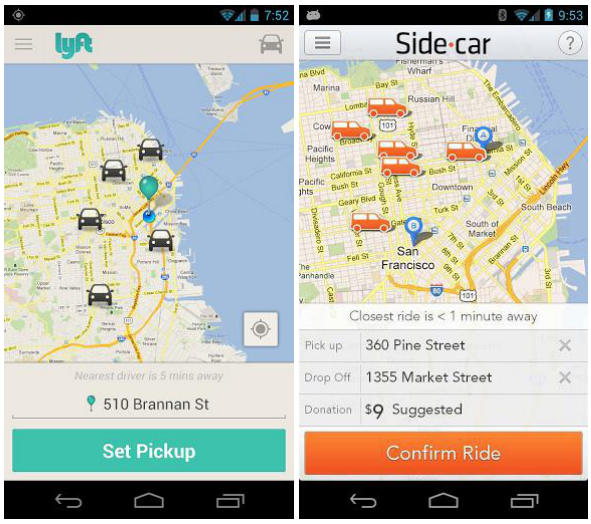


Fig. 1 - Screenshots of mobile applications from lyft.me and side.cr

Both of the listed mobile apps are available for iOS as well as for Android mobile platform, from which screenshots are taken, but neither currently present a web browser user interface. This is probably intentional since both mobile apps are natively written, and by our observations both use TCP sockets to communicate with their respective backend services, so they would both need some changes to make them web-friendly. Unfortunately, those code changes could include some rather tedious transformations because native TCP socket traffic has not been well suited for consumption in web browsers relying dominantly on HTTP, until recently. Another popular application and website named Waze [7], which isn’t predominantly used for carpool and ride-sharing, but for gridlock traffic reporting and avoidance, seems to have however taken another approach. Although their mobile apps are currently natively written, the website does present a "live map" user interface which maps events reported by their users. Such events include pickup requests and replies of passenger and vehicle driving Waze users who are otherwise linked in a popular social network. Unfortunately, access to those real-time events is currently only provided to its iOS app users and not through backend GeoRSS feed via HTTPS. The original GeoRSS XML format is transformed to JSON for easier web browser JavaScript consumption, and presumably for traffic overhead reduction, but is still limited in update interval time. To our knowledge, there are no other globally popular websites and mobile apps that currently allow for carpool and ride-sharing uses using any other drastically different approaches.

B. Current carpool and ride-sharing papers

Because static variety carpool still represents the majority of existing solutions, almost all of the available papers and literature on carpool and ride-sharing mainly tackle the static ridesharing issues, whereby users must pre-schedule their trips, neglecting the dynamic aspect. Despite much of the progress experimented on dynamic carpooling and ride-sharing concepts thanks to the current solutions, it still remains in the early stages regarding publicly available works and literature that deal with its real-time automation. In order to make up for that shortfall, some of the papers which mention carpooling and ride-sharing, and even some that did considered the dynamic aspect [8], in majority also considered other issues beside the static and dynamic carpooling and ride-sharing problems at the same time. Some papers are especially involved in the concepts of traceability, communication and security services, which their authors feel that none of the current solutions evoked, identifying the security issues as one of the main reasons hindering their success [9]. All cited current papers admittedly still provided us with a lot of beneficial ideas and food for thought transferred onto this paper, its findings and conclusions, and out of that still quite disorganized literature which tackles a lot of issues, we have identified some yet non-tackled, laid out in the following sections. Mainly, we take issue with web browser user interfaces and standardized web technologies which seem to be the unifying way forward, putting ubiquity in the grasp of every hybrid web and mobile application.

III. DESIGN

In the previous section we cited some of the current solutions, ideas and issues tackled in carpooling and ride-sharing recently. In this section we are building up on those solutions and ideas, proposing some of our own design concepts for a global dynamic real-time carpooling and ride-sharing solution. Subsection A describes some of design concepts we feel are suitable for a real-time dynamic carpooling and ride-sharing solution. Subsection B further extends on A, allowing for the proposed real-time solution to tackle the problem of being able to serve up to a global user base, adding cloud and distribution design concepts. Finally, subsection C tries to deal with the ubiquity problem, considering the client user interface technology we feel will be future-proof and available on almost all new mobile and desktop platforms.

A. Real-time dynamic solution design concepts

As it was noted in section II, real-time dynamic carpooling and ride-sharing solutions are becoming more common amongst the current carpooling and ride-sharing solutions, although it takes more designing effort to achieve real-time dynamic capabilities than for mere static carpooling and ride-sharing. The reason for the recent increase is obviously because real-time dynamic solutions are more convenient, and thus more likely to be used in greater numbers by end users, but also because some technologies previously used for seemingly real-time communication on the web, have only recently matured and have been standardized.

In the begging of the so called Web 2.0, at the time when real-time updating websites were only just starting to appear, most of those websites used Asynchronous JavaScript and XML (AJAX) [10], which is a group of interrelated web development techniques used on the client-side to create asynchronous, seemingly real-time web applications. Most of those techniques relied upon regular HTTP, a simple request-response and stateless protocol. Having to achieve what was usually two-way communication took some effort for websites and web applications, using various workarounds, techniques involving the use of the browser XmlHttpRequest object or some other web browser plugins.

The first workarounds developed into techniques known as: frequent polling, long-polling and the so called forever-frames. Although all of those techniques were, and still are, very much usable for seemingly real-time web page updates without requiring full page refreshes, they had drawbacks. Their primary drawback was, notwithstanding client-side implementation difficulties, the amount of server-side and network resources they consume. The server is either forced to respond to a large number of frequent requests, or it opens up a number of long running responses, which additionally occupy its hardware resources. On the other hand, using workarounds such as various browser plugins, although less network and server-side resource demanding, turned out to be non-practical, because of the lack of plugin support on current mobile devices. For such reasons, new techniques were developed, and recently standardized by the W3C. As part of the HTML5 specification Server-Sent DOM Events (SSE) were standardized in 2011 [11], but have not yet been implemented by all desktop browsers, namely, Internet Explorer. However, Web Sockets API [12], drafted a protocol back in 2009 currently supported by all major web browsers. Web Sockets provide a full-duplex communication channel over a single TCP connection, thus allowing for a lower network latency time due to less traffic overhead compared to HTTP. Compared to SSE and other polling techniques Web Sockets provide the best option for building real-time communication on the web, and that is why such a protocol is part of our proposed design concept.

B. Distribution and cloud design concepts

Having chosen Web Sockets (WS) as a preferred means of communication, although helping solve latency issues which can lead to a great number of performance problems in building real-time solutions, left another issue unsolved. WS based communication, as all others, still has a limit on the maximum number of simultaneous clients connected to a single server node. Even though that number may be greater when using WS, it still depends on available server hardware resources. Since vertical scaling of server hardware resources can be expensive and still limiting, the solution to the problem is horizontal distribution, across multiple server nodes. Ideally, any global real-time solution would be best served in one's own server farm, but given hardware and its maintenance costs, renting cloud resources works as well. However, for horizontal scaling, one needs to be able to scale data also. Since traditional, i.e. relational data scaling is much harder [13], we have turned to non-relational data (NoSQL). NoSQL databases besides easier scaling, offer better performance in data writes, as well as a possibility of scaling reads onto multiple database nodes, combining sharding and some parallelism approaches. Utilizing the two in a document oriented NoSQL data store that supports geospatial data indexing would make it a perfect fit for our proposed solution and storing our users' location based data. Also, a key-value memory caching NoSQL data store could be used as a messaging backplane for communication between our individual server nodes, but that use is trivial.

C. User interface architecture design concepts

To make a website or mobile app truly ubiquitous, one needs to support as many different desktop and mobile platforms as possible, ideally all of them. Although client applications can be natively written for each platform, there are some unifying user interface technologies for almost all current desktop and smartphone mobile platforms today.

So, to achieve ubiquity, we propose the use of combined HTML5/CSS3 for user interface (UI) rendering. Building the UI around streamed real-time data flows of state changes created by passenger and vehicle driving user events (users requesting rides, user driving busy, user driving free, etc.) is why the paradigm of reactive programming seems to be the perfect choice. Reactive programming is not to be confused with responsive web design, which is also utilized, for same UI reuse across various device screen resolution sizes.

Any changes in state registered by user client applications are asynchronously transferred via WS to distributed load balanced cloud server nodes which are displayed in Fig. 2.

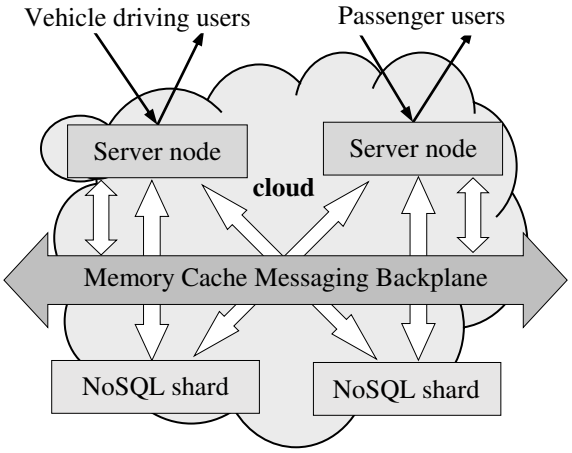


Fig. 2 - Basic design diagram

IV. Prototype Implementation

This section describes in more detail some of the implementation choices we used to build the prototype of our distributable cloud-based dynamic real-time carpooling and ride-sharing solution. Subsection A describes our prototype's real-time communication transport library choice. Subsection B deals with our use of NoSQL data stores for geospatial indexed data and fast memory caching messaging backplane implementation, along with our choices of NoSQL technology and products used. Subsection C goes into some of our UI implementation details.

A. Real-time communication

Having helped develop the first online taxi dispatching solution in Serbia, realized in .NET and being cloud-hosted on Microsoft Azure [14], influenced a lot of our primary technology choices for the prototype of a real-time dynamic carpooling and ride-sharing solution to be described here.

As noted in the design section, the need to have a real-time dynamic carpooling and ride-sharing solution is imperative, since those solutions are what most users currently wish to use. To make our prototype solution real-time capable, the choice to implement it using a library capable of WS protocol communication in .NET came down to a library named SignalR [15]. SignalR is an open-source library for ASP.NET to add real-time web functionality to .NET applications, adding the ability for server-side code to push content to the connected clients as it happens, in real-time. SignalR server is capable of supporting clients written in .NET, JavaScript and some other programming languages. The server-side code can push content to those connected clients via a number of transport techniques, most suitable being bi-directional WS, if available. For a server-side the WS transport requirements are either a self-hosted ASP.NET 4.0 application or one hosted within Internet Information Services (IIS) 8. Server-side hosted on earlier versions of IIS fallbacks to other means of message transports. For clients, WS requirement issue is a bit lengthier to describe, so it is listed in better detail in the client UI subsection.

B. NoSQL implementation

Since IIS 8 was our prototype's hosting platform of choice, it should be noted that the server node could then only have been hosted within the Windows 8 / Server 2012 OS platforms. Fortunately, Windows Server 2012 was made available to end-users on the Windows Azure cloud platform as a virtual machine operating system choice since late 2012, and it is deployable onto an Extra Small machine instance. Extra Small Windows Azure instance, which entails a shared core processor with only 768MB RAM, may not be the first choice from a performance standpoint. However, it is sufficient for proof-of-concept deployments and for limiting cloud-hosting costs.

Deploying SignalR server-side code alongside a NoSQL key-value memory cache data store named Redis [16], with the minimum amount of RAM allocated to Redis, produced a fully functioning server node capable of serving a test number of simultaneous users. Since a new server node can be cloned, and any cloned node's Redis instance can then be easily subscribed to a Redis instance of an existing node, we can easily increase the number of new server nodes to meet all of our scaling needs. Scale out is so easily achieved in part due to SignalR's in-built scaling mechanisms, which uses Redis pub/sub features for a messaging backplane. Each SignalR server node could then, through its Redis instance, be notified of any new WS communication channel needed in real-time. The load balancer of connected computing cloud instances, which is built into Windows Azure, takes care of diverting traffic to a SignalR server node best able (least busy) to process any incoming new or reoccurring real-time request. But since each node has then been notified, by its Redis instance, that such communication has taken place each channel should be reusable by any other SignalR node, so each node is capable of replying to any real-time request.

Beside Redis NoSQL, our prototype incorporated another NoSQL document-oriented geospatial indexing data store for ease of scaling, named MongoDB [17]. MongoDB allows for rapid storing of user current locations by extensions to the stored location data, incorporating another key value used for sharding of that data across multiple MongoDB store instances. Data sharding approach allows for quicker reads since our shard key represents a geographical area within which our users are seeking or soliciting ride request during their relatively short trips. Thus, sharding increases read performance by reducing the amount of indexed data being queried in each MongoDB instance, because that data gets spread out across multiple data store instances. By adding MongoDB's built-in MapReduce (MR) to our queries the tasks of searching through sharded data execute in parallel.

Also, MongoDB by itself incorporates some replication set mechanisms, giving it a highly-available aspect as well, which bodes well with the SignalR's ability to distribute via Redis as a messaging backplane, reducing single points of failure.

C. User interface

Finally, for ubiquity reasons, our choice of prototype client UI rendering technology incorporated HTML5/CSS3. Having previously built a fully functional HTML5/CSS3 client for a commercial online taxi dispatcher, which was wrapped as an app using a mobile development platform named PhoneGap [18] runnable across various mobile platforms, we felt confident that HTML5/CSS3 was also a right choice. Both desktop and mobile web clients shared the same JavaScript logic codebase which offered unified access to geolocation [19] features of the devices they all ran on, a must for a dynamic carpooling and ride-sharing applications. The look and feel across smaller resolutions changes accordingly, but not drastically, by utilizing responsive CSS3 design incorporated in jQuery mobile [20] as of version 1.3.

All the clients also use a reactive programming paradigm, connecting to the backend via code using the Reactive extensions for JavaScript (RxJS) library [21]. This means the user interface responds asynchronously to user actions and events which they result as, either events which are streamed from the server-side generated by other users and fed via WS to all supporting clients, or user's own events. If, per chance, the mobile device's web browser does not support WS transport, SignalR client in JavaScript will gracefully fall back to other means of seemingly real-time transports, which RxJS will still continue to process as asynchronous events.

Support for WS as a mean of communication transport, depends primarily on a platform web browser's capabilities which is for current desktop and mobile web browsers given in Table I.

TABLE I.
WEB BROWSER SUPPORT FOR WEBSOCKETS

Web browser	**Supported since version**	**Supported**
Internet Explorer	10.0 (fully)	Yes
Firefox	4.0 (partially) 6.0 (fully)	Yes
Chrome	4.0 (partially) 14.0 (fully)	Yes
Safari	5.0 (partially) 6.0 (fully)	Yes
Opera	11.0 (partially) 12.1 (fully)	Yes
iOS Safari	11.0 (partially) 12.1 (fully)	Yes
Opera Mini	-	No
Android Browser	-	No
BlackBerry Browser	7.0 (fully)	Yes
Opera Mobile	11.0 (partially) 12.1 (fully)	Yes
Crome for Android	25.0 (fully)	Yes
Firefox for Android	19.0 (fully)	Yes
Firefox OS Boot2Gecko	1.0.0-prerelease (fully)	Yes
Tizen OS	2.0.0a-emulator (fully)	Yes

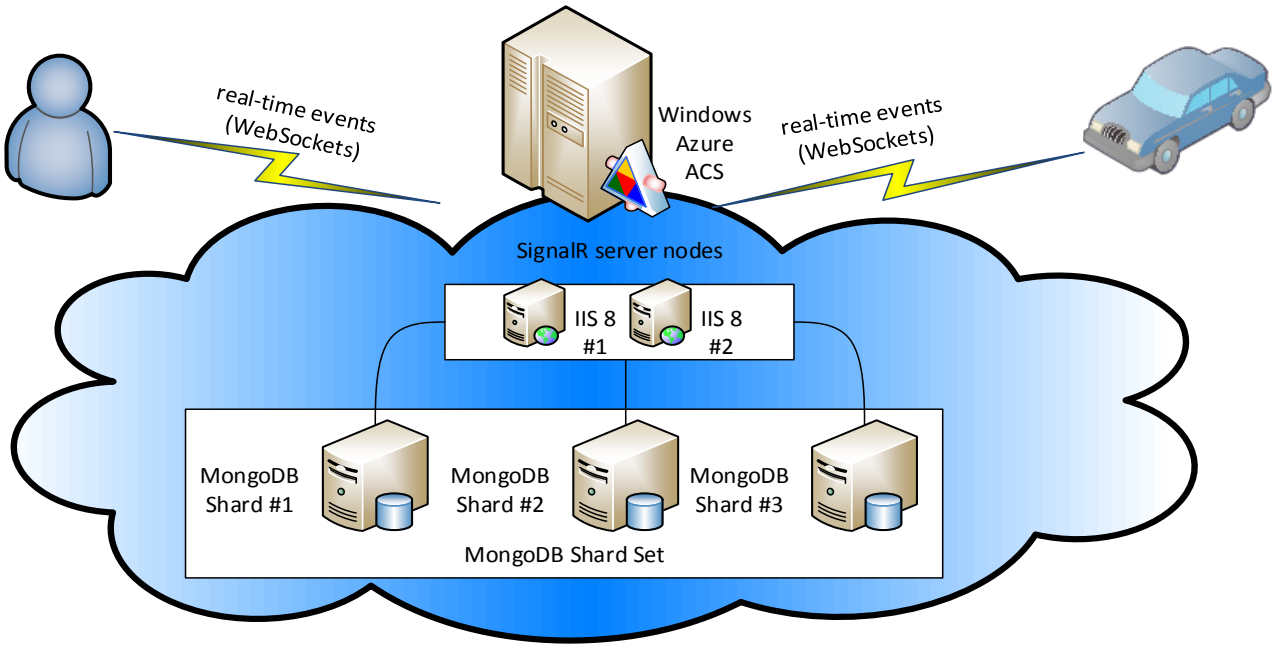


Figure 3 – Windows Azure hosted prototype design diagram displaying some implementation details

V. Future Work

Having described some our prototype's implementation details (Fig. 3) our future work and plans envision for it to be deployed and tested in the real-world. Since the prototype clients were based on previous work done for a commercial online taxi dispatcher, it will be initially tested and deployed as part of that solution in limited numbers. Early adopters of the online taxi dispatching service will get the benefit of being able to track a few assigned taxis in real-time. Drivers of those taxis will be either issued mobile devices with pre-installed HTML5/CSS3 web clients and/or those client apps will be installed on their own devices. Such real-world tests will hopefully lead to identifying problems not yet foreseen. Once a stable solution is reached the prototype could and will become a standalone service, open for public use and not just for taxi dispatching and the cost of its operational maintenance could then also be better estimated. If deemed low enough to be offset by ad support according to [20], its use could be completely free for end users unlike [21, 22].

To reach that point however, some other issues, such as security and privacy, will also need to be tackled. In [5] the solution for the security and privacy issue was implied by use of a 3rd party location based service (LBS), which used OAuth protocol to authenticate and subsequently authorize which exact set of users would be allowed access to the authorizing user's location. Unfortunately, come February 2013 the 3rd party LBS was shut down, and an alternative solution should either be found or developed prior to prototype's launch as a standalone service. Trying to avoid the repeat of having to find alternatives to a 3rd party components not being operational any more, the focus could be on building up own LBS features respective of privacy, but relying on information which can be provided from popular social networks. To aid us in that endeavor, instrumental part of the puzzle could be Windows Azure built-in Access Control Service (ACS), allowing for users to single sign-on to the proposed carpool and ride-sharing service just as if they were signing into the aforementioned social networks. If those users comply, their location data could then only be made accessible to a subset of their social network friends, a widely acceptable solution from a current privacy standpoint.

VI. Conclusion

This paper tried to underscore the need for developing dynamic real-time carpool and ride-sharing solutions, instead of already outdated static ones, by employing some novel web technologies and approaches. Since a prototype has been successfully developed following the outlined design concepts, distribution and cloud strategies, it is obviously possible to build other such solutions using the same approaches. Especially interesting is the possibility to develop a web platform application that runs across multiple devices and their web browsers, be they mobile or desktop. Using an open-source jQuery mobile library and Apache Cordova [23] mobile developer platform, which was derived from PhoneGap, is what interests us the most and we feel could be the unifying tools for any future service supposedly usable across multiple operating systems, current and future. Combining those with some other frameworks which use the HTML5 UI elements such as the canvas tag thus adding the ability to render graphical data such as street level maps for carpool should, by our position, be the leading way forward.[i]

VII. References

[1] Ozanne, L., & Mollenkopf, D. (1999). "Understanding consumer intentions to carpool: a test of alternative models." In Proceedings of the 1999 annual meeting of the Australian & New Zealand Marketing Academy. smib.vuw.ac.nz (Vol. 8081).

[2] Fraichard, T. (2005). "Cybercar: l'alternative à la voiture particulière." Navigation (Paris), 53(1), 53-74.

[3] Dargay, J., & Hanly, M. (2007). "Volatility of car ownership, commuting mode and time in the UK." Transportation Research Part A: Policy and Practice, 41(10), 934-948.

[4] Massaro, Dominic W., et al. (2009) "CARPOOLNOW: Just-in-time carpooling without elaborate preplanning." the 5th International Conference on Web Information Systems and Technologies. Lisbon, Portugal. 2009.

[5] Dimitrijević, D., & Luković, I., & Dimitrieski, V., & Vasiljević, I. (2013) "Orchestrating Yahoo! FireEagle location based service for carpooling" 3rd International Conference on Information Society Technology and Management, Kopaonik, Serbia, 2013.

[6] The largest car sharing network for cheap, green travel in Europe. Web - carpooling.com

[7] Outsmarting traffic, together. Web - waze.com

[8] Sghaier, M., Zgaya, H., Hammadi, S., & Tahon, C. (2011). A Distributed Optimized Approach based on the Multi Agent Concept for the Implementation of a Real Time Carpooling Service with an Optimization Aspect on Siblings. International Journal of Engineering (IJE), 5(2), 217.

[9] Sghaier, M., Zgaya, H., Hammadi, S., & Tahon, C. (2010, September). A distributed dijkstra's algorithm for the implementation of a Real Time Carpooling Service with an optimized aspect on siblings. In Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on (pp. 795-800). IEEE.

[10] Garrett, J. J. (2005). Ajax: A new approach to web applications.

[11] Hickson, I. Server-Sent Events, W3C Working Draft 20 October 2011.

[12] Hickson, I. (2010). The Web Sockets API, W3C Working Draft 29 October 2009.

[13] Cattell, R. (2011). Scalable SQL and NoSQL data stores. ACM SIGMOD Record, 39(4), 12-27.

[14] Najbrži put do slobodnog vozila. Web - taxiproxy.com

[15] ASP.NET SignalR : Incredibly simple real-time web for .NET. Web – signalr.net

[16] Redis. Web – redis.io

[17] MongoDB. Web – mongodb.org

[18] PhoneGap. Web – phonegap.com

[19] Popescu, A. (2010). Geolocation api specification. World Wide Web Consortium, Candidate Recommendation CR-geolocation-API-20100907.

[20] Goldstein, D. G., McAfee, R. P., & Suri, S. (2013, May). The cost of annoying ads. In Proceedings of the 22nd international conference on World Wide Web (pp. 459-470). International World Wide Web Conferences Steering Committee.

[21] Lyft. Web – lyft.me

[22] SideCar. Web – side.cr

[23] Apache Codrova. Web – cordova.apache.org

[i] This is to acknowledge all of the not so small help of our PhD advisors.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 787–793

Emerging technologies for interactive TV

Marek Dąbrowski
Orange Labs Poland
ul. Obrzeżna 7
02-691 Warsaw, Poland
+48 22 699 5706
marek.dabrowski@orange.com

***Abstract*—Advances in web services and open network interfaces enable development of new user-oriented applications in different areas of digital life. Among them, digital entertainment is one of fastest growing networked application domains. Technologies like CDNs (Content Delivery Networks) and HTTP streaming opened the way for new models of TV and video consumption over the Internet. Thanks to emergence of new networking technologies and cooperation paradigms the distinction between traditional TV and the Web is slowly blurring, creating new category of TV entertainment which is more interactive, delivered over multiple screens and inviting users to participate and collaborate.**

This paper presents research on selected social and interactive TV services, focusing on innovations related with incorporation of network interfaces and web protocols. Two use-case scenarios are investigated: one of them explores telco 2.0 interfaces for interactive videos, and the other allows users for social exchanges while watching live sports events. The evaluation results by qualitative user study are presented for the second use-case. A technical architecture is introduced, focused on seamless integration into a typical WebTV platform of telco operator. As important contribution of the paper, some new emerging network protocols and features (like telco 2.0, WebRTC, WebSockets, HTML5 video) are evaluated as potential enabling technologies for future social and interactive TV.

I. Introduction

With the proliferation of IPTV and WebTV content delivery, the TV viewing experience is changing from predominant "lean-back" (passive) viewing habit into the new era of "lean-forward" behavior. People demand more interaction, participation, and social features in addition to just watching the TV. The actors on TV services market (content providers, service aggregators, telco and cable operators) are facing a number of challenges, from the point of view of usage, business and technology, how to integrate the new social and interactive experience into their established service platforms. This paper aims to study selected scenarios for future evolution of social and interactive TV programs, focusing on opportunities brought by new telco network interfaces and web services technologies.

A technical architecture for proposed use-cases is studied from the point of view of telco operator, aiming to integrate new features by leveraging and extending existing TV distribution systems. Special focus is put on OTT (Over-The-Top) content delivery model. The OTT model usually refers to content providers and 3rd party companies who directly access their customers over open Internet. For telecommunication network providers, OTT model starts to be considered as a viable alternative, allowing for addressing additional customers and complementary to the managed IPTV solutions that are still predominant [1]. The OTT content delivery architectures typically use HTTP streaming protocol (e.g. MS Smoothstreaming or Apple HLS) and a CDN (Content Delivery Network) for video distribution over unmanaged Internet with optimal QoE (Quality of Experience).

This paper aims to answer the following research questions:

- From service point of view: what are the realistic and compelling usage scenarios for social and interactive TV?
- From architecture and technology point of view: what are the opportunities brought by new emerging web protocols and network interfaces?

These questions are addressed through analysis of two use-case scenarios, namely:

- Scenario#1: interactive immersive video,
- Scenario#2: social exchanges between spectators of live sports events.

II. Interactive content

The first concept presented in this paper belongs to the category of interactive videos. With interactive video, the movie becomes a kind of immersive game where different story variants may be displayed depending on previous decisions of a viewer. Although interactive videos will probably not enter the mainstream movie industry in the short term at least, they could be appealing for selected types of video production like short artistic movies, instructional scenes, or video games.

The interactive video concept has been known for some years and it has been used for example in DVD interactive games. More recently, it has become quite popular with YouTube service, which offers tools for creating interactive videos. The user may control the video by clicking buttons displayed after the end of particular clip and leading to different variants of the story (see e.g. semi-amateur video [9]). The work presented in article [10] follows the same concept, although removing some of its limitations. The transitions

978-83-60810-53-8/$25.00

between video clips are seamless and more deeply embedded into narrative of the movie. The authors of [10] provide some guidelines for scriptwriting and shooting interactive videos and identify interaction goals for the viewer: to influence the story, to change the perspective, or to see additional material.

The related works mentioned above have a common limitation: the interaction tool and interface for a user is basically a computer mouse. The viewer has to click a button or an area on the screen to make his decision regarding the storyline. This is not very natural, and feeling of immersion is broken at the time of interaction. On the contrary, interesting video referenced in [11] presents an experiment, where mobile phone has been used as interaction tool. A natural conversation has been imitated between the viewer and the character on screen. The outcome of this conversation has been taken into account in the movie storyline. Interrogated user has been selected among the cinema theater audience. A computer system has called his or her mobile phone and conducted a simple dialog using voice recognition technology.

The approach taken in our project is similar to [11], but not limited to the audience of cinema theater. We propose integration of interactive video system into the WebTV content distribution architecture. Any client located in the Internet could watch and control the interactive video. Like in [11] we use the mobile phone (which is a natural tool for inter-personal communication) to imitate a conversation between the viewer and character on the screen, preserving the feeling of immersion into the fictional story.

A. Usage scenario

Fig 1. exhibits our Interactive Video prototype in use. We have produced a short example movie, not targeted for wide distribution, just for demonstrating the potential of proposed system. The movie protagonist consults the viewer before taking crucial decisions for the storyline. For that, in certain moment the system calls the viewer's mobile phone number, previously entered on a web page. The viewer can hear character's voice on his mobile phone (e.g. a phrase "*tell me if I should take the elevator or use the stairs*"). Then he may answer "*yes*" or "*no*" and in the next video scene the protagonist follows the viewer's advice. All actions are synchronized between the video and real life, i.e. the viewer's phone rings at the exact moment when movie character makes a call; the protagonist starts speaking when the user picks up the call, and during conversation the same voice can be heard on the phone and in the video. The movie character interrogates viewer for his advice in several crucial moments in the story, and the movie has three different endings depending on user's answers.

Thanks to our system the user can feel an illusion of real communication with fictional character while watching the video and is empowered with possibility to influence the story.

B. Architecture and enabling Technologies

A prototype has been developed to demonstrate service concept and serve as playground for evaluation of some new

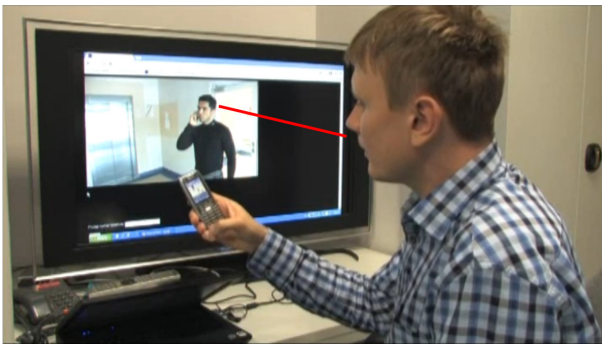


Fig. 1. Interactive video prototype

interesting technologies that may impact future interactive TV. The prototype schema is presented in Fig 2.

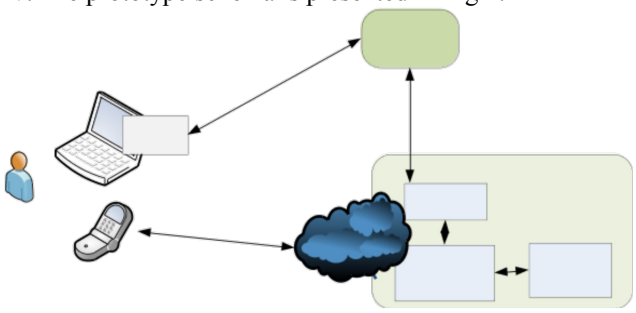


Fig. 2. Architecture of interactive video system

WebSocket and HTML5 video

As design goal, the Interactive Video service should be available for any Internet user equipped with PC computer with a standard web browser, without need of additional software or plugins. Thus, in our architecture the server side hosts logic of interactive movie and controls the sequence of videos that should be played to particular user, depending on his individual interactions with the system. So, the server must be able to send some kind of commands to user's video player (which is embedded in his web browser). This command has to be sent in asynchronous way, i.e. without preceding request from the client. Asynchronous communication is not a usual mode in the client-server world of Internet. There are some techniques to overcome this limitation, e.g. by frequent polling of server by the client, but they can be very inefficient in terms of traffic and processing overhead if one wishes to achieve a timely response. In addition, popular technologies for implementing video player plugins, e.g. Adobe Flash or MS Silverlight, do not allow easily (or not at all) for server-side control of the video player. The player can be controlled from the user-side by mouse clicking on displayed buttons, but the external program has no possibility to change the video that is played, in arbitrary moment.

HTML5 framework [7] comprises some prospecting technologies for solving the limitations mentioned above. The HTML5 video tag enables playing the video natively by a browser. In addition, the video player can be externally controlled by JavaScript code, which in fact facilitates implementing interactive videos over the web. Another HTML5

feature we have investigated is WebSocket [12], an asynchronous protocol allowing the server for sending messages to the client in arbitrary moment. To establish a WebSocket connection, client browser sends a WebSocket handshake request (special HTTP GET message) to the server. Once the WebSocket connection is established, data frames can be sent back and forth between the client and the server in full-duplex mode. Thanks to HTML5 and WebSocket we could implement interactive video scenario as a standard web application, not requiring installing any additional software on the client machine.

For watching an interactive movie, user opens a web page hosted on Interaction Controller server (see Fig. 2). A form for entering a phone number is displayed, as well as "Start" button for launching the video. When a user clicks it, the browser opens WebSocket connection with the Interaction Controller server. A timer process is started for the new user, to control video playback. The timer process will trigger a sequence of user interactions (phone calls at specific times) and video clips, according to pre-programmed scenario.

To play certain video, the Interaction Controller has to send a command to the client browser to open appropriate video file. For that, the server uses WebSocket connection opened at the beginning of the process. The command is interpreted by JavaScript function run by the browser, and it tells browser to stop playing current video and start a new one, available under specified link. The video file is retrieved and played either until it ends, or until next command from Interaction Controller server is received.

Telco2.0 API

The server triggers interaction with a user by invoking advanced Telco 2.0 API functions. Telco 2.0 is a concept of exposing telecommunication services (SMS, call control, location, etc.) through open APIs to external developers. HTTP REST or SOAP is usually used as invocation protocol. Telco 2.0 interfaces are implemented by many network operators and API standardization efforts are recently taken by GSMA organization [13].

Advanced voice API service that we have used allows for establishing a voice call between user's phone and IVR (Interactive Voice Response) machine. IVR is a telecommunication system which controls automatic dialogs with the user. It is typically used by customer relation systems, e.g. when you call a hotline and hear a recorded message *"press 1 to reach customer service, or 2 to report technical problem…"*. In our case, the IVR system has been programmed to play an audio file with recorded voice of the movie character. A user may answer to question asked by the character with almost natural language (in practice, for the purpose of the prototype just simple phrases were used like *"yes/no"*). The ASR (Automatic Speech Recognition) system is then engaged for voice analysis and matching the answer with pre-configured phrases. Finally, the Interaction Controller server is notified about the user's answer to the interaction question. The server, knowing the user's response, chooses appropriate variant of the story and accordingly commands playback of a video file using HTML5 features as described previously. The interaction process is executed in real-time, i.e. there is no perceived delay between the decision expressed in phone conversation, and effect observed in the video.

Evaluation

The concept has not been tested by real users, as we recognized that user's perception and perceived usage benefit depends a lot on a story that is being told by the interactive video. As we only had possibility to produce a simple amateur video as a proof of concept, the user tests would be too much biased by assessing the video (narrative and artistic quality) instead of potential offered by new technology. However, internal tests done with the prototype gave us some interesting insights on the technical and usage aspects.

On the technical side, new web-oriented technologies and protocols like HTML5, WebSockets, Telco2.0 API, have proved their merits for supporting interactive video scenarios. The proposed server-based architecture correctly manages execution of interactive video and allows for providing it as a service to WebTV users. The voice API has been confirmed as powerful tool for carrying out automated voice-based dialogs with users.

Some usage tests have been performed, however only by project members and thus the trustfulness of results is limited. Anyway, based on the feedback of this limited experiments we can confirm that the interactivity is fluid and switching between consecutive videos is almost seamless. Feeling of immersion in the story is genuine and experiment participants could admit that conversation with fictional character was really funny and enjoyable for them. As a drawback we can mention that the automated dialog and voice recognition technology is still not perfect when user says a phrase or sentence that is not understandable or unexpected. As a hint for future work, such situation should be covered in the interactive video scenario, e.g. the movie character should keep on asking viewer for repeating his answer in the case it had not been properly recognized by the system.

III. Social Interactions Around Sports Events

Social TV is recently an often discussed trend in media industry. In fact, this term covers a wide range of scenarios and applications. As a general consensus, Social TV refers to services where TV watching experience is augmented with some forms of inter-personal and inter-audience communication.

The scenarios where content (i.e. a TV show or a movie) is a trigger for social communication, have been studied in numerous related works. For example, paper [2] distinguishes two types of communication around content: synchronous (while watching TV, e.g. through text or voice chat) and asynchronous (after watching, e.g. by leaving notes or recommendations for people who will watch the same content at another time). The authors of [2] have analyzed impact of TV genres (like movies, news, sports, soap operas) on social activities of people. Apparently, sports is one of the genres that is the most "talked about" while

watching, i.e. is a good candidate for social TV system supporting communication in synchronous mode. Intuitively, we can expect that watching sports triggers intensive emotions, and it is natural for us to share these emotions with others, especially friends and relatives. In addition, contrary to films or documentaries, sports do not have strict plot-oriented structure and do not require full viewer's attention all the time. There are usually some breaks and low-activity periods when the action on screen is not so absorbing and the social exchanges at that time would not interrupt watching experience so much.

The social communication aspect can be realized through (1) existing social networking services like Facebook or Twitter, (2) a dedicated application, or (3) enhanced TV service delivery platform. As an example of the first approach, paper [3] presents exhaustive report on how Twitter has been used by producers of popular TV reality show X-Factor for enabling inter-audience communication during and after the show. Interaction with the audience by interactive communication channels and 2nd screen has become an intrinsic part of TV show format. The example of second approach is GetGlue [4], a popular application for sharing and exchanging comments while watching TV shows.

In our approach, the communication features are even more closely integrated into the TV viewing experience thanks to implementing them as part of live TV content distribution architecture. We propose a Social TV system that is specially designed for sports programs and adapted for using during live TV transmissions. Our solution enables social exchanges among friends and fan communities.

A. Usage scenario

In our scenario, social communication around sports events has two "dimensions". The first one concerns exchange of messages, photos and user-generated videos between people who are physically present on the tribunes (e.g. of a football stadium), and those who stay at home and watch live transmission on their TV set or PC computer. Using special mobile application, spectators at the stadium may share emotions and comments with viewers at home. After launching the application, user may select the "event" he attends to. The "event" is configured by service provider and it corresponds to particular match or concert. Then, the application role is twofold. First, it allows for browsing and watching the messages, photos and videos sent by other spectators of the same event. Each message may be commented (by adding text) and ranked by clicking on an icon of a thumb. Second, most important, the application allows for quick upload of your own content after clinking on "send message", "send photo" , or "send video" icons.

Uploaded messages are displayed in real-time on top of the video watched by users of WebTV service. Special video player has been developed using MS Silverlight framework. It displays incoming messages as a scrolled list on the right side of the screen, with preview of picture or video content appended to the message. Viewer may enlarge each message to see its rank and associated comments. He may also add his comment and rank the message in the same way as users of mobile application. Thanks to our system people at home may have a glimpse of stadium atmosphere, which is shaped by fans supporting their players. Example screenshot of WebTV player, displaying a gallery of photos provided by fans at the stadium, is presented in Fig 3.

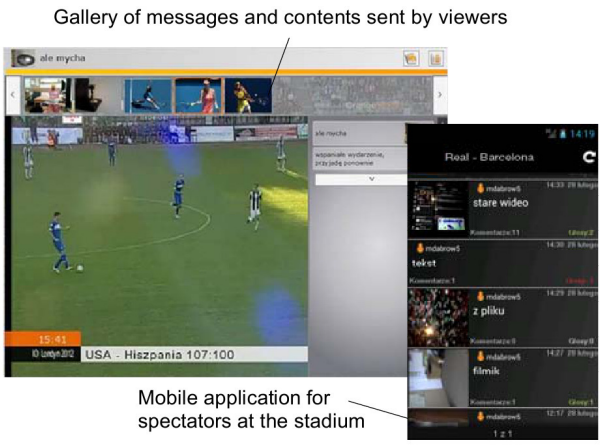


Fig. 3. Sharing photos/videos between spectators at the stadium and viewers at home

Remark that our service concept is a bit similar to FanFeeds, described in article [5]. FanFeeds is a system for aggregation and presentation of additional content-related media on 2nd screen while watching the program. This supplemental information can be contributed by people (from our social group, or strangers) who are watching the same content. The authors of [5] present the prototype concept and results of field trial, which thoroughly investigated people's motivations and incentives to communicate over content. The need to exchange comments during watching the content has been confirmed through trial results. Basically, we enable sharing comments between viewers of a live event in similar way as FanFeeds. However, in our approach we specifically focus on connecting two groups of people: spectators physically present at the stadium, and TV viewers at home. In this way, a fan community can be created by spectators of a sports event, regardless if they attend it physically or at home on TV.

Another dimension of social exchanges around sports events, that we consider in our system, is real-time interpersonal communication. With our service, friends or relatives can watch the televised event together despite being apart in their homes. Real-time communication by text chatting while watching TV has been often mentioned in literature and for example article [6] presents an advanced social TV system based on chat. Real-time voice and video communication is less popular, as it is more challenging from the usability and also technical point of view (a user device equipped with camera is certainly required). The experience of videoconference while watching TV may be replicated by using external application, like e.g. Skype, launched on a PC computer. We take a different approach, by integrating videoconference into the TV content delivery system, providing user with a single interface to access TV service together with the social features. PC and camera is currently required to run the service, while in the future an STB

equipped with camera could also be envisioned. Thanks to group management and presence service a user may see that his friend is watching the same sports program, and invite him to watch together. Fig. 4 presents exemplary screen layout with two participants watching together a tennis match.

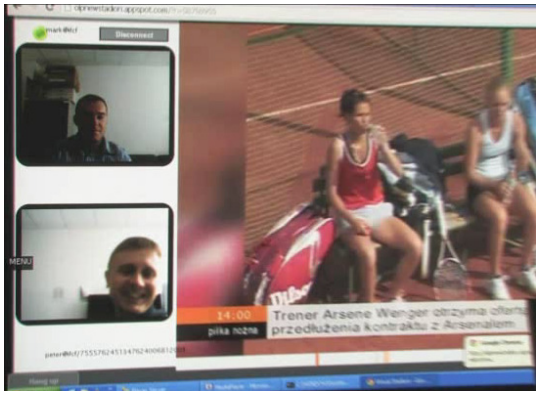


Fig. 4. User-to-user communication while watching TV show

We can say that people who have joined the service are gathered on a "virtual stadium" and constitute a kind of social circle of virtual supporters. They may communicate by voice an they can see each other. In addition, they may take part in quizzes or small games. In our developed prototype, the quiz questions are moderated by service operator and sent to all participants at the same time.

B. Architecture and enabling Technologies

A proof-of-concept prototype has been developed to verify viability of the use-case (concerning both dimensions of communication as described above) and maturity of enabling technologies. It extends typical OTT-based content delivery system with components for managing social exchanges and inter-personal communication. The prototype schema is presented in Fig. 5. On the server-side we distinguish the WebTV platform, which groups content preparation and distribution functions (encoding using H.264 standard, and streaming with adaptive protocol MS Smooth-streaming). The streams are delivered over CDN (Content Delivery Network). The social TV features are provided by two servers. The "Virtual Stadium" server provides clients with a main service web page and manages real-time communication: users identification and presence, as well as groups and videoconference participants matching. The videoconference media is exchanged in peer-to-peer mode between communicating clients by WebRTC technology. The "UGC server" handles functions related with non-real time communication mode. It aggregates messages submitted by user's smartphones (an Android application) and stores associated contents as JPG and H.264/MP4 files. For enabling unified display of UGC messages overlaid on TV content, a special video player has been developed with Microsoft Silverlight.

Although most of the technologies used for the prototype are standard and widely known, the WebRTC merits some attention as it is relatively new and can be considered as very promising for communication services over the Internet.

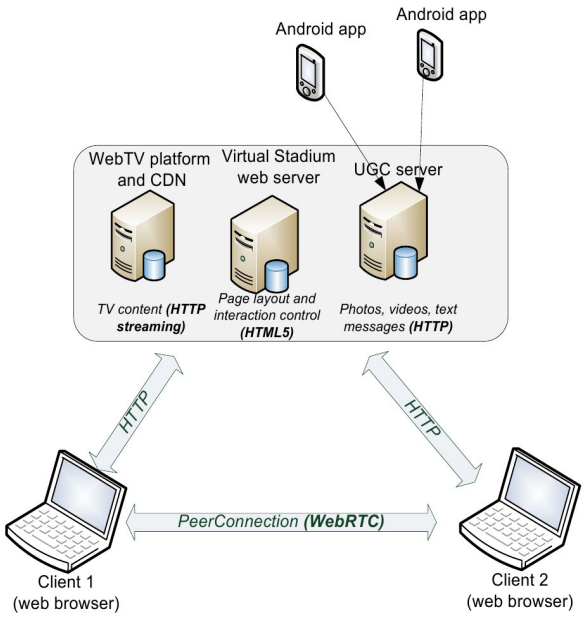


Fig. 5. SocialTV prototype architecture

WebRTC

WebRTC is part of HTML5 framework [7] and it enables embedding real-time communication component within a web page. The communication session is handled directly by web browser, without need of any software plugin. WebRTC [8] defines a standardized browser API that can be invoked using JavaScript by web applications for services such as voice and video calls. WebRTC is subject of ongoing standardization works in IETF and W3C.

The client (see Fig. 5.) opens a "Virtual Stadium" web page on his PC computer. A Silverlight player downloaded from the web page provides access to live TV content through connection with WebTV platform and CDN (Content Delivery Network). When a user decides to connect with his friend, Virtual Stadium server is invoked to generate a session token that is shared by web browsers of communicating parties. A WebRTC peer-to-peer connection is then established between them, for direct exchange of videoconference media (audio and video), all displayed on the same web page as TV stream.

Thanks to WebRTC technology the videoconference service could be seamlessly and easily integrated into the WebTV content distribution architecture.

C. Evaluation

Apart from assessing technical maturity, important goal of research work was to evaluate potential market adoption of the Social TV service. Qualitative tests with real users were conducted for this purpose, using methodology of Focus Groups Interviews (FGI). Four sessions were carried out, each with 6 participants recruited by external agency with regard to typical Social TV user profile that is: age 21-40, people watching matches, concerts, etc. either live or via

web TV, active users of social networking, video-chat applications, internet chat rooms and discussion forums. The goal of the test was to survey reaction of potential users to the Social TV service concept, determine factors which encourage people to use the services and identify any usage barriers. During interviews, service concept and demo videos were shown to participants, followed by a guided discussion on perceived benefits and flaws of the presented concept.

We can conclude from the interviews that users positively perceive the service as meeting their need of sharing comments and emotions while watching live events on TV. The opinions gathered during experiments confirm findings of [2], saying that sport is one of the TV genres that especially trigger social exchanges. Watching live sports is related with strong emotions, which are best experienced in a group. We may quote of the test participants: "*When you're alone, you don't feel the thrill. You have no one to experience it with.*".

Referring to possibility of sharing messages and photos between spectators at the stadium and viewers at home, interviewed persons appreciated the feeling of being a part of live event, thanks to access to additional content sent by participants in real time. Moreover, they liked the possibility of seeing more than is shown by the camera in a typical broadcast: "*At such an event, you cannot be everywhere at once. And we all know that everybody pays attention to slightly different things.*"

The aspect of sharing emotions and experiences has been appreciated by test participants. The impression of watching the event in the company of others, being in connection with the fan community, is important.

The respondents pointed also to some limitations of the concept. A bias has been detected between the benefit perceived by viewer at home, who enjoys additional supplemental content, and benefit of people at the stadium, who may not have sufficient motivation to use our service. We think that possible ways to increase the perceived benefit for users at the stadium are: a competition for spectators (win a prize for best photo from the tribunes), or adding a gamification scheme (collect points and rewards for sharing). Test participants advocated also for possibility of creating user groups and sharing comments within a closed social circle instead of general population of TV viewers. Comments collected during the tests certainly give us interesting hints for future development of the service. A field trial is planned in spring 2013, where the prototype will be evaluated during a real football match event.

As main benefit of Virtual Stadium use-case people mentioned possibility to share opinions and comments with friends immediately during live TV broadcast. The interviewees emphasized the difference between exchanging comments on the event while it was still on, with sharing them after the event is over. The respondents agreed that they saw a greater need to share emotions during the event than after it was over.

Although some concern has been raised that interactivity may disturb perception of the event, participants noticed that while watching sports there are usually some less exciting moments, when conversation with friends would be hardly disturbing for the viewing experience.

Some test participants have previously used other tools for communicating with friends while watching TV: ("*I once watched a match over Skype with a friend who was in South Africa then.*"). Nevertheless, the idea of all-in-one service, integrating several currently used tools has been appreciated as making the experience easier to use. The respondents liked also the fact that it would be available simply in a web browser, without need to install additional software.

Quizzes and games during the match have been judged rather controversial. People find it annoying if the quiz question pops up suddenly on the screen, covering the broadcast. The quiz component has to be designed with great care, being less intrusive and providing sufficient benefit for the users to participate (rewards, competition, possibility to show-off your knowledge in the area, etc.).

Another critical comment concerned current limitation of videoconference to only two participants. Sometimes there is a need to comment or discuss in a larger group. In future works the videoconferencing system should be extended for multiple users.

IV. Conclusion

The paper has discussed several exemplary realistic scenarios for TV services of the future, focusing on opportunities brought by new advances in network interfaces and web protocols. The interactive video use-case assumes imitating a voice dialog between the viewer and fictional character of the movie. Social TV for sport programs allows for two types of inter-audience communication: sharing of comments and photos between spectators at the stadium and viewers at home; and videoconference connection between friends while watching live sports event. The scenario has been evaluated, with positive feedback, by small scale user test.

For all presented use-cases a technical architecture has been designed and prototyped, assuming full integration with OTT-type content delivery architecture of a telco operator. The service logic is implemented by dedicated server components, and standard web browser is used as unified client device for content consumption as well as for social features. In the case of interactive video, mobile phone is employed as second device, using its most natural function, that is a voice call.

Several emerging technologies have been evaluated considering their potential for enabling interactive TV. Among them, HTML5 and WebSocket overcome limitations of traditional web architecture for interactive videos controlled from the server side. Advanced Telco 2.0 interfaces can be used to engage with the viewer through his phone. Mobile applications and open web services interfaces allow for quick and easy upload of user-generated text and contents. Finally, WebRTC is a promising framework for integration of content delivery and inter-personal communication within a web page.

REFERENCES

[1] VideoNet report: The hybrid network operator, March 2013

[2] D. Geerts, P. Cesar, D. Bulterman, The Implications of Program Genres for the Design of Social Television Systems, uxTV 2008, USA

[3] M. Lochrie, P. Coulton, Sharing the viewing experience through Second Screens, EuroITV2012, Berlin, 2012

[4] GetGlue application: http://getglue.com/ (last accessed 10.07.2013)

[5] S. Basapur et al., FANFEEDS: Evaluation of Socially Generated Information Feed on Second Screen as a TV Show Companion, EuroITV2012, Berlin, 2012

[6] F. Martins et al., SentiTVchat: Sensing the Mood of Social-TV Viewers, EuroITV2012, Berlin, 2012

[7] W3C HTML5 specification, http://www.w3.org/TR/html5/

[8] WebRTC initiative web page: http://www.webrtc.org/ , Decemeber 2012

[9] Interactive horror, YouTube video: http://www.youtube.com/watch?v=D2mMIxXGvEE (last accessed on 10.07.2013)

[10] Interactive Movie Demonstration Die Hobrechts, Drei Regeln (Three Rules). In Adjunct proceedings of EuroITV2013. Video available at: http://preview.3regeln.de/player_en.html (last accessed 10.07.2013)

[11] Interactive Horror Movie Last Call, YouTube video: http://www.youtube.com/watch?v=386VGKucWDo (last accessed 10.07.2013)

[12] IETF RFC6455: The WebSocket Protocol, December 2011

[13] http://www.gsma.com/oneapi/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 795–799

Communication in Distributed Database System in the VANET Environment

Ján Janech, Štefan Toth
University of Žilina
Faculty of Management Science and Informatics
Department of Software Technologies
Univerzitná 1, 010 26 Žilina
Email: {jan.janech, stefan.toth}@fri.uniza.sk

***Abstract*—This paper describes principles of the data communication in the distributed database system AD-DB developed by paper authors. The database system is designed to function properly in such a complex and dynamic network as the VANET is. That way, vehicles connected to the VANET could distribute traffic related data to others.**

I. Introduction

VANET (Vehicular Ad-hoc NETwork) is the field of important research nowadays. Many are trying to develop new principles to make it possible to distribute information through this network. Applications for VANET could be divided into two categories: safety applications and comfort applications. Safety applications are more important ones. They are focusing on distributing information about traffic accidents, obstacles and other safety hazards to as many vehicles as possible [13][14].

VANET is defined to be a special case of MANET (Mobile Ad-hoc NETwork) where network nodes are represented by vehicles in a road traffic. But problems with distributing data in VANET are completely different from the MANET ones. MANET nodes as computers with limited power source and limited computing resources have to communicate in small time frames to preserve as much power as possible. All research of the MANET communication is about minimizing communication and computing time and about conserving node power.

On the other hand, almost all of VANET nodes (vehicles in road traffic, road infrastructure) have good power source. So research in this area is focusing on the best way to distribute information for all nodes that are interested in it.

II. State Of the Art

A. Classic Architecture of Distributed Database System

Architecture of DDBS from data organization point of view is shown at the figure 1. It is simple layered model with four layers. Each one of them represents some view on data itself.

There are four layers of distributed database system, each modeling one kind of view on distributed database [1]:

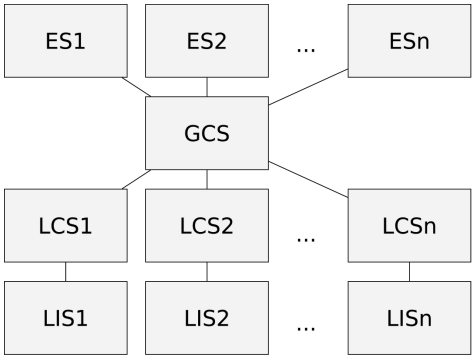


Fig 1: DDBS reference architecture [1]

1. *LIS (Local Internal Schema)* represents physical representation of data stored at one node. It is analogy of internal schema from centralized databases.
2. *LCS (Local Conceptual Schema)* describes logical organization of data at one node. It is used to handle data fragmentation and replication.
3. *GCS (Global Conceptual Schema)* represents logical organization of data in whole distributed database system. This layer is abstracting from the fact that the database system is distributed.
4. *ES (External Schema)* represents user view into distributed database. Each external schema defines which parts of database is user interested in.

The fact that the user is using only global conceptual schema through views defined in external schema, assures that the user can manipulate with the data regardless of its position in the distributed database system. Therefore it is necessary to have a mapping from every local to the global conceptual schema. This mapping, named GD/D (Global Directory/Dictionary), is defined as part of the distributed database system.

The main role of GD/D is to provide access to mapping between local conceptual schemas and the global conceptual schema. So it has to be accessible from every node sending queries to the system. There are several ways to ensure it [4] [5]:

978-83-60810-53-8/$25.00

1. *Centralized directory* – Whole GD/D is stored centrally at one node. The advantage of this solution is that it makes GD/D manipulation simpler. However, one central node represents single point of failure for the whole distributed system and can be bottleneck as well.
2. *Fully redundant directory* – Replication of the whole GD/D is stored on every node. That way it can be quickly accessed whenever needed. But its modifications are more complicated due to its multiple occurrences in the system.
3. *Local directory* – Every node stores only its own part of the GD/D, so its management is very simple. On the other hand, global query requires communication with other nodes to make possible to create the query plan.
4. *Multiple catalog* – In the clustered distributed database system it is possible to assign whole GD/D replication to one node in each cluster. It is combination of first two ways.
5. *Combination of 1. and 3.* – Every node has its own GD/D replication and there is one global replication as well. Each of this possibilities has its pros and cons. But they have all something in common: the system needs to recognize all of its parts.

Whether the GD/D is stored at one node or somewhat distributed through the system, there needs to be some way how to access it as a whole. This is not possible in VANET as there is no way to ensure communication between all of the nodes. In this situation, GD/D cannot be used to locate requested data.

As of the present time there is not solution designed specifically for VANET known to the authors. But there are few solutions for MANET, so we will describe them in next sections of the article.

B. TriM protocol

The TriM protocol is the one of first attempts for solving the problem of data distribution in MANET environment the generic way. It was designed as a part of PhD thesis at the University of Oklahoma [6]. The main focus of the protocol is to minimize power consumption and to utilize all three modes of communication [7]:

- *Data Push* represents data distribution using broadcast messages.
- *Data Pull* represents on demand data distribution.
- *Peer-to-peer communication* for querying data.

The main disadvantage of the TriM protocol is its requirement to have same data on all nodes. This requirement makes it practically unusable in the VANET environment.

C. HDD3M protocol

HDD3M protocol tries to solve TriM protocol problems. As in the original protocol, HDD3M aims to use all three modes of communication and to conserve as much power as possible. The main difference from the TriM protocol is possibility to manage database fragments and to modify distributed database in transactions.

HDD3M divides nodes into 3 categories:

- *Requesting node (RN)* is sending queries to distributed database system.
- *Database node (DBN)* is containing database fragments.
- *Database directory (DD)* stores GD/D for distributed database.

This protocol must solve problems with distribution GD/D. There is no guarantee that all of database directory nodes receive the GD/D update request. Some of the nodes could be inaccessible through MANET or shut down due to lack of energy. When the network is fragmented, keeping the data accurate and actual might be impossible.

The biggest problem for deployment of distributed databases in the VANET environment is the necessity of the knowledge of all the nodes in the system. This problem persists in this solution as well because the GD/D is still used.

III. Principles of Proposed Solution

So the only way to make sure it is possible to use the distributed database system in the VANET environment is to remove the GD/D from the system and replace it with a different principle. As it has been said already, the GD/D describes the mapping between the local and global conceptual schemes. Without the mapping the system does not know where the data are located and how to query them.

Using the GD/D in the VANET environment is impossible because it requires knowledge of the whole system (*global* directory). In a VANET every node knows its immediate surroundings only. So querying of a distributed database is fairly limited in such environment. The only nodes which can be addressed to using queries are those in the immediate surroundings in the network. So the system naturally creates virtual clusters of nodes that can communicate with each other. The clusters might overlap, so each of the nodes of the cluster can communicate with another set of nodes.

The only possibility to introduce principles of distributed database systems into VANET environment lies in allowing to query data only from clusters containing the query node. That way we can replace GD/D with another principle – CD/D (Cluster Directory/Dictionary). But there is still question, how to store CD/D and how to distribute it throughout the database system. The possibilities are same as they were for storing GD/D. They were described in the subsection A. in section I of this paper.

Best possibility for VANET seems to be to store they own part of CD/D at each of network nodes. Other possibilities would be complicated to implement due to highly dynamic nature of VANET.

This is the way distributed database management system AD-DB is working. AD-DB was created as the result of PhD thesis at University of Žilina [2] by one of authors.

IV. Query Processing in AD-DB

As we already said, it is impossible to keep CD/D as a whole and distribute it throughout the VANET. Instead of that, AD-DB is using broadcast messages for data communication and lets each data node to decide whether it has requested data or not by looking to its own part of CD/D.

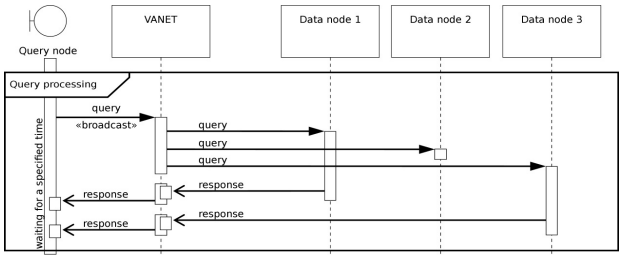


Fig 2: Query processing using the pull method [2][3]

AD-DB supports two methods of communication each based on slightly different principle:

- *Pull method* is application of pull mode of data communication into AD-DB database. It allows each node to query data from cluster.
- *Push method* is application of push mode of data communication into AD-DB database. It allows to share own data to other nodes without any prior query.

A. Pull method

Pull method represents the standard method of query processing in classic distributed database systems. One of nodes sends query to the system and waits for the results.

The method could be used in such situation where a client does not have to update data periodically and it needs to query it once instead. One time search for nearby cinemas could be taken as an example of such a situation.

It is also possible to use the pull method as a mean for data replication but it is much more ineffective than using the push method [15].

The query principle is shown on Fig. 2. Communication is done in the following steps:

1. *Global query optimization*. It is important to optimize a query to minimize the size of resulting data.
2. *Sending a query*. Query node sends optimized query using broadcast message. That way all of the data nodes in cluster receive the query. The query node waits for the specified time.
3. *Query fragmentation*. Every data node which received the query fragments the query and searches for subqueries the node is able to execute.
4. *Local subquery optimization*. Data node optimizes each of found subqueries and prepares it for execution.
5. *Subquery execution*. Data node executes each of subqueries.
6. *Sending the result*. Data node sends back the resulting data together with identification of executed subquery using unicast message.
7. *Results evaluation*. After the specified time runs out, the query node evaluate all results received from data nodes and merges them to one complete result.

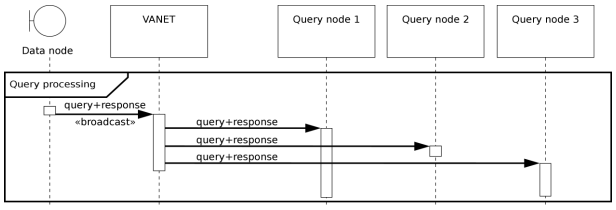


Fig 3: Schematic illustration for push method [3]

B. Push method

Taken that organization of the network structure is changing rapidly in VANET, it is clear that sometimes there is need for querying the same data repeatedly. Possibility of using push method of data communication in AD-DB can be handy in such situation.

That is also the reason why the push method is more effective to be used in data replication algorithms than the pull method [15].

Schematic principle of push method is shown on Fig. 3.Communication is done in the following steps:

1. *Local query optimization*. Data node optimizes query and prepares it for execution.
2. *Query execution*. Data node executes the optimized query.
3. *Sending the data*. Data node sends resulting data packed with the query through VANET as broadcast message.
4. *Results evaluation*. When query node receive the data, it analyzes attached query to determine whether it needs the data or not. If it needs the data, it forwards the data to user application to process it.

V. High Level Communication Protocol for AD-DB

Schematic representation of the communication protocol used in AD-DB is shown on Fig. 4. Data node can process message processQuery. That is sent by a query node in form of broadcast message in the pull method of communication.

The query message has following structure [2]:

- *Schema uuid* is a unique identifier of the current database schema. It is important to include this for data node to be able to determine whether it should process the query or not.
- *Serialized query* represents the query itself. The best way to transfer the query is in a form of serialized abstract syntax tree of it, as it is easy to process by data node.

There is no need to transfer session identifier of any kind, because query and uuid could be used as a unique identifier of the request.

The response message structure is as follows [2]:

- *Schema uuid* as a part of response unique identifier.

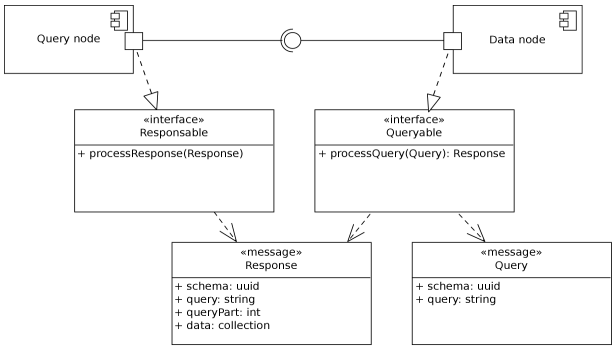


Fig 4: Schematic representation of communication protocol used by AD-DB [2]

- *Serialized query* as a part of response unique identifier. It is possible to use query as a part of unique identifier, because query processed by the database system is expected to be simple and short. If this assumption was not true, it is still possible to use value computed from a query by some hash function instead.
- *Query part* is the identifier of processed subquery.
- *Data* as a collection of the resulting objects.

Using the schema uuid and query pair as a unique identifier of a request has one advantage over using surrogate identification number. This way the response message format can by the same for the pull and the push methods of communication.

Important part of a response message is the query part identification. It represents a unique identifier of query part processed by a data node as a subquery. This identifier is needed by a query node to be able to merge all responses from all responding data nodes.

There are two possibilities how to use the same system of numbering for all query parts by the both query and data node:

- Inserting the identifier directly into serialized query. Process of identifier inserting is done directly by the query node after the global optimization.
  Example of query with identifiers (syntax of the query language used by AD-DB is published in multiple publications by authors [2][3][8]):

  $$^{(1)}\bowtie(^{(2)}projects//(\lambda x|x\triangleleft name{=}\text{KANGO}),^{(3)}employees)$$

  where $^{(1)}$ identifies the whole query as one part, $^{(2)}$ identifies collection of all projects with the name KANGO, and $^{(3)}$ identifies collection of all employees.
- Automatic numbering of all operations by theirs priority. Priority of an operation can not change as it is defined by the query language, so the numbering will be same on both query and data node. This system is preferred and it is used by AD-DB as it does transfer slightly smaller quantity of data between the query and data node.

Push method is using the processResponse message. It is sent by the data node in a form of broadcast message.

VI. The OSACP Protocol

OSACP (Object Structure Aware Communication Protocol) is an application protocol designed specifically to transfer structured data through VANET. It is designed as a part of PhD thesis at University of Žilina [9]. OSACP is using UDP transport protocol on top of IPv6 network protocol. Its design allows it to transfer any structured data through VANET and reconstruct it on the other side even if part of data was not transferred correctly [10][3].

Missing parts of the structure are replaced by special object UNKNOWN to indicate incomplete message. It is up to the user of the distributed database (person, or another application) to decide whether it can process the message or not.

VII. Conclusion

There is no known distributed database systems, that would be possible to operate in the VANET environment. There are few attempts to do so for MANET, but they are unusable for VANET.

The paper presented the communication system of the distributed database system AD-DB. The database system is designed to be used in VANET environment and so its basic principles had to be altered for such usage.

In the nearest future we would like to focus on enhancing query optimization algorithms, but there are many other areas which would be interesting to explore. For example, many of data in VANET are of highly temporal character, e.g. current weather, traffic flow speed, traffic obstacles, etc. It would be interesting to have possibility to query current state of those temporal data.

We have some accomplishments in this area even now. We have designed system to query visual objects recognized by vehicle cameras through VANET [11][12]. So the next logical step would be to integrate this system into distributed database system AD-DB.

Acknowledgment

This contribution/publication is the result of the project implementation:

Centre of excellence for systems and services of intelligent transport II., ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

"Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ"

References

[1] M. T. Ozsu, P. Valduriez, *Principles of Distributed Database Systems*, 3rd ed. 2011.

[2] J. Janech, *Riadenie procesov pri distribúcii databáz (Data distribution process control)*, PhD disertation, Dept. of Software Technologies, University of Žilina, Žilina, Slovakia. 2010

[3] J. Janech, T. Baca, A. Lieskovsky, E. Krsak, K. Matiasko, "Distributed Database Systems And Data Replication Algorithms For Intelligent Transport Systems", *Communications: Scientific Letters of the University of Žilina*, Vol. 15, No. 2. 2013.

[4] P. Sokolovský, J. Pokorný, J. Peterka, *Distribuované databázové systémy (Distributed Database Systems)*, 6th ed. 1992.

[5] C. J. Date, *An Introduction to Database Systems,* 8th ed.. 2003.

[6] L. D. Fife, *TriM: Tri-Modal data communication in mobile ad-hoc network database systems*, Ph.D. disertation, University of Oklahoma. 2005.

[7] L. D. Fife, L. Gruenwald, "Research issues for data communication in mobile ad-hoc network database systems", *SIGMOD Rec.,* Vol. 32, No 2. 2003.

[8] J. Janech, A. Lieskovský, E. Kršák, "Comparation of Strategies for Data Replication in VANET Environment", *26th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. 2012.

[9] T. Bača, *Optimalizácia prenosu správ v ad hoc sieťach (Optimization of Message Distribution in Ad-hoc Networks)*, PhD dissertation, Dept. of Software Technologies, University of Žilina, Žilina, Slovakia. 2012.

[10] T. Bača, "Optimisation of message distribution in Ad-hoc networks", *Information Sciences and Technologies : bulletin of the ACM Slovakia*, Vol. 4, No. 4. 2012.

[11] Š. Toth, J. Janech, E. Kršák, "Query Based Image Processing in the VANET", *5th International Conference on Computational Intelligence, Communication Systems and Networks,* CICSyN2013. 2013.

[12] Š. Toth, *Spracovanie obrazu s využitím dopytov v prostredí VANET (Query Based Image Processing in the VANET)*, PhD dissertation, Dept. of Software Technologies, University of Žilina, Žilina, Slovakia. 2013.

[13] E. Kršák, P. Hrkút, P. Vestenický, "Technical infrastructure for monitoring the transportation of oversized and dangerous goods", *Federated Conference on Computer Science and Information Systems, FedCSIS 2012*. 2012.

[14] S. Badura, A. Lieskovsky, "Intelligent traffic system: Cooperation of MANET and image processing", *1st International Conference on Integrated Intelligent Computing, ICIIC 2010*. 2010.

[15] T. Bača, "Data replication in distributed database systems in VANET environment", *Proceedings of 2011 IEEE 2nd international conference on software engineering and service science*, Beijing, China. 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 801–805

Content Delivery Network Monitoring with Limited Resources

Krzysztof Kaczmarski, Marcin Pilarski
Faculty of Mathematics and Information Science,
Warsaw University of Technology
ul. Koszykowa 75, 00-662 Warszawa, Poland
Email: k.kaczmarski@mini.pw.edu.pl
Email: marcin.pilarski@mini.pw.edu.pl

Bogdan Banasiak, Christophe Kabut
Orange Labs Poland
Telekomunikacja Polska S.A.
ul. Obrzeżna 7, 02-691 Warszawa, Poland
Email: bogdan.banasiak@orange.com
Email: christophe.kabut@orange.com

***Abstract*—This article presents results of designing a Content Delivery Network monitoring system for resource limited applications. CDN monitoring is important both for content providers (media companies) and administrators (Internet Service Providers). It is a challenging task since network traffic may generate huge volume of data which must be parsed and analysed in real-time. This paper describes the design of a prototype system that uses a small resource footprint, scalable Big Data solution, which is motivated by real world use cases.**

I. Introduction

DISTRIBUTED internet systems monitoring is already an important branch of industrial computer systems. Network data transfer, hardware state and many other aspects of distributed systems need to be constantly tracked in order to detect anomalies, prevent failures and measure hardware and software load. In this field Content Delivery Network (CDN) monitoring becomes a fast developing branch of telecommunication. Network caching is used in a variety of different contexts to provide cost savings due to decreased bandwidth consumption, but also to reduce network latency. Due to the techniques described in [1] that makes RAM/HDD resources ratio important especially for use cases in developing regions of the world.

CDN Monitoring requires detailed information, both statistical and sensor-reported to be available real-time and cover any given period of time while not loosing any detail. For example, if a malfunction is detected one may need to track it back behaviour up to the unlimited time point in order to find possible coincidences with other events. Therefore all relevant data concerning CDN operation must be stored in exact shape.

There are several highly efficient systems able to collect and store raw logs. Flume [2] for example is able to collect log lines, send them to a HDFS storage and generate reports on them. Hive [3] can evaluate SQL-like queries over large data volumes collected in Hadoop HDFS [4]. All these classical Big Data technologies require large hardware configurations while storing data inefficiently from the time series point of view. Much better approach is applied in OpenTSDB [5], a dedicated time series database running on Hbase [6]. Its data model can be effectively tuned to achieve both very fast querying and limited hardware. This paper describes optimizations for a CDN monitoring system based on experiences of its deployment in one of the biggest telecommunication companies in Poland which operates commercial CDN services.

A. Limitations of Network Traffic Monitoring

The CDN monitoring becomes especially challenging in developing regions where storage footprint for monitoring is very limited. The situation at those regions is especially unfortunate since the transmission of data logs into public clouds or third party services cannot be performed for the reasons like cost, security and privacy.

This type of monitoring system may be useful in environments where the link bandwidth is a constrained resource e.g. the CDN nodes deployed at libraries, schools, remote communities, etc., basically all locatlizations where uplink bandwidth is limitted.

Also, the security and privacy constraints are general in this scope e.g., most of ISPs are reluctant to deploy subset of CDN systemts called Transparent Caching solutions primarily because of logging concerns in third party services.

The goal of this research was to develop CDN monitoring tool for usage in developing regions with limited resources. We present the system which provides the same level of flexibility of traditional heavy weight monitoring tools.

II. Time Series Database

From a theoretical point of view the system needs to store time series understood as a collection of observations made sequentially in time [7]. These discrete observations T are represented by pairs of a *timestamp* and a *numerical value* (t_i, v_i) with the following assumptions:

- number of data points (timestamps and their values) in one time series is not limited,
- each time series is identified by a name which is often called a *metric name*,
- each time series can be additionally marked with a set of *tags* describing measurement details,
- observations may not be done in constant time intervals,
- storage should not limit time series to be piecewise constant or linear (see Fig.1).

978-83-60810-53-8/$25.00

A. Architecture

The general architecture of the system is composed of three components: data collection, data storage and querying engine.

As in many other systems data are inserted by distributed *data collectors*, which may work directly on data sources and push data into the data storage. Usually data collection is a result of the ongoing measurement or monitoring processes (the operating system, databases and web servers or network devices: see OpenTSDB monitoring tools [8]). The system does not limit possible data which can be inserted and analysed. The only requirement is that it must have a form of *time series*.

According to the existing taxonomies (like in FAME system [9]) measured values are of two types [10]:

- level values stay the same from one period to the next in the absence of activity. For example, inventory is a level value, because inventory stays the same if you neither buy nor sell.
- flow values are zero in the absence of activity. For example, energy expenses go to zero if there is no consumption.

This distinction turns out to be important when interpolating missing values and for time scale conversion. Our system is open for both solutions by inserting zeros when necessary during data collection time.

As we described in [11] our CDN monitoring system is built on OpenTSDB which uses HBase and HDFS as a data storage. OpenTSDB is responsible for storing data in two HBase tables: the first one contains main data compacted into one hour blocks, the second one keeps time series and tag IDs. OpenTSDB is also evaluating queries by getting proper data blocks from HBase and aggregating them into a time series according to the query semantics. Performance of all the system is influenced by all the three components: HDFS, HBase and OpenTSDB.

B. CDN Metrics and Metrics Querying

For the purposes of CDN monitoring the most important metrics from an operational point of view are:

- bandwidth – average number of bits per second transferred from the platform within a given time period [kbps, Mbps, Gbps]
- traffic – sum of bytes transferred in a given time period [kB, MB, GB]
- sessions – number of active sessions in a given time period

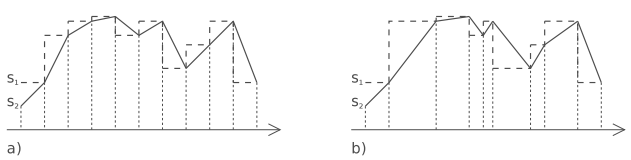


Fig. 1. Time series with constant (a) and variable (b) sampling, piecewise constant (S_1) and piecewise linear (S_2).

- unique clients – number of unique IP addresses
- url hits – number of content download events content from given URL addresses
- byte cache hit ratio – percent of data volume sent from the cache [%]

Additionally, all metrics must be further divided into the following dimensions:

- node name – CDN node name or IP which sends data to a client
- cluster group name – group of nodes logically grouped together
- http response code – HTTP code sent in response to a client request
- country ISO code – client's geographical location
- AS code – Autonomous System code of a client's Internet Service Provider
- CDN instance name – a logical grouping of nodes working in one or multiple CDNs
- provider name – name of a content provider
- origin server name – name of an origin server which contains the original data cached by a CDN
- url – url which is tracked by the system

We use the following metric naming schema:
`[infrastructure].[measurement].[aggregation]`
where:
`infrastructure` indicates a system which is measured (in our case CDN instance, but could also be CPU),
`measurement` indicates a name and type of the measurement being done (for example Mbps, MB, requests, etc.),
`aggregation` describes the type of aggregation done during the data collection which results in one data point for a given time interval.

For example, one of the clients' requests in Poland downloading the content from a hypothetical disco-tv could be stored as (a metric name, timestamp and value followed by a list of tags):

```
cdn.mbps.avg5min 2013-02-20-21:04:20 2.50
node=node01.waw.cdn-lab.pl group=waw02 httpcode=200
country=pl as=AS5617 cdn=lab provider=disco-tv
origin=share.disco.pl url=clip463421-doda
```

which means that within 5 minutes period the collector calculated average value of Mbps to be 2.5 for server named **`node01.waw.cdn-lab.pl`** being in cluster group **`waw02`** for a client located in Poland in **`AS5617`** downloading content from a provider named **`disco-tv`** originating from a server **`share.disco.pl`** with the use of CDN instance named **`lab`** and referencing url **`clip463421-doda`**.

This approach, characterized by defining all possible dimensions for each metric, enables very flexible querying of collected data. For example, one could ask for an average Mbps for given provider in Germany in a given time period or number of error codes returned for a given provider's network (identified by AS number) when accessing certain CDN node by given URL. Please note, that all tag values combination defines one time series. Total number of time series stored for

tag name	min values	real-life
node	10	100
group	5	10
httpcode	5	12
country	1	90
as	1	30
cdn	1	5
provider	1	5
origin	1	20
url	10	50
total time series	2500	$810 \cdot 10^9$

TABLE I
APPROXIMATION OF POSSIBLE NUMBER OF TIME SERIES FOR ONE METRIC IN MINIMAL AND REAL-LIFE SCENARIOS.

one metric is then given by $v_1 \cdot v_2 \cdot \ldots \cdot v_i,\ i = 1 \ldots l$ where l stands for the number of tags attached for a metric and v_i is maximal number of distinct values possible for tag i. Let us analyse how many time series can be generated by an average metric and tags set. An approximation of minimal and real-life scenarios are given in Table I.

The real-life approximation was done using production data from one of the working average CDN systems. It assumes it is possible that accessing all URLs will be tracked with possibly all http response codes and hitting all possible nodes from each country. Since there might be some dependencies between the dimensions the real observed values may be smaller, however, the database must be prepared for the worst scenario which may appear for example during malfunction. Actually, anomalies are the most important from the analytical point of view. Therefore the system should be able to present data including all possible dimensions, and tag values ranges.

C. Data Volumes

Another aspect of CDN monitoring is the estimation of data volumes which must be processed by the system in real-time. CDN traffic monitoring may be based on at least three information sources: CDN proprietary logs, Apache http logs and Syslog events. Due to the latest IETF standarization efforts [12] and [13] we may expect fourth information source of the logs from CDN interconected systems. One request for content generates one line in the log entry (about 0.2 kB). One Smooth Streaming video generates up to 300 entries per 5 minutes (about 60 kB) 10k users watching 90 minutes smooth streaming video generates about 10GB of log data. In some cases log coalescing funciton may be used, however that techniques may reduce size of log by a factor of 10 in average. Another example is 100k users downloading a 1GB game with 5Mbps connection may generate up to 1GB of log data. These data need to be downloaded and analysed and cannot wait for batch processing at night or weekend days. CDN systems offering content worldwide may be equally loaded all the time in which case a monitoring system must collect and process data in real-time.

III. OPTIMIZATIONS

It is a typical HBase performance design pattern to build clusters of at least 11 nodes. In more demanding environments 50 nodes is an average number. This approach would lead to building a monitoring infrastructure more expensive than the monitored CDN itself. Therefore the time series database must be deployed on a single node cluster with a pseudo-distributed configuration on one hand allowing for possible quick extensions in the future and generating sensible running costs on the other hand.

A. Single Node Configuration Performance

Let us now analyse the performance of a single node configuration. We executed two types of queries:

- Aggregating all available time series for a given metric into a one time series in a given time window. For example[1]: `select sum:cdn.bandwidth.15min from 01.01.2012 to 31.12.2013`
- Performing a selection of time series using tags before the aggregation and within a given time window. For example: `select sum:cdn.bandwidth.15min from 01.01.2012 to 31.12.2013 where provider=disco-tv and country=de`

Due to the CDN logs behaviour we used an equal sampling with 5 minutes time period therefore each day is described by $24 \times 12 = 288$ time points. Please note that the time series may not be continuous and therefore their number may vary throughout a queried time window. In Table. II we can see that an average time series processing speed for a query which is scanning all time series in given a time window is about 660k points per second. Since we scan all data, the number of points and time series is constantly increasing. The processing of 4 days data takes almost 40 seconds. Queries for the periods longer than 4 days failed due to an out of memory error. Similar situation appears for the filtered time series querying presented in Table III but we may observe that processing speed is much worse due to more complex data selection from the database. However, the number of processed points is much smaller allowing for better response times. Although it was possible to run queries covering even 16 days, the response time took 108 seconds, which is totally unacceptable from a user's points of view.

B. Reduction of data complexity

One of the conclusions from the previous section is a need to reduce the number of time points and time series in the database. This can be achieved in two ways:

- by aggregating the points in time with downsampling, which can be called *horizontal compaction*
- by aggregating time series with grouping the tags for one metric, which can be called *vertical compaction*

Both solutions lead to a reduction of available information but may be acceptable if consulted with the user queries. For example, in a most typical example a user wants to get brief information about the system and would like to drill down if needed. Therefore both the detailed and coarse information

[1]In this paper we use an abstract query language with obvious semantics to express platform independent queries.

TABLE II
QUERYING AGGREGATING ALL TIME SERIES, DAYS: 1...4.

queried days	proc. time	data points	periods	series	pts/sec.
1	2.9	2,225,293	288	7,726	767,342
2	13.2	9,497,624	576	16,488	719,516
3	28.5	17,137,873	864	19,835	601,328
4	38.1	25,099,032	1,152	21,787	658,767

TABLE III
QUERYING FILTERED TIME SERIES, DAYS: 1...16.

queried days	proc. time	data points	periods	series	pts/sec.
1	6	793,907	288	2,756	132,317
2	8.4	1,295,810	576	2,249	154,263
3	14.1	2,284,509	864	2,644	162,021
4	20.6	3,250,740	1,152	2,821	157,802
16	108.9	12,479,745	4,608	2,708	114,598

should be kept in the system. In case of short time periods a user is interested in a detailed information, for the longer ones the details would not be visible anyway, therefore the coarse plots are just fine.

1) Downsampling Collectors: There are two possible ways to calculate the downsampled time series. The first one, the easiest, is to perform downsampling right during the data collection process. However, this would require to store aggregates for a long time according to the downsampling period (even one week or more). Keeping a collector's state for a longer time may be dangerous in case of a server malfunction. In industrial prototype we prefer solutions which are stateless, run frequently and returning results as quickly as possible. Therefore, we propose another way by implementing additional collectors which process data already stored in the database and send it back into another metrics after aggregation. The basic architecture of this solution is visible in Fig. 2 where the new collector is in gray colour. Obviously there should be one downsampling collector running for each downsampling time period. It is not working continuously but rather started every given time interval. For flexible querying it produces several metrics containing four different aggregations if sensible: *minimum*, *maximum*, *sum* and *average* of values for the sampled time period. Additionally to enable further average calculation it also counts a number of processed points and put the result into another metric called *count*. For some metrics (like Mbps) the summation downsampling does not make any sense and should be omitted.

2) Tags Grouping During Data Collection: During the prototype evaluation phase, we have observed that although a client could run queries containing any combination of tag values he is usually interested only in a subset of two or three tags. The queries concerned: geographical location; traffic per AS, http response code, origin server; and download speed. All time series need to be further divided per content providers. Therefore, instead of one metric with 9 tags, as it was described in II-B, we propose five metrics with six tags. In case of bandwidth metric (cdn.mbps) it could be:

- **`cdn.mbps-country`** with tags:
  `node, group, cdn, provider, origin, country`

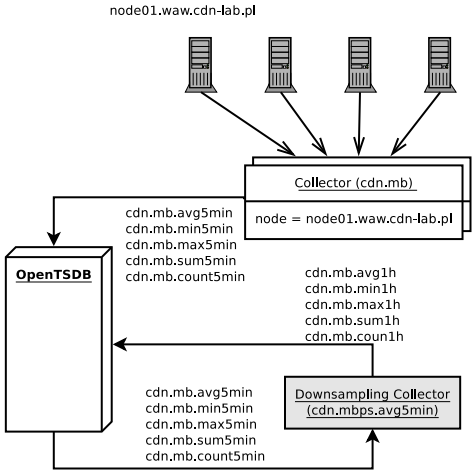


Fig. 2. The basic architecture of a downsampling collector.

- **`cdn.mbps-group`** with tags:
  `node, group, cdn, provider, origin, group`
- **`cdn.mbps-network`** with tags:
  `node, group, cdn, provider, origin, network`
- **`cdn.mbps-speed`** with tags:
  `node, group, cdn, provider, origin, speed-group`
- **`cdn.mbps-httpcode`** with tags:
  `node, group, cdn, provider, origin, httpcode`

Obviously, a user cannot display for example http codes for given AS or country with the above metrics and tags. However, these detailed queries can be processed with the original unoptimized metrics since selecting many tags greatly reduces the number of queried time series. Furthermore the query gets sensible number of data points. The optimized metrics should be used for general queries aggregating many time series. This optimization is a kind of a pre-aggregation done for certain query types.

This vertical aggregation can be run as a post-processor alike in downsampling or during the data collection process. Due to the architecture presented in our previous publication [11] it is much simpler to build it as a MOLAP cube with a reduced number of dimensions.

IV. RESULTS

Adding the vertical compaction optimization by aggregating time series during the data collection time has increased the log processing time for one CDN log covering data transfer for 5 minutes period from a single node by a factor of 3 from about 10 to 30 seconds. The log rotates every 5 minutes (300 seconds) so the time left can be used for processing log files even 10 times larger which will not be the case of a small or medium system.

The horizontal time compaction by the downsampling of the existing data and putting it back into the database as a new time series does not introduce any additional cost in an on-line log processing and does not need to be further studied here.

TABLE IV
SPEED-UP FOR QUERIES ON OPTIMIZED TIME SERIES.

days	1	2	3	4	16
speed-up total	3.2	11	17.8	16.5	
speed-up filtered	12	12	11.75	15.8	13.4

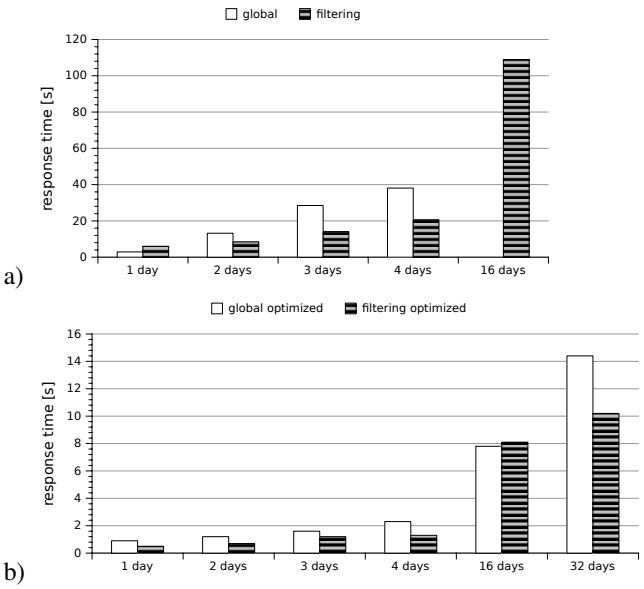


Fig. 3. Response times for initial unoptimized system (a) and final optimized version (b). The charts present two types of queries: global aggregating all time series and filtering performing selection before aggregation. For unoptimized system global query for more than 4 days could not be run due to out of memory error. The queries were run on time series without downsampling.

The reduction of time series and tags describing one metric resulted in an outstanding improvement of the database querying performance (see Table IV and Figure 3). This effect was possible first by the reduction of Hbase tables lookups. When processing queries, OpenTSDB first looks for IDs of the time series in a meta table. The more time series and tags, the more expensive this initial query can be. Then, the second great speed up is achieved by reducing the complexity of the time series needed to be grouped and interpolated in order to get the final time series which answers the query.

The response times for the same queries, but working on remodelled time series, are around 15 times faster for the longer time periods. This optimization decreased the amount of memory necessary to process the time series during the aggregation and therefore allowed for the processing of the times ranges longer than 4 days, which is obviously important for the real-life monitoring platforms.

V. CONCLUSIONS

We presented the problem of a small and medium scale CDN monitoring system in a case of limited resources which does not allow for real Big Data storage cluster. The initial state could not be accepted by the industrial requirements since the short term queries were processed too slowly and longer term queries could not be evaluated at all. Adding more resources by inserting computational nodes or increasing size of memory was not possible due to budget constraints. A significant improvement was achieved by introducing optimizations of the time series stored in the system. All initial properties of the system including the real-time processing and a fine grained data store allowed for detailed queries.

As the next step we plan to study the number of time series reduction allowing for arbitrary queries on an optimized series. This could be achieved by adding more metrics with reduced tag sets.

VI. ACKNOWLEDGMENTS

We appreciate many valuable comments from the anonymous reviewers of Federated Conference on Computer Science and Information Systems 2013. We thank Sara Oueslati, Paris, France for her great support with setting up the logging environment in the ISP CDN networks, and Marc Fiuczynski, New Jersey, USA for additional technical feedback.

REFERENCES

[1] A. Badam, K. Park, V. S. Pai, and L. L. Peterson, "Hashcache: Cache storage for the next billion.," in *NSDI*, pp. 123–136, USENIX Association, 2009.
[2] The Apache Software Foundation, "Apache Flume." http://flume.apache.org, 2013.
[3] The Apache Software Foundation, "Apache Hive." http://hive.apache.org, 2013.
[4] The Apache Software Foundation, "Apache Hadoop." http://hadoop.apache.org, 2013.
[5] B. Sigoure, "OpenTSDB scalable time series database (TSDB)." http://opentsdb.net, 2012. Stumble Upon.
[6] The Apache Software Foundation, "Apache HBase." http://hbase.apache.org, 2013.
[7] C. Chatfield, *The analysis of time series: an introduction*. Florida, US: CRC Press, 6th ed., 2004.
[8] OpenTSDB, "Whats opentsdb." http://opentsdb.net/, 2010-2012.
[9] "MARKETMAP ANALYTIC PLATFORM." http://www.sungard.com/fame, 2013.
[10] D. Shasha, "Time series in finance: the array database approach." http://cs.nyu.edu/shasha/papers/jagtalk.html.
[11] K. Kaczmarski and M. Pilarski, "Content delivery network monitoring," in *FedCSIS* (M. Ganzha, L. A. Maciaszek, and M. Paprzycki, eds.), pp. 633–639, 2012.
[12] L. Peterson, J. Hartman, and M. Pilarski, "A simple approach to cdn interconnection," *Internet-Draft*, no. draft-peterson-cdni-strawman-00, pp. 1–27, 2011.
[13] G. Bertrand, I. Oprescu, F. L. Faucheur, and R. Peterkofsky, "Cdni logging interface," *Internet-Draft*, no. draft-ietf-cdni-logging-04, pp. 1–41, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 807–810

The control on-line over TCP/IP exemplified by communication with automotive network

Elzbieta Grzejszczyk
Warsaw University of Technology Pl. Politechniki
1, 00-661 Warsaw, Poland , Email:
egrzejszczyk@zkue.ime.pw.edu.pl

***Abstract*—One of the more important stages in the development of wireless networks was developing protocols, procedures and systems providing packet data transmission. Packet data transmission enables sending sets of measurement data and other information over long distances, and thus integrating with other available networks, for example the Internet. The aim of this article is to demonstrate the implementation of communication algorithms of the above networks as exemplified by communication with an automobile on-board network. The part of the article which deals specifically with the functionality of the discussed transmission types addresses the issues of remote control as exemplified by low-power executable devices**

I. Introduction

THE first automotive microcontroller built by the Bosch company in 1979 was a 4-bit microprocessor microcontroller implemented in BMW 732 and BMW 633 Csi cars. Its function was solely to control ignition and fuel injection systems. From the time perspective it can be said that it was the first ECU system (Engine Control Unit) in the world –computer controlling and monitoring of engine operation.

Constant development of Information Technology (IT) and microprocessor evolution enabled further development of controlling and monitoring of automotive subsystems by introducing information buses. Fast and error-free data transmission to main controllers is possible thanks to digital communication buses installed in cars which on the basis of this information make further control decisions , (e.g. correcting engine operation, ABS or EPS systems.). At present, the most popular bus used in the car industry is the Control Area Network (CAN). Its standard was made public in 1986 by Bosch and the first CAN controller (made by Intel) was made available in 1987.

The first implementation of the system took place in the mass-produced Model S of Mercedes-Benz.

Continuous efforts of designers to increase passenger safety and comfort have resulted in installation of several buses managing the work of on-board microcontrollers in modern cars, which control both operation and measurement systems. These buses create on-board information networks, such as LIN, K-Line, Flex-Ray or MOST (Fig.1)

Each of the aforementioned networks and buses controls different executable systems both while the vehicle is in motion and when it is stationary – for diagnostic purposes. (Fig.2)

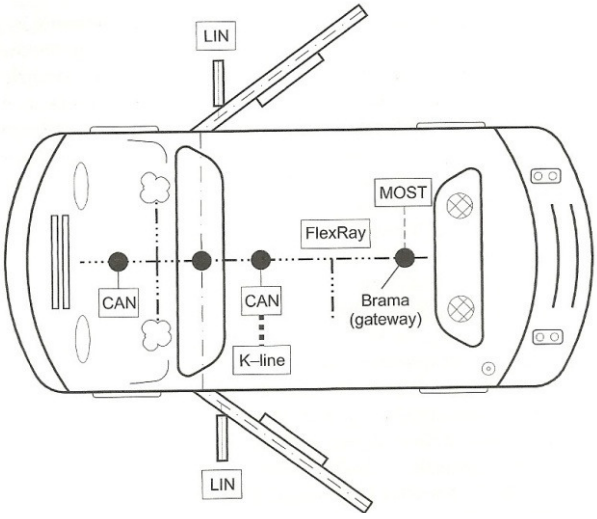


Fig. 1 On-board computer network [1]

II. Remote Control of an Object Over TCP/IP and GPRS Network

Information technology tools available at present, as well as wireless telecommunications networks, allow for easy remote monitoring and controlling of various vehicle functions [4]. Such control is made possible, i.a., by the development of the GPRS network concept, which enabled

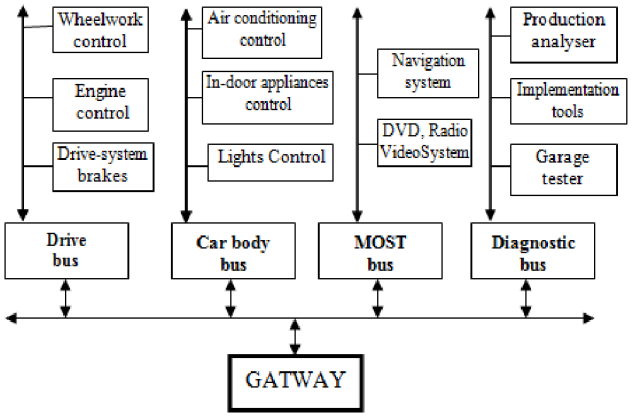
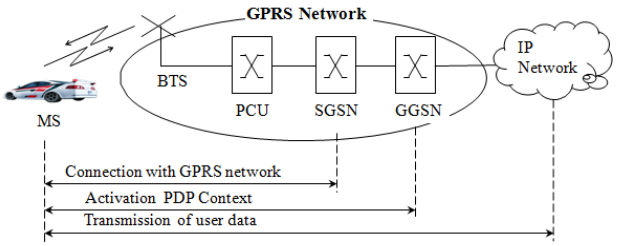


Fig. 2 Digital bus networks controlling run-time systems (according to [2])

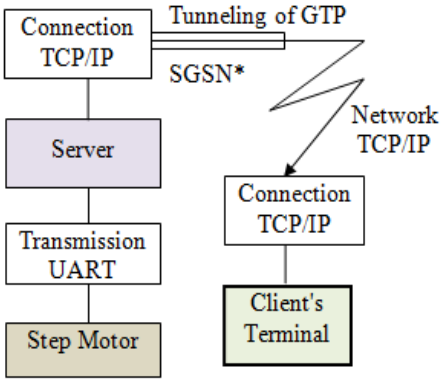
packet data transmission.[3] Figure 3 illustrates establishing connection between a MS (Mobile Station) and the Internet.



MS - Mobile Station, **BTS** - Base Transceiver Station; **PCU** - Packet Control Unit;**SGSN** - Serving GPRS Support Node; **GGSN** - GPRS Gateway Support Node;**GPRS** - General Packet Radio Service

Fig. 3 Packet data transmission over GPRS between a mobile station (MS) and the Internet [1a]

The sequence of establishing a connection involves activating and connecting an MS to the GPRS network and then defining the MS's PDP Context, including the assignment of an IP number. Then packets may be mutually exchanged, which enables the assignation of the sender's and the receiver's addresses to each connected user , , both in the IP and the GPRS network. How the GSM/GPRS network operates is described in [3]. The model of remote controlling of a selected object over TCP/IP is shown in Fig.4.



* SGSN - controller Serving GPRS Support Node, GTP - GPRS Tunnelling Protocol

Fig. 4 Model of remote control over the TCP/IP protocol

The end user (Client) controls the object via the GTP protocol (GPRS Tunneling Protocol) tunelled by TCP/IP protocol. Data exchange between the controlled object and the Central Computer (Server) uses UART.

(Tunneling means establishing a connection between two different protocols by encapsulating one protocol in another. By means of tunneling a connection between distant hosts can be established, giving the impression of direct connection.)

III. AN EXAMPLE OF SELECTED OBJECT CONTROL

Figure 5 illustrates an example of object controlling over TCP/IP. This method of controlling was developed for purposes of laboratory teaching during lectures in Wireless Data Transmission Systems delivered by this paper's author to the Computer Science students.[5]

The object of control is a stepper motor (EDS 10) linked to a computer by means of a microcontroller 8051. The executable system is remotely controlled by over a dozen methods. Parameters of controlling are entered from the keyboard of a remote computer. (Client)

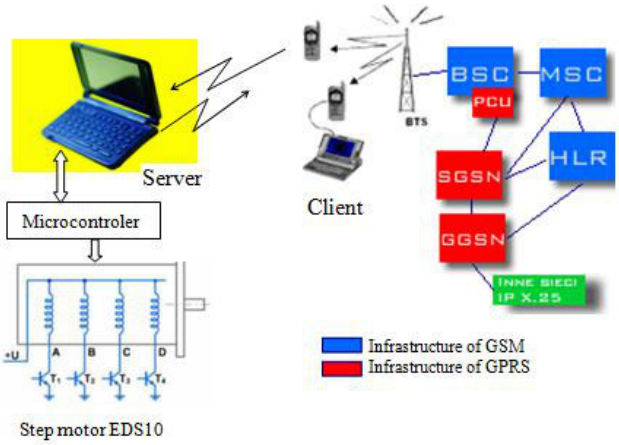


Fig. 5 Diagram of a remotely controlled step motor

Providing software for the above system involved the following stages:

- programming TCP/IP interface between a remote computer, the so-called Client Computer and the Central Control Computer - Server (initialized by the Client)
- designing and programming an interactive user interface, enabling the introduction of selected values of control parameters.
- developing programs in the microcontroller's language for controlling a given object (i.e. the step motor).

The user's (Client's) interface is shown in Fig.6.

The remote control program is run in two modes:

- Client mode and
- Server mode

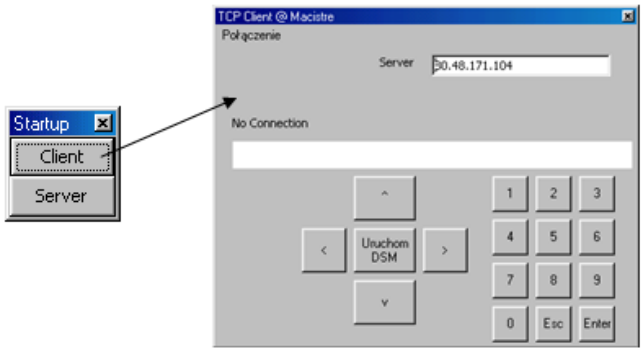


Fig. 6 User interface in the remote control system in Client mode

Client runtime mode

There are two stages in this mode:

- establishing connection with the remote server by entering its IP number in the text field of the form (the top right corner of the interface) and choosing the option "Connect" from the system menu

- handling the interactive form by means of which the user enters control parameters (keys 0-9, direction arrows, key ESC and Enter.)

The system was designed so that messages from the microcontroller 8051 could be displayed on the client's desktop. These messages show both the state of the running control program and the parameters required for control purposes. They are displayed in the white field of the interface. (They also appear on the remote LCD of the microcontroller.)

Server runtime mode.

The Server's aim is to handle communication between the Client and the microcontroller and concerns the transformation of the controller's Assembler language to messages shown on the screen of the Client's remote interface.

Handling the Server requires defining the connection between the Server and the Client as well as the Server and the microcontroller.

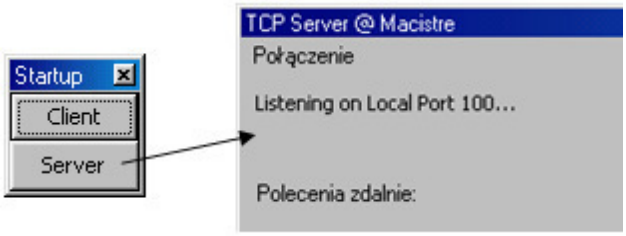


Fig. 7 User interface in the remote control system - Server mode

Client – Server Communication

The above communication was carried on the basis of one of the available controllers working in the TCP/IP standard. Port no 100 was selected. (Cf. Fig.7)

Server - Controller DSM8051 communication

Communication between devices was set up to use the following control (counting from the Server). Transmission parameters were set to 1200 bauds (Theoretically - even for 57 600 bauds)

Microcontroller DSM8051

The DSM8051system [6] employs a Micromade 8-bit microcontroller . It is a programmable training controller equipped with both analogue and digital input/output ports, for connecting and handling peripheral devices.

, The stepper motor controlling software, which controls the step motor in different ways, was developed on the basis of available systems . (about 10 various programs) (Fig.8) . The software coded in microcontroller's assembler may be launched from the Client's remote desktop. User Interface for Client/Server mode was written in Microsoft Visual Studio 2010 Software.

Description of control program operation

There are three stages in the program:

- establishing connection between individual devices (Client, Server, Microcontroller)
- providing parameters to control the step motor from a remote control desktop
- carrying out the motor movement according to a preset set of characteristics.

Below there are examples of characteristics controlling the EDS 10 system.

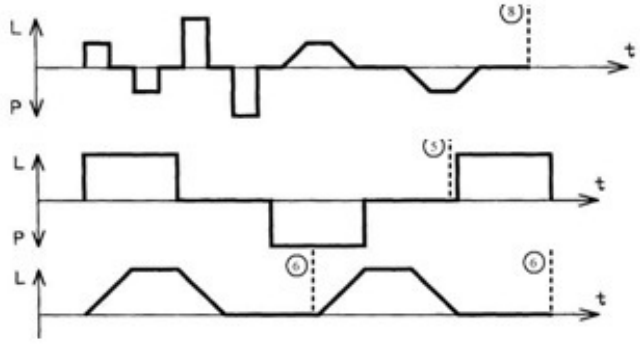


Fig. 8 Selected control characteristics of a step motor

System software has been presented as a UML diagram (use cases) below. (Fig. 9).

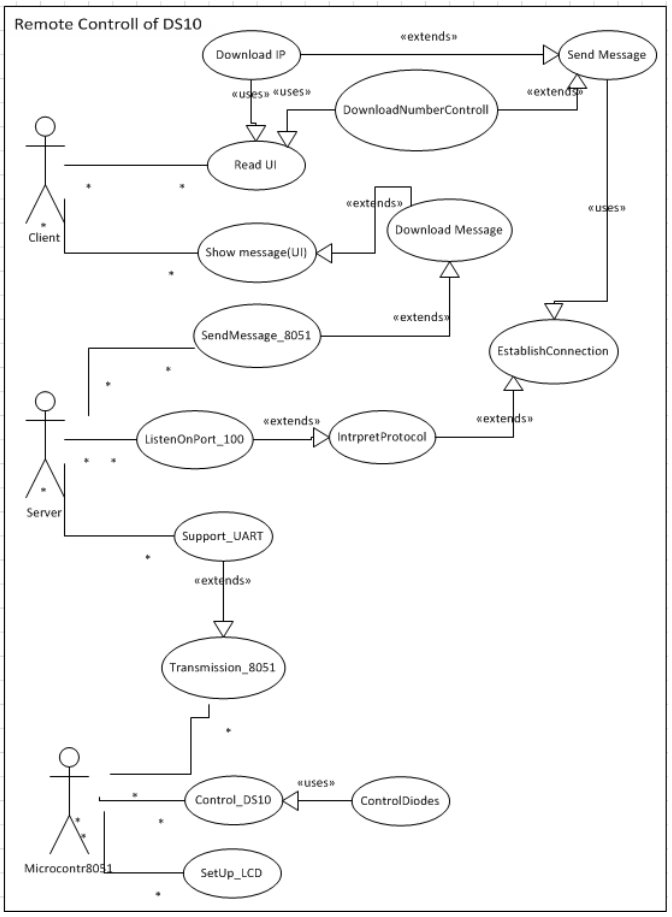


Fig. 9 Use cases for operating software

IV. CONCLUSIONS

The presented solution of remotely controlled communication with an executable system allows for:

- remote monitoring of an executable device – both continuous and on demand
- remote read-out of the system parameters - both continuous and on demand
- remote (e.g. cyclic) recording of measurement data in the microcontroller's memory, from where the data can be read out at any moment (periodically or on operator's demand).

Assuming that the server (Fig.5) symbolizes the on-board computer, there are many ways in which the described system can be used in a motor vehicle and they depend only on the designer's creative imagination and automation systems (microcontroller's) installed in the motor vehicle.

Connecting the Gateway (Fig.2), handling interbus communication to the on-board Computer (Server-Fig.5) instead of the executable device/EDS 10 enables the implementation of many telematic services in vehicles. Some of them are: BMW Assist or BMW Tracking ([3], [4]) and similar. Telematic BMW services are currently available in Austria, France, Germany, Great Britain and the United Arab Emirates.

Summing up, it should be stated that the main objective of the presented solution is long distance data transmition , which can be used in all fields of technology, e.g. telemetry, positioning or remote control.

REFERENCES

[1] B.Fryskowski, E.Grzejszczyk, ., "Systemy transmisji danych" (Systems of Data Transmission), Wydawnictwa Komunikacji i Łączności,Warszawa, 2010, pp.24, [1a] pp. 201

[2] W. Zimmermann, R. Schmidgall, "Magistrale danych w pojazdach" (Data buses in vehicles), Wydawnictwa Komunikacji i Łączności, Warszawa, 2008, pp. 24 - 101

[3] E. Grzejszczyk.,"Teleserwisowa komunikacja z pojazdem samochodowym" (Teleservice communication with a motorised vehicle), Electrical Review, ISSN 0033-2097, R. 87 NR 12a/2011,pp. 94-98

[4] E. Grzejszczyk, "Analiza wybranych usług teleinformatycznych serwisu BMW" (Analysis of selected telematic services), Electrical Review, ISSN 0033-2097, R. 88 NR 7a/2012, pp. 294 - 296

[5] Thesis, Gryszpanowicz B., "The implementation of Wide Area Network for remote communication with motor vehicle", Electrical Dept., Warsaw University of Technology,Warsaw 2004

[6] E. Grzejszczyk., A.Sęk., "The control on-line of selected functions of car system over GSM network", Scientific Papers of the Warsaw University of Technology, Electricity, No. 129, Publishing House of Warsaw University of Technology, Warsaw 2003

[7] 2. E.Grzejszczyk, "Introduction into Visual Web Developer Express 2008 and ASP.NET 2.0 Technology", Publishing House of Warsaw University of Technology, Warsaw 2010

[8] E. Grzejszczyk, "Basic tools of Information Technology", Publishing House of Warsaw University of Technology, Warsaw 2006

[9] Microprocessor Educational System DSM8051-www.micromade.com

[10] Information about the Author:

Elzbieta Grzejszczyk, Ph.D., M.Sc.(Eng.), Docent at Warsaw University of Technology (WUT) Warsaw,Poland;
1974- Degree of Master of Science (Eng.) WUT, - Electrical Engineering Institute of Control and Industrial Electronics;
1979- Degree of Ph.D/WUT- a thesis on artificial inteligence;
2007 - Docent/WUT Electrical Department;
Doc. E. Grzejszczyk has worked for about 10 years at the Industrial Research Institute for Automation and Measurements in Warsaw (Poland) as a software designer of microprocessor systems. Currently she works at the Faculty of Electrical Engineering WUT Poland, as a lecturer and teaches classes in Wireless Data Transmission Systems, Computer Systems in motor vehicles and C# object-oriented programming for students majoring in Computer Science.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 811–818

How to use the TPM in the method of secure data exchange using Flash RAM media

Janusz Furtak
Military University of Technology
ul. Kaliskiego 2,
00-908 Warszawa, Poland
Email: jfurtak@wat.edu.pl

Tomasz Pałys
Military University of Technology
ul. Kaliskiego 2,
00-908 Warszawa, Poland
Email: tpalys@wat.edu.pl

Jan Chudzikiewicz
Military University of Technology
ul. Kaliskiego 2,
00-908 Warszawa, Poland
Email: jchudzikiewicz@wat.edu.pl

***Abstract*—This document describes how to use the Trusted Platform Module (TPM) in the method of secure transmission of data stored on the Flash RAM through insecure transport channel. In this method the sender of the file specifies the recipient and the recipient knows who is the sender of the file. The idea of a solution that uses symmetric and asymmetric encryption is described. The TPM is used to safely generate symmetric and asymmetric keys, and theirs the safe collection, storage and management in order to protect files during transfer. The way of organizing data in a cryptographic keystore for users authorized to use the system for the secure transmission of files stored in Flash RAM is described.**

I. Introduction

IN THE era of widespread use of the Internet moving data from one location to another is a natural phenomenon. Data may be transferred with the use, for example, of HTTP protocol (files describing the presentation of Websites in HTML) FTP protocol or others. Transferring a non-confidential data over the Internet is not born of trouble. The problem occurs when the transferred data are sensitive and can be made available only for selected recipients. Then securing the data transferred on the route between the sender and the recipient is necessary. For this purpose, in Internet TLS protocol (Transport Layer Security) at the application layer or IPSec protocol (Internet Protocol Security) at the network layer are commonly used. The applied solutions in this area use symmetric and asymmetric cryptography. In both approaches from the point of view of security, the most important task is authentication both sides communicating against each other. This can be achieved by using one of the following methods:

- password known for both sides (shared secret);
- manually exchange of public keys of both sides;
- using public key certificate (X.509).

Normally is used the last solution, which is most comfortable, but requires the access to the Certification Authority.

Transfer data in the data processing systems of different classification levels is much more complicated problem [4]. In such systems, it is important not just who is the data recipient, but also whether the data recipient has adequate security clearance certification. Due to the legislation regarding of classified data processing very often it is not possible using a computer network to data transfer and building a mechanism for user authentication [12]. Then the transfer of sensitive data over insecure transport channels becomes a necessity. In such systems, the only way to safe data transfers is secured a removable media (i.e. Flash RAM or removable disk) connected to the system by USB interface [3].

This paper presents a solution to such preparation the data stored in Flash RAM so that storage medium can be used for secure transmission of the data [3]. In this solution a sender of the data (ie, the creator of the protected media content) can be sure that the data will only be available for a designated recipient, and the recipient can be sure that the received data originates from the expected sender. The described mechanism uses both symmetric encryption algorithms and asymmetric. The presented solution uses a *filter driver* [1][5][6][9]. In this solution, it is assumed that in terms of operating system the data can be processed in two directions: from plain text form stored on your hard disk, to protected form on removable media (e.g. Flash RAM, hard drive) connected to the system via the USB bus and, conversely, from protected form on removable media to plain text form on the hard drive.

The process of securing exchange of data is transparent for the user except for moment at which the system will recognize new storage media Flash RAM connected to the station. Then, the user is forced to specify his preference of data encryption algorithms, origin of cryptographic keys and location of protected files. For this reason to implement the process securing data it is necessary using separate operating system modules (drivers) running on the system kernel mode [5][9][10]. Schematic diagram of developed solution is shown in Fig 1.

User have access to the described solution through the Control Applications (***CApp***). The basic elements of constructed system are cooperating with each other drivers: encryption driver [1][2][3] and driver supporting, which are compatible to the Windows Driver Model [5][9]. Both

This work was supported by The National Center for Research and Development, Project OR00014011

978-83-60810-53-8/$25.00

elements work in operating system kernel mode and communicate with each other using the internal mechanisms of the operating system (in the figure these mechanisms are labeled as IRP (Input-Output Request Packet) [5][7][9][10]. The purpose of the encryption driver **(EnD)** is realization of the process of encryption / decryption data and determination a value of hash function for these data. The tasks of the Driver Supporting **(DSu)** are among other things, creation of signatures for protected data and intermediation in transmitting messages / commands between **DSu** and **CApp**. Both drivers cooperate with the Trusted Platform Module (TPM). The purpose of the TPM are a secure cryptographic key generation, storing them and sharing, and supporting cryptographic procedures. Management Console for Cryptographic Module (MCCM) is used to management of TPM and cryptographic keystore management. An important element of the system is a DLL library that provides the encryption functions.

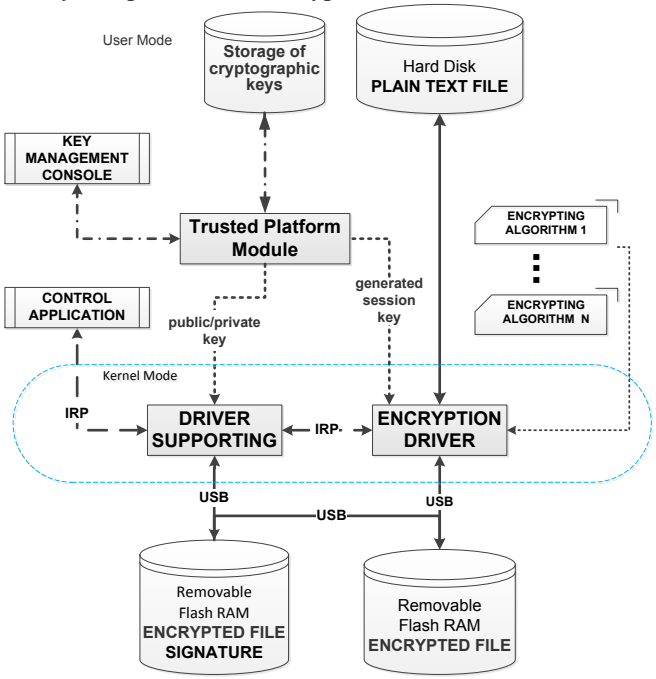


Fig. 1 Schematic diagram of securing data stored on removable media

The products of the process of securing data are two files: a file with encrypted data and signature of that file describing the file containing the encrypted data. The both files can be saved on one medium, or each file on different medium. Choosing a storage location of the signature file is defined by the user through **CApp**. It should be noted that saving the encrypted file and the signature file on separate media increases the security of stored data, but it is cumbersome to use. In the next sections of the paper is presented the process of creating and reading a protected file and the concept of using the capabilities of TPM to management of the cryptographic keys for procedures of protecting the file on a removable Flash RAM.

II. THE PROCESS OF CREATING AND READING A PROTECTED FILE

In the process of creating a protected file on removable flash memory (that is a creating an encrypted file and the signature of this file) and reading (decrypting) the file from the removable flash memory are necessary attributes of the user who created the protected file (this user will be called the sender) and user for whom the protected file was created (this user will be called the recipient). When a protected file is created the user logged into the system is the sender, and he specifies the file recipient using the **CApp**. When reading a protected file with using the **CApp** logged user plays the recipient role, a sender's attributes are read after successful decryption of signature of this file, using a private key of the logged on user. Permissible is a situation in which the logged user is simultaneously sender and the recipient of data.

The process of creating a protected file includes the step of encryption, and then creating a signature for that file. However during the process of reading a protected file in a first step the attributes needed for decrypting this file are obtained from the signature. In the second step the file is decrypted.

A. Creating a protected file

The process of writing the file, including file encryption and hash generation, is performed by the **EnD**. Operation of **EnD** has been presented in [1]. The diagram describing the process of writing the file is shown in Fig. 2. Dashed line in Fig. 2 indicates operations implemented by the **EnD**. During the process of file encryption is determined the value of the hash function to ensure the integrity of the file.

Determined value of the hash function, and the generated session key after completion of record are transferred to the **DSu** in order to generate a signature for the stored data. The process transferring the hash function value and the session key transferring is implemented using the system mechanisms marked in Fig. 2, as the IRP.

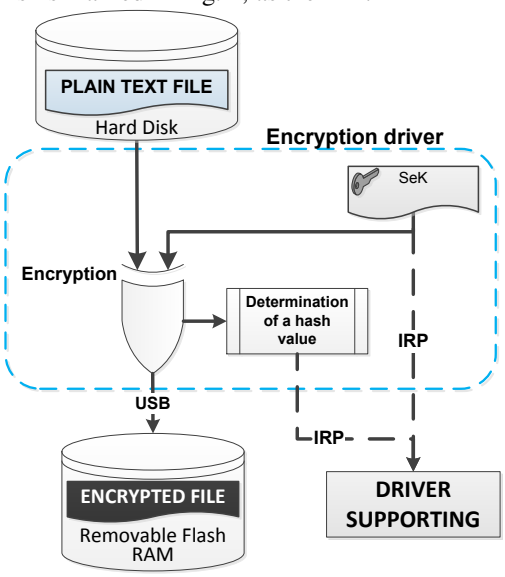


Fig. 2 The process of writing data to removable flash memory

B. Determining the signature

For each of the protected file the signature is generated which containing the information needed to read of this file. Signature of the file contains the following fields:

SeK - random key to encrypt / decrypt the secure file;
HASH - value of hash function which is determined based on the content of protected file after encrypting this file;
H_ID - identifier of the algorithm used to generate the hash;
En_ID - identifier of the algorithm used to encrypting;
O_ID - identifier of the logged user (the sender) who initiated the operation of data write - this identifier is required to determine the public key of sender when the file is read;
TMS - time stamp of file creation - this value corresponds to the date of file creation.

The structure of signature is shown in Fig. 4, and process of signature creation proceeds according to the diagram is shown in Fig. 3.

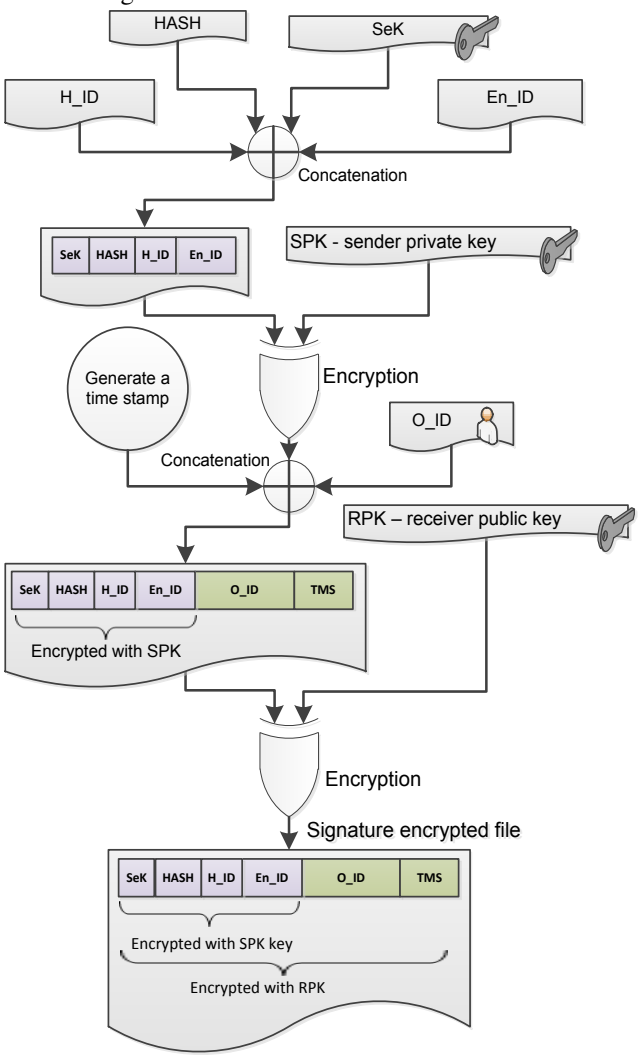


Fig. 3 The algorithm of signature generation for protected file

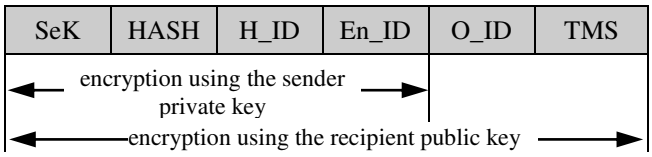


Fig. 4 The structure of signature secure file

C. Reading a protected file

The process of reading the file requires that the signature was read before and then decrypted. These activities are performed by the logged user (recipient of file) using **CApp**. The process starts with decrypting the signature file using the private key of the logged user, then reading time stamp and user identifier (O_ID) which assumed the role the sender creating a protected file. The time stamp protects the an encrypted file before moving it to another medium than that on which was originally written. Incompatibility of date and time stored in the time stamp and date and time, when the file was created, causes displaying the message and terminating the procedure of file reading. With compatibility of the parameters the next part of the signature is decrypted using the user public key which identifier (O_ID) has been read. The next steps of file decoding are schematically shown in Fig. 5.

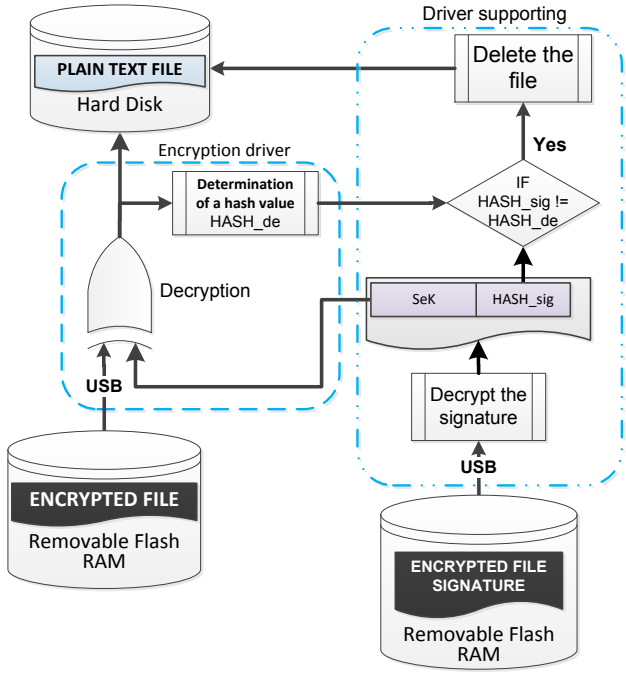


Fig. 5 The process of data reading from an external file

On the Fig. 5 operations performed by the **EnD** are marked using thick dashed line, and the operations performed by the **DSu** are marked using the thinner line (two dots dash).

During the read data the value of hash function (HASH_de) is determined. If the value HASH_de is different from the values obtained from the signature (HASH_sig) a message is displayed and the decrypted file, which was saved on hard disk, is being deleted [7].

III. THE CONCEPT OF USING TPM

Trusted Platform Module (TPM) is an implementation of a standard developed by the Trusted Computing Group [11]. This module is designed to support the cryptographic procedures and protocols that can be used for securing data [12]. Trusted Paltform Module, software TrouSerSwin consistent with TSS[1] and OpenSSL[2] library is a full-featured encryption system that provides the following functions:

- generating an asymmetric key pair;
- secure storage of keys;
- generating an electronic signatures;
- encryption and decryption;
- implementation of an operation defined by the standard PKCS # 11 cryptographic.

The TPM uses the mechanism of Root of Trust (RT) of secure data transmission in a computer system and is the basis to perform trusted cryptographic operations. This mechanism consists of the following components [8]:

- Root of Trust for Measurement (RTM) – uses Platform Configuration Registers (PCR) to record the state of a system;
- Root of Trust for Reporting (RTR) – uses PCR and RSA signatures to report the platform state to external parties in an unforgettable way;
- Root of Trust for Storage (RTS) – Uses PCR and RSA encryption to protect data and ensure that data can only be accessed if platform is in a known state.

The following algorithms are typically implemented in TPM: RSA, SHA-1, HMAC and AES[3]. In addition, each TPM chip stores a unique serial number and your RSA private key that is never available to read. TPM components are shown in Fig. 6.

The most important element of TPM is the cryptographic coprocessor. It performs the following actions:

- generating keys for asymmetric and symmetric cryptography using a Random Numbers Generator;
- encryption/decryption of data using the RSA algorithm;
- supporting of the integrity protection (SHA-1 Engine) and authentication (HMAC Engine).

The non-volatile memory of the module stores the asymmetric keys[4]: Endorsement Key (EK), Storage Root Key (SRK) and symmetric key NV_KEY. The EK key is usually generated during the production of the TPM and is used to decrypt the certificates for the other keys generated by the TPM. The SRK key is generated during the initialization of the TPM and is used as a master key to secure store of users' keys. In the key store are stored a keys of users who can create protected files (file sender) and users who will be recipients of protected files. The RK key is used to import the public key of recipients protected files. The NV_KEY key is designed to encrypt a data resource of users.

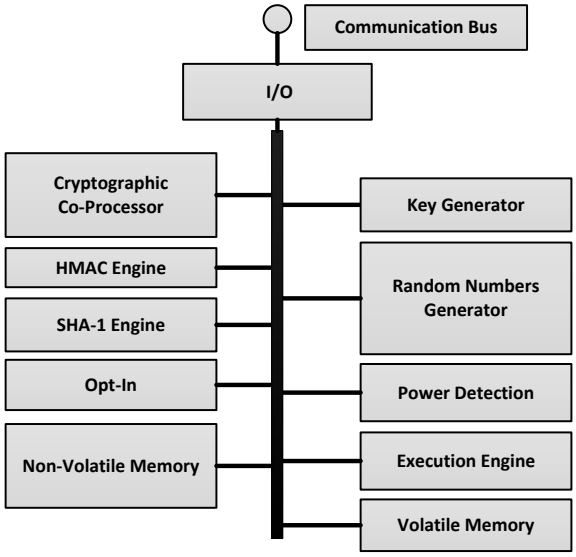


Fig. 6 TPM Component Architecture (based on [11])

For secure storage of user data involved in the exchange of protected files, especially cryptographic keys for that users, in the system created two protected resources: cryptographic keystore and TPM users' table.

A. Cryptographic keystore

The cryptographic keystore is designed to securely store, manage and share keys of all users involved in the exchange of protected files. These users include:

- local users who are on the station will be acted as the sender of uploaded files or recipient of received files;
- external users who are adequately recipient uploaded files and sender received files.

Keys for local users is generated by the TPM. The private key and public key of local users are stored in the cryptographic keystore. For each external user is kept only the public part of the key. The keys of external users are imported from other stations. Each key in the cryptographic keystore is identified by Universally Unique Identifier (UUID)[5].

Cryptographic keystore has a hierarchical structure (Fig. 7). At the top of the hierarchy is placed Storage Root Key (SRK). That key is generated by the TPM during the initialization of the module and is always stored in the non-volatile memory of TPM. Public part of the SRK key is used to encrypt the keys present in the hierarchy below the SRK key. The SRK key is the predecessor of the following keys:

[1] TSS – Trusted Computing Group Software Stack – specification of the software that allows the implementation of applications based on the TPM. For Windows OS, it is a project "TrouSerSwin", and in the case of Linux systems, it is possible to use the project "TrouSerS".

[2] OpenSSL – OpenSource cross-platform library that contains the implementation of protocols and general-purpose cryptographic algorithms.

[3] TPM uses a symmetric algorithm AES to protect the confidentiality of the session in which it participates following the recommendations of the TCG. However, symmetric encryption functions are not normally accessible outside the TPM.

[4] The Endorsement Key (EK) and the Storage Root Key (SRK) are mandatory for TPM while the Roaming Key (RK) and NV_KEY are necessary for the implementation of security procedures for files on Flash RAM.

[5] In the present solution as the UUID is used Globally Unique Identifier (GUID) of objects in Windows which is assigned to individual users.

- Roaming Key (RK) - used in the procedure to import the public keys of external users;
- Storage Key (SK) of local or external user - is used to encrypt the encryption keys used by individual users.

The safest place for storing cryptographic keys is non-volatile memory of the TPM, but due to the small size of the memory only the most important keys are stored there. Other keys are placed on your hard disk. The arrangement of keys is show in Fig. 7 to cryptographic keystore management is used the TrouSerSwin library with service TCSD[6] which works according to recommendations of the Trusted Computing Group.

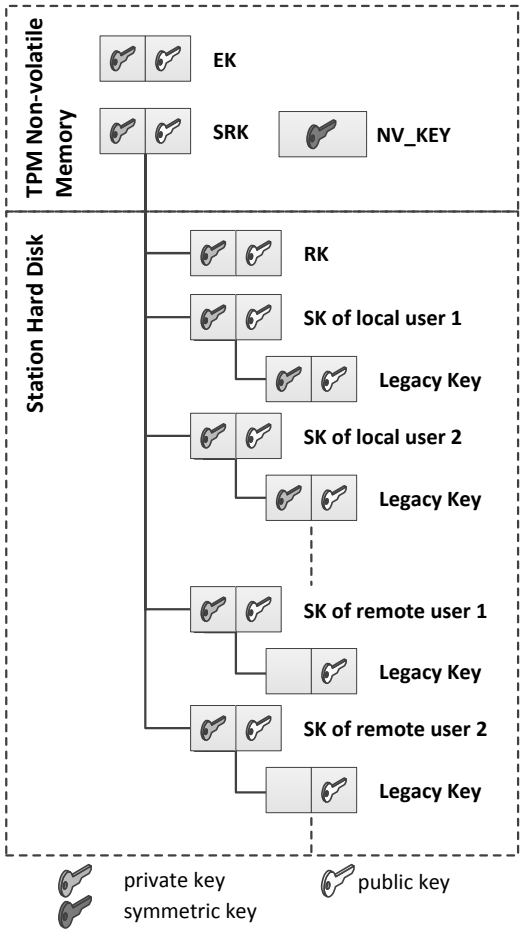


Fig. 7 Hierarchy of cryptographic keys supported by TPM

B. TPM users' table

The TPM users' table stores users data involved in the exchange of protected files. These refers to local and external users. A user with an access account to resources of the station may be considered as the local user. Otherwise an external user is the user with an access account to another station and whose data were imported from that station. The following record is combined with each TPM user:

- name of access account;
- display name of user;
- user Security Identifier (SID) on the local station - identifier used in Windows systems to determine access rights - USER_SID;
- Globally Unique Identifier of User - is used to identify the SK key of user – USER_GUID;
- identifier of user Legacy Key - this key is used to encrypt data intended for that user – USER_LEG_GUID.

The data of TPM users are encrypted with a symmetric key NV_KEY stored in non-volatile memory of TPM.

C. Cryptographic module

The cryptographic module is used to manage TPM users. This module consists of the following components:

- TPM;
- cryptographic keystore and a table of TPM users;
- TrouSerSwin library with service TCSD and Open SSL library;
- Management Console for Cryptographic Module (MCCM).

Architecture of cryptographic module is shown in Fig. 8.

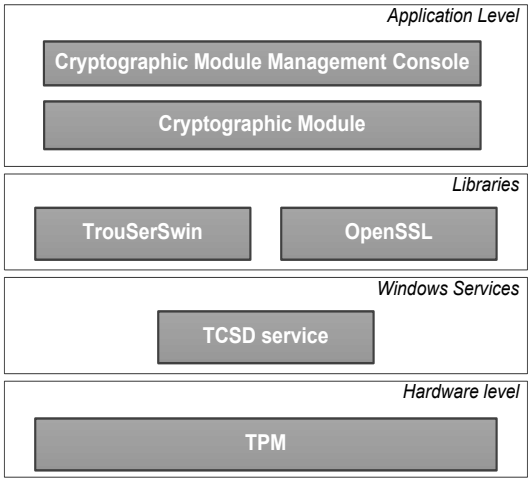


Fig. 8 Architecture of cryptographic module

The management console can be operated in a cryptographic module configuration mode and in TPM user management mode. The following functions are available in cryptographic module configuration mode:

- **login/logout** – opening / closing session of the cryptographic module configuration (entering the password for the owner of the TPM is necessary to opening the session);
- **reset TPM** – cleaning of TPM, i.e. delete all data from the cryptographic keystore and from TPM users' table;
- **initiate TPM** – creation of data structures for cryptographic keystore and for TPM users' table, and in particular the following keys: SRK, RK and NV_KEY.

A screenshot of the management console in configuration mode is shown on Fig. 9.

The following functions are available in user management mode:

- **Add user** – add a local user;
- **Remove user** – remove the user together with his keys stored in the cryptographic keystore;

[6] TCSD – software for management of TPM resources and support of local and remote requests. The software is part of TrouSerSwin software or TrouSerS software.

- **Generate keys** – generating the cryptographic keys for a local user;
- **Remove keys** – removal of previously generated cryptographic keys for the user;
- **Initiate import** – preparation of removable media to import external users data;
- **Import users** – import external users data;
- **Export users** – save the data of local user on previously prepared USB stick to transfer the data to another station.

A screenshot of the management console in user management mode is shown on Fig. 10.

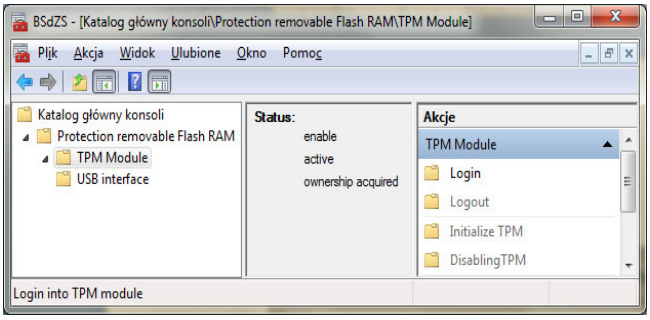


Fig. 9 A screenshot of the management console in Configuration Mode

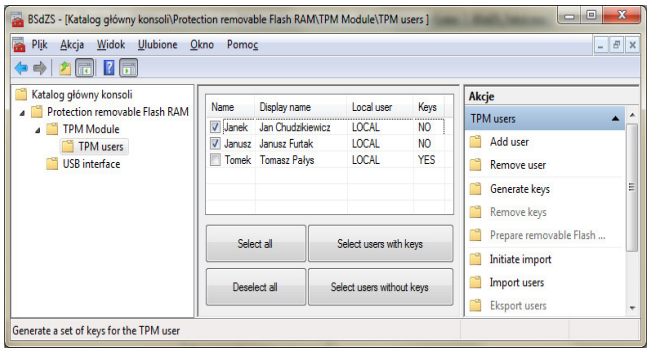


Fig. 10 A screenshot of the management console in User Management Mode

The "Add user" is designed to select from among the users defined in the system users who will be able to transfer protected files. For each selected user the GUID is generated. Data for these users are placed in TPM users' table.

The "Generate keys" is designed to generate the SK key and Legacy Key for local TPM user and put the keys in the keystore.

D. Export / import of user data

The procedure for preparing a protected file requires public key of the external user (file recipient) The source from which the keys are obtained, are stations where target user is a local user. The procedure for transfer of selected data about the local users of station ST2 to station ST1 includes the following activities (Fig. 11):

- transfer the public part of the RK key of ST1 station (e.g. via Flash RAM) - preparing a Flash RAM using the action "initiate import";
- providing media to the station ST2;
- preparation of a protected file that contains data about the local users of station ST2 using the action "Export users";
- providing the file and signature to the station ST1;
- decrypt user data, adding data these users (as an external users) to TPM users' table, and keys of that users to the keystore using the action "Import users".

Transfered protected file contains data of only selected local users of station ST2. The data of each user include:

- display name of user;
- Globally Unique Identifier of User – USER_GUID;
- identifier of the user Legacy Key - USER_LEG_GUID;
- public part of the user Legacy Key.

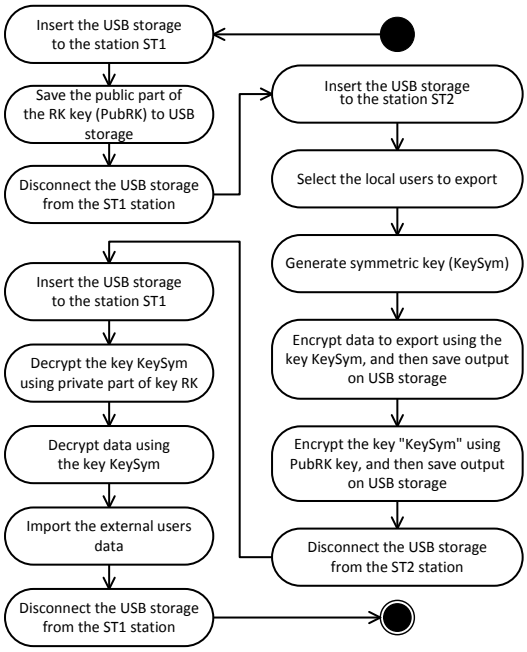


Fig. 11 Import data of users from station ST2 to station ST1

The file containing the user data is encrypted using a symmetric key generated at the station ST2. For that file is prepared other file (signature) containing symmetric key used to encrypt that file and value of SHA1 hash function for that file. The file (signature) is encrypted using the previously supplied public part of the key RK ST1 station.

At the station ST1 the received file containing signature is decrypted using the private key RK of the station ST1. In the next step the integrity of the file is checked up, then the file is decrypted using the symmetric key contained in the signature file.

The new records are created in TPM users' table basing on contents of received file. For each new external user in cryptographic keystore creates a hierarchy of keys containing the key SK and Legacy Key. The key SK is generated and will be identified by the sent user identifier USER_GUID. Lower in the hierarchy will be placed sent the Legacy Key.

IV. HANDLING FOR CREATING AND READING A SECURE FILE

The process of creating protected file requires first of all connection one or two (depending on where the file with the signature will be stored) removable Flash RAM memories to a computer through USB interface. The devices are automatically detected by **EnD**, which transmits information about them via the **DSu** to **CApp**.

Logged user (file sender) configures the parameters of the process of creating a protected file using **CApp**. These parameters are as the following:

- location of public key data recipient file ("Location of Public Key" field) – TPM is default (see Fig. 12), but it is possible provide the recipient public key from a disk (see Fig. 13);
- identifier for user (receiver) encrypted data ("Data Recipient" field) – one of the users whose keys are stored in the assets managed by the TPM;
- drive in which will be stored the protected file ("Data Drive" field);
- drive and path to the directory in which will be stored the file with the signature ("Signature Drive" field);
- identifier for the algorithm used to encrypt ("The encryption algorithm" field);
- identifier for the algorithm used to generate hash value for the protected file ("The Hash function algorithm" field).

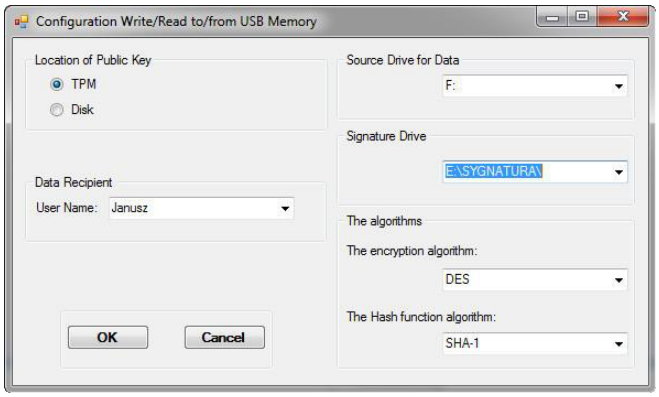


Fig. 12 The window of control application CApp for write data - recipient's public key is derived from the TPM

Identifier (O_ID) and the private key of the sender (the elements required to generate the signature) are automatically retrieved from the cryptographic keys storage managed by TPM. After determining the data configuration logged user can begin the process of copy the file using , e.g. Windows Explorer. The name of file which stores the signature will be concatenation of the name of protected file and string "SIG". The process of creating a file with the signature is started after the encryption process is finished and is, just as the encryption process, invisible to the user. When next file for the same recipient is being encrypted it does not need to change the configuration data unless the other parameters (that is the identifier of encryption algorithm or identifier of algorithm generating of hash value) will be changed. Always for the next file a new session key will be automatically generated.

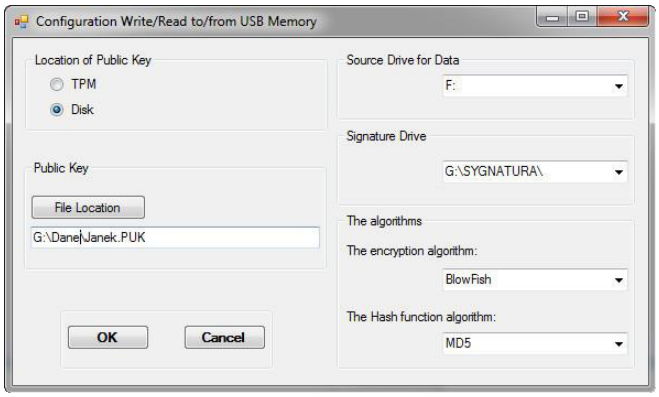


Fig. 13 The window of control application CApp for write data - recipient's public key is derived from disk

The process of reading protected file requires connection to a computer through USB interface one or two (depending on where is stored the file with the signature) removable Flash RAM memories. The devices are automatically detected by **EnD**, which transmit information about them via the **DSu** to **CApp**. The logged user (recipient of the data) using **CApp** has to specify the drive on which is stored encrypted file and indicate the file with the signature corresponding to the encrypted file. He accomplishes this by selecting (see the Fig. 14):

- drive on which is stored the protected file ("Data Drive" field);
- drive and path to the directory on which is stored the file with the signature ("Signature Drive" field).

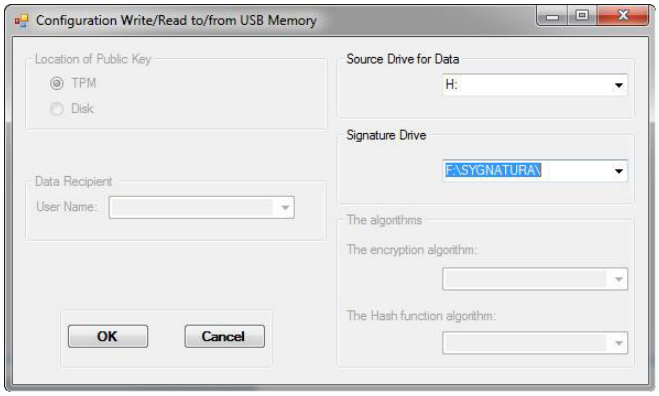


Fig. 14 The window of control application CApp for read data

Other parameters required to decrypt the file are determined based on the signature. After initializing by the logged user the process of copying a file **EnD** sends to the **DSu** the name of the copied file and pauses the copy process to the moment when receives data required to decrypt the file (that is the identifier of encryption algorithm, session key and identifier of algorithm generating of hash value). Based on submitted by the **EnD** the name of encrypted file, **DSu** identifies a file containing the signature and performs

the process of signature decryption and reading the configuration data. Then performs the verification process read out TMS with the date and time of the creation of an encrypted file. In the case of inequality of these values message is displayed and the file reading process is interrupted. In the case of equality of those values other configuration data read from the signature are passed to ***EnD***, which resumes the process of decryption. During the process of decrypting the file the ***EnD*** determines the value of a hash function for that file. After completion the process of copy ***EnD*** transmit to ***DSu*** determined value of the hash function for verification. If the designated hash value is not equal to the value read from the signature a message is displayed and the ***DSu*** deletes the file which has just been decrypted.

II. Conclusion

This paper presents a method for securing data on removable Flash RAM used to transfer sensitive data over insecure channels of transmission, such as a courier, traditional mail system and so on.

The developed method ensures the integrity and confidentiality of protected file. File sender (i.e., the creator of the protected content) is to ensure that the data will be available only to the specified recipient. The recipient can be sure that the resulting data are derived from the appropriate sender. Protecting the original file consists in preparing two files. The first one contains the encrypted contents of the original file (symmetric algorithm is here used). The second file is the signature for the encrypted file and contains the encryption key and the hash function value for encrypted file. The second file is encrypted using asymmetric cryptography. This approach ensures the safe transfer of the encryption key between stakeholders.

Both parties involved in transferring such protected data are forced to safely generate asymmetric keys and the safe theirs collection, storage and management. Particularly important is the proper procedure to generate symmetric key. For this purpose, the TPM has been used. This module supports (by using the hardware) many activities related to securing data on removable media and provides secure management of cryptographic keys.

From the safety point of view the most sensitive point of presented system for protecting the files transferred using removable Flash RAM is the procedure for export / import keys of external users. Particularly sensitive operation is "manual" transfer of a public key of station to which the data of users should be imported. The presented method is intended for use in an environment where it is not possible to use a computer network for data transfer. Such restrictions may apply in case of data transferring between systems, which belong to different security domains (i.e. with different levels of classification)[2][4] [12]. For this reason, it is not possible to use public key infrastructure (PKI) and Certificate Authority (CA). On the other hand organizational requirements for how to process data in such systems and how to use the removable media give a guarantee for safe transfer of the public key by the described procedure of export/import data. It should be noted that in cases where there are no obstacles to the use of computer networks in the presented procedure of export / import data is possible to use PKI and CA.

References

[1] J. Chudzikiewicz, "Zabezpieczenie danych przechowywanych na dyskach zewnętrznych" in *Metody wytwarzania i zastosowania systemów czasu rzeczywistego*, 2nd ed. vol. 3, J. Peters, Ed. Warszawa: Wydawnictwo Komunikacji i Łączności, 2010, pp. 211–221.

[2] J. Chudzikiewicz, J. Furtak, *Cryptographic protection of removable media with a USB interface for Secure Workstation for Special Applications,* Journal Of Telecommunications And Information Technology, no. 3, 2012, pp. 22-31.

[3] J. Chudzikiewicz, J. Furtak, *The method of secure data exchange using Flash RAM media*, Proceedings of the Federated Conference on Computer Science and Information Systems, 2012, pp. 621-626.

[4] A. Kozakiewicz, A. Felkner, J. Furtak, Z. Zieliński, M. Brudka, and Małowidzki M., *Secure Workstation for Special Applications*, Lecture Notes in Computer Science, Vol. 187, pp.174-181, Springer 2011.

[5] *Microsoft Windows Driver Kit (WDK,)*, Technical Documentation, Redmond, Microsoft Corporation, 2009.

[6] R. Nagar, *Filter Manager*. Redmond, Microsoft Corporation, 2003.

[7] R. Nagar, *OSR's Classic Reprints: Windows NT File System Internals*. Redmond, OSR Press, 2006.

[8] R. Ng, Trusted *Platform Module. TPM Fundamental,* APTISS, August 2008 (http://www.cs.unh.edu/~it666/reading_list/Hardware /tpm_fundamentals.pdf)

[9] W. Oney, *Programming the Microsoft® Windows® Driver Model*, Redmond, Microsoft Press, 2003.

[10] M. E. Russinovich, D. A. Solomon, A. Ionescu *Windows Internals Part 2*, Edition VI, Redmond, Washington, Microsoft Press, 2012.

[11] *TPM Main Part 1 Design Principles. Specification Version 1.2. Revision 116*, Trusted Computing Group, Incorporated, 2011

[12] *TCG Software Stack (TSS) Specification Version 1.2 Part1: Commands and Structures* (http://www.trustedcomputinggroup.org /files/resource_files/6479CD77-1D09-3519-AD89EAD1BC8C97F0 /TSS_1_2_Errata_A-final.pdf).

[13] Z. Zieliński, at all, *Trusted Workstation for Processing of Multi-level Security Data*, Journal of Telecommunications and Information Technology, 2012, pp. 5-12.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 819–823

LocFusion API – Programming Interface for Accurate Multi-Source Mobile Terminal Positioning

Piotr Korbel[1]
[1]Institute of Electronics
Lodz University of Technology
ul. Wólczańska 211/215, 90-924 Łódź, Poland
Email: piotr.korbel@p.lodz.pl

Piotr Wawrzyniak[1,2], Sebastian Grabowski[2], Dorota Krasińska[2]
[2]Orange Labs (Poland)
ul. Obrzeżna 7
02-691 Warsaw, Poland
Email: piotr.wawrzyniak@orange.com

***Abstract*—The aim of this paper is to present a prototype LocNet API programming interface for indoor positioning systems and a prototype LocFusion API interface enabling joint use of terminal positioning data from mobile operator's GMLC and the LocNet API. The use of data from complementary information sources can improve the accuracy of user terminal positioning in large buildings, where coverage of satellite systems is weak.**

***Index Terms*—Location Services (LCS), Application Programming Interface (API), Open Middleware, Mash-Up Services**

I. Introduction

THE ABILITY to determine the geographical position of a wireless terminal allows telecom operators to implement new Location Based Services (LBS) [1], [2], [3], [4]. Mobile operators can provide their subscribers with a variety of information related to their actual geographic location, e.g. the nearby points of interest such as ATMs, hotels, restaurants, gas stations, traffic information, etc.

Terminal localization techniques used in contemporary mobile communications networks vary in accuracy offered, implementation complexity and cost. The simplest positioning methods, such as Cell-ID or received signal strength (RSS) based, are usually easy to implement in existing networks and do not require any modifications to the user terminals. The accuracy offered by these methods strongly depends on the size of the cells as well as on radio wave propagation conditions. In typical scenarios, the positioning accuracy varies from a few hundreds to tens of meters. Much better positioning accuracy may be achieved with the use of satellite positioning systems, such as the popular GPS. Satellite systems allow to determine user location with an accuracy of up to a few meters. However, the use of satellite receivers for accurate positioning of mobile terminals in indoor environment or in densely built up city centers is difficult or impossible. Thus, due to the weak coverage, satellite systems are not useful in areas with a large concentration of users (i.e. where such services are needed) [5]. These limitations of contemporary positioning systems contribute to the search for alternative high accuracy and availability methods suitable for complex environments.

One of the possibilities to increase accuracy of positioning in indoor areas is to use dedicated local positioning systems [6] along with contemporary Gateway Mobile Location Centre (GMLC) service. An attempt to integrate local indoor positioning systems with UMTS network architecture was described in [7]. The authors proposed integration of local indoor positioning networks (Assistant Location Networks) directly into core network.

In this paper we present a prototype LocFusion API. This programming interface allows joint use of information obtained from the GMLC of the network operator and from an indoor positioning system. In the prototype implementation, the access to the GMLC data is provided by Orange Labs Poland with the use of TerminalLocation API Telco 2.0 interface [8]. The information from local positioning systems is available through the proposed LocNet API interface.

The remainder of this article is organized as follows:

- Section II presents the basic concept and principles of operation of LocNet API prototype. LocNet API allows to access the results returned by local indoor positioning systems extending the range of satellite based localization services.
- The architecture, the operation, and the implementation of LocFusion API prototype are described in section III. LocFusion API allows joint use of location information from the operator's GMLC and the LocNet API.
- Section IV provides possible usage scenarios for LocFusion API.
- Section V summarizes the paper.

II. LocNet API

LocNet API is a universal programming interface used to communicate with compatible local indoor positioning systems. The interface retrieves the coordinates (x, y, z) describing the user's location inside the building along with supplementary position related information, e.g. description of the zone of the building, floor number etc. The interface specifies only communication protocol and does not impose implementation of any particular positioning methods and algorithms. Thus, the LocNet API can be used with many different classes of positioning systems.

The prototype implementations of LocNet API described in the paper were tested with two different positioning systems: LocNet indoor positioning and tracking system developed and maintained by Lodz University of Technology [9], [10], and LocNet-PW terminal positioning system developed by Warsaw

978-83-60810-53-8/$25.00

University of Technology and Orange Labs Poland. Both systems used throughout the tests employ GSM and/or Wi-Fi received signal strength (RSS) measurements [11] to estimate the actual position of the terminal [12]. However, the two systems employ various classes of positioning algorithms and are implemented with the use of different server technologies.

To enable interaction between local positioning servers and LocNet API, the servers should be capable to accept LocNet API compatible requests and return positioning results in LocNet API compatible data format.

LocNet API application server is responsible for registration and management of local positioning servers providing terminal location information. This is achieved with the use of a database storing information on LocNet locations (i.e. sites where local positioning systems are available) and corresponding configuration parameters.

LocNet API application servers also triggers GSM/Wi-Fi RSS measurements required to estimate actual terminal position. To achieve that, the server communicates with user terminals. To support the RSS measurements, a dedicated LocNet Android application was developed. The application designed for Android based mobile devices performs GSM/UMTS/Wi-Fi signal strength measurements and passes the results back to the server. After successful initialization, LocNet Android application runs in the background. The application can be activated with the use of SMS message sent by LocNet API. After activation, the application starts the measurement procedure and sends the results back to the LocNet API. RSSI measurement results are sent in JSON format.

III. LocFusion API

The main goal of the LocFusion API interface is to allow joint use of terminal location data from different, complementary sources [13]. LocFusion API permits co-operation with local indoor positioning systems through LocNet API and with the GSM/UMTS network Gateway Mobile Location Centre (GMLC) via RESTful API exposed by PLMN operator. The possibility of common use of the location information from different sources makes it possible to determine more accurate information on the probable location of the user terminal.

LocFusion API enables to determine the user's location information in one of the two modes. Basic operation mode involves determining the user's location based on the results of the received signal strength measurements collected at a single point of the building. The other mode involves the possibility of determining the position of the terminal based on a sequence of signal strength measurements collected along the path describing the movement of the user [14].

LocFusion API interface makes it possible to easily incorporate additional sources of positioning information, such as RFID, dedicated WLAN systems using other communication protocols (ZigBee, 6LoWPAN), inertial sensors, etc.

A. API Architecture

The system is composed of three main modules: local positioning networks coupled with local positioning server, dedicated application server that manages all local positioning networks and exposes their features via web-based API, and main application server that manages all of the elements and communicates with the user. A database of sites where wireless local positioning networks are deployed is managed and supervised by LocNet API application server. The server also provides indoor positioning data for the API. An important part of the prototype system is Telco 2.0 Web Gateway that is used to incorporate external GMLC into the system. The architecture of the proposed solution is presented in Fig. 1.

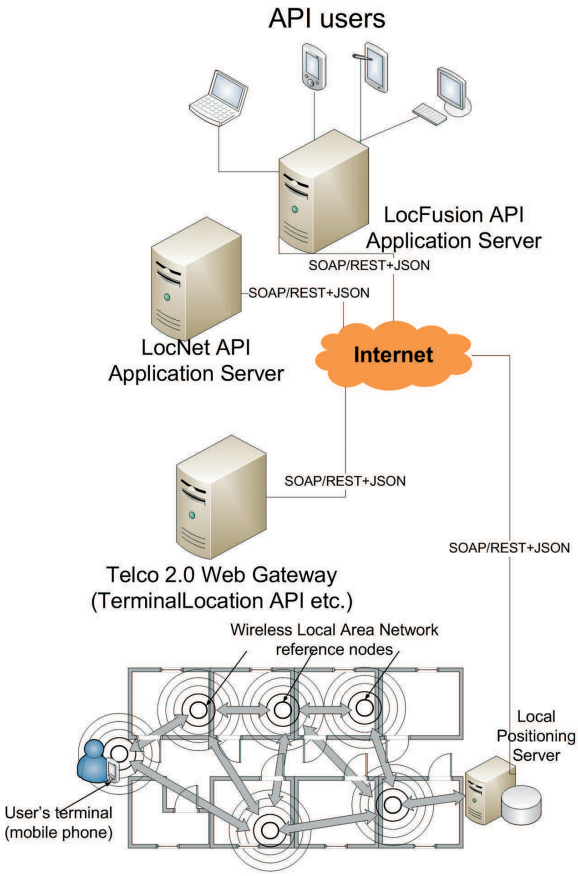


Fig. 1. LocFusion architecture.

Main application server is also responsible for managing security policy of the LocFusion system, especially for user authorization and maintaining legal affairs of estimating user's position in PLMN.

B. Software Architecture

LocFusion API has a modular software architecture. Every module can be developed separately and independently of each other until interfaces interoperability is kept. Moreover, at functional level LocFusion software is layer-organized as shown in Fig. 2.

Modular software architecture permits to expose certain positioning modes over the northbound interface. For prototype implementation three different API modes have been

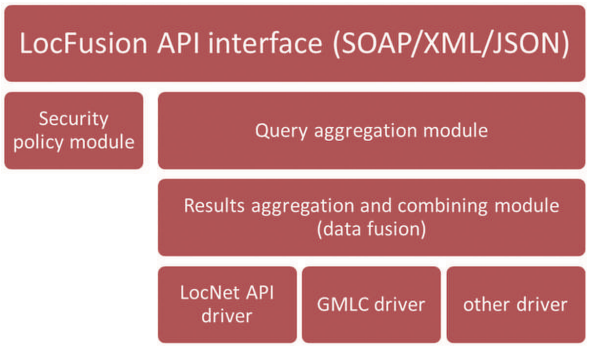


Fig. 2. LocFusion Software Architecture Functional Diagram.

implemented:

1) Single position estimation:
   - With remote call address (callback) – in this mode request identifier is returned immediately while entire terminal position is passed to provided callback address as soon as it is available,
   - At single HTTP transaction – in this mode HTTP response is sent to the user directly with estimated position. Server response contains only position estimator, request identifier is not passed to the end-user.
2) Multiple position estimation:
   - With remote call address (callback) – in this mode request identifier is returned immediately while entire terminal position is passed to provided callback address as soon as it is available. Terminal position is estimated periodically, the period between each two estimations is provided within request along with the total number of position estimations.

In each mode at least two parameters have to be provided in order to perform user positioning: valid authorization key (API-key) and phone number. Each API-key is then verified in Security Policy Module against phone number to verify whether positioning permission has been granted by the user. In particular, terminal can be localized only if user granted permission for MSISDN to be localized.

C. LocFusion Query Processing And Aggregation Algorithm

In many positioning system applications developers and designers focus on accurate positioning in means of estimating absolute geographical location of the end user. This approach is the most convenient way to mark position in outdoor scenarios, hence it is easy to integrate with GIS systems.

On the other hand, indoor positioning does not rely on absolute location identifiers since in-building descriptors (e.g. “office room 312” or “conference room”) can be more convenient to utilize. Moreover, when 3D positioning is considered accurate and reliable altitude identification is necessary. When considering outdoor positioning absolute altitude above the sea level is the most popular. For indoor positioning the most applicable altitude identification could be floor index

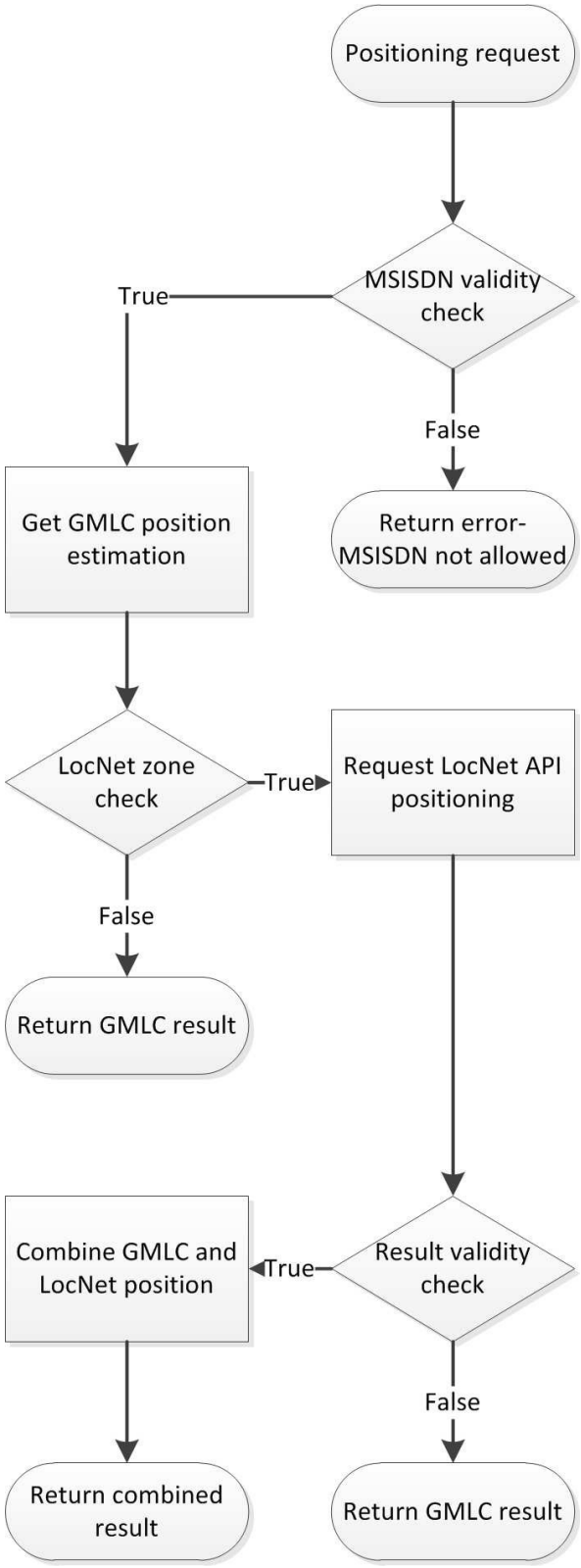


Fig. 3. Query processing algorithm.

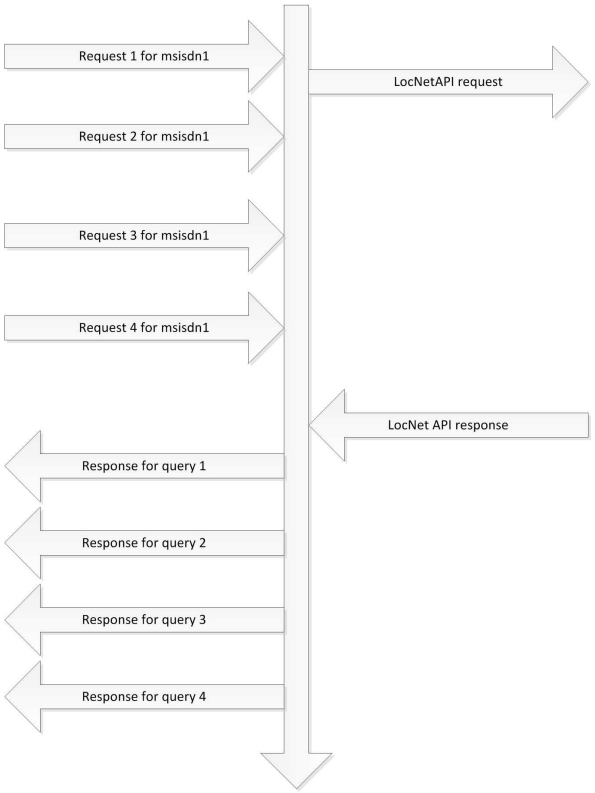


Fig. 4. Query aggregation flow diagram.

(e.g. ground floor, 1st floor etc.), hence it is more convenient building altitude identification rather than absolute values above the ground or sea level.

Therefore, LocFusion API is designed to provide additional descriptors for indoor position including floor index or name, zone within a building (e.g. "north wing"), room or office number or its name (e.g. "kitchen" or "office room 312") and site-related information (if available), including site features etc.

Moreover, contemporary absolute geographical position is calculated and returned for API backward compatibility. Absolute coordinates are also be the only result if no additional indoor positioning could be invoked.

Therefore every valid positioning request is internally processed with the use of an algorithm presented in Fig. 3. In order to increase API efficiency and to minimize LocNet API calls internal API request aggregation module has been implemented. Main feature of this module is to collect all valid positioning request for certain mobile subscriber and satisfy all the requests with single LocNet API call and thus single terminal call. Aggregation module operation idea is presented in Fig. 4.

IV. Usage Scenarios

LocFusion API might be seamlessly incorporated in a number of Location Based Services (LBS) but it is primarily designed to increase positioning accuracy inside the buildings. Therefore, possible use scenarios include but are not limited to monitoring and navigating the elderly or disabled people indoor, services allowing parents to localize their child, monitoring the trace of the employees (e.g. couriers or security guards).

The ability to determine reliable user position inside building makes it possible to aid elderly or visually impaired people in navigating inside public buildings like offices, hospitals or shopping centers. In this case LocFusion API might serve as reliable and accurate network-based user location source. Moreover provided context- related information might be used directly to inform users on their current location.

There are already developed applications that use positioning capabilities of the mobile network (offered by GMLC) to help parents in finding their child. These services offer rough location estimation and thus do not make possible to find a child for example inside big shopping center. In this case LocFusion API might provide both: rough location outside building (as accurate as the one offered by GMLC) and accurate context-related position inside the building. It is worth to mention that returned results include description of the building area (like "left wing corridor", "3rd floor") which makes them easy to use in finding people indoor.

LocFusion API makes it possible to send periodic position updates to the remote service and thus allows to implement a variety of tracking and monitoring LBS directly on the basis of the API. Possible use of the API includes monitoring and tracking of the employees that should follow a desired trace like couriers, security guards etc. It is also possible to use LocFusion API for Machine-To-Machine (M2M) applications, mainly for tracking of important and precious equipment (machines, electronic devices etc.) coupled with GSM module.

V. Summary

This article presented the prototype Application Programming interfaces that allows joint use of the dedicated indoor positioning system and GMLC in order to improve accuracy and usability of the location information. Modular system architecture and detailed description of the key components has been provided as well as detailed summary of developed query processing algorithms. The article also provides description of the possible use of LocFusion API.

Future works on the API includes evaluation of API performance in both computing and network usage aspects. It is also planned to undertake experiments to analyze positioning error when LocFusion API is used for tracking services.

References

[1] Ku Wei-Shinn, Chen Haiquan, Wang Chih-Jye, Liu Chuan-Ming, "GeoStore: A Framework for Supporting Semantics-Enabled Location-Based Services," IEEE Internet Computing, vol.17, no.2, pp. 35–43, March-April 2013.

[2] I. Maly, Z. Mikovec, and J. Vystrcil, "Interactive Analytical Tool for Usability Analysis of Mobile Indoor Navigation Application," *in Proc. 3rd International Conference on Human System Interaction,* Rzeszów, Poland,2010, pp. 259–266.

[3] Park, Jun-geun, et al., "Growing an organic indoor location system," *in Proceedings of the 8th international conference on Mobile systems, applications, and services,* ACM, 2010. pp. 271–284.

[4] M. Ibrachim, M. Youssef, "CellSense: An Accurate Energy-Efficient GSM Positioning System," IEEE Transactions on Vehicular Technology, vol. 61, no. 1, pp. 286–296, 2012.
[5] L. Zekeng, I. Barakos, and S. Poslad, "Indoor location and orientation determination for wireless personal area networks," Mobile Entity Localization and Tracking in GPS-less Environments, 2009, pp. 91–105.
[6] M.A. Islam, F. Belqasmi, R. Glitho, and F. Khendek, "The design and implementation of OMA restful location services in wireless sensor environments," IEEE Communications Magazine, vol. 51, no. 4, pp. 122–131, April 2013.
[7] F. Gil-Castineira, F. Gonzalez-Castano and J. Pousada-Carballo, "Integration of indoor location networks into the UMTS architecture: Assistant Location Networks," *in Vehicular Technology Conference,* 2004.
[8] Open Middleware 2.0 Community portal, http://www.openmiddleware.pl/, Accessed 19 May 2013.
[9] P. Korbel, P. Wasilewski, and P. Wawrzyniak, "Positioning Systems for WiFi and ZigBee Networks Aiding Visually Impaired in Indoor Navigation," ("Systemy lokalizacji terminali sieci WiFi i ZigBee do wspomagania niewidomych i słabowidzących we wnętrzach budynków"), Telecommunication Review – Telecommunication News (Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne), vol. 8–9, 2011, pp. 735–737. (in Polish)
[10] P. Wawrzyniak, P. Korbel, "Indoor positioning system for short range radio communication networks," *in Proc. II Forum Innowacji Młodych Badaczy 2011,* Łódź, Poland, 2011.
[11] J. Stefański, "Radio Link Measurement Methodology for Location Service Applications," Metrology and Measurement Systems, vol. XIX, no. 2, pp. 333–342, 2012.
[12] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, *WLAN Positioning Systems,* Cambridge University Press, Cambridge, 2012.
[13] P. Barański, M. Polańczyk and P. Strumiłło, "Fusion Of Data From Inertial Sensors, Raster Maps, And GPS For Estimation Of Pedestrian Geographic Location In Urban Terrain," Metrology And Measurement Systems, vol. XVIII, no. 1 pp. 145–148, 2011.
[14] J. Figueiras, S. Frattasi, *Mobile Positioning and Tracking,* John Wiley & Sons Ltd, Chichester, 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 825–828

Mobile Applications Aiding the Visually Impaired in Travelling with Public Transport

Piotr Korbel, Piotr Skulimowski, Piotr Wasilewski and Piotr Wawrzyniak
Institute of Electronics
Lodz University of Technology
ul. Wólczańska 211/215, 90-924 Łódź, Poland
Email: {piotr.korbel, piotr.skulimowski}@p.lodz.pl

***Abstract*—The paper presents a set of mobile applications aiding the visually impaired in using the public transport. A user equipped with a modern smartphone with mobile data transmission and positioning capabilities can access location related context information. Keeping up the connection with dedicated system servers gives the user access to additional services, e.g. enables the use of passenger information system and provides access to services facilitating the navigation in urban areas. The paper describes an overall architecture of the system for guidance and public transport assistance of the visually impaired. Then, the details of the applications developed for Android based smartphones are presented. The applications are mainly focused on aiding in urban navigation and provide various ways of accessing data from public transport passenger information system.**

***Index Terms*—Context-aware services, mobile computing, personal communication networks, pervasive computing, radio navigation**

I. Introduction

ACCORDING to the World Health Organization there were more than 285 million visually impaired living in the world in 2012 [1]. Even moderate vision impairment may strongly affect their everyday activities and often leads to social exclusion. Inability to sense the surrounding environment, poor orientation and navigation capabilities, difficulties in accessing textual information result in a limited mobility of the blind and the visually impaired [2]. Travelling becomes especially challenging in urban areas. Lack of good spatial orientation makes difficult to find a safe path among obstacles, and to locate and identify points of interest (POI) like bus stops or pedestrian crossings. Inability to access textual information like street name, public transport timetables, numbers of vehicles gives rise to additional difficulties. Recently, a number of electronic travel aids (ETA) addressing the needs of the visually impaired have been developed. The devices are used to overcome difficulties associated with everyday activities, i.e. problems with spatial and geographical orientation, navigation, accessing visual information. Electronic systems aid the visually impaired in mobility and in accessing various public services. One of the applications of the electronic aids is to facilitate access to public transport services. Precise information on user location can be used to retrieve position related data from public transport passenger information systems, e.g. bus or tram arrival times, information on routes, temporary changes to the timetables, etc. A number of electronic systems aiding the visually impaired in urban travelling involve various beaconing techniques to identify landmarks like bus or tram stops, entrances to public buildings, etc. [3][4][5][6][7]. Transmitters installed in the landmarks send signals uniquely identifying the place. System information can then be decoded with the use of a dedicated handheld receiver and presented to the user as voice messages. Another approach to guidance of the blind involves the use of dedicated user terminals equipped with GPS receiver, GSM transceiver and inertial sensors [8][9]. With the growth of popularity of advanced mobile phone terminals, more and more smartphone applications aiding the visually impaired in navigation and travelling appear on the market [10][11][12]. Some of them, like OnTheBus project [10], address also the problem of public transport accessibility. Significant number of various ETAs have been developed so far trying to solve different mobility difficulties. However, many of the assistive devices and applications address only selected aspects of the mobility problems, and hence have not gained wider acceptance of their target group of users so far. Therefore, there is still a pursuing need to develop complex solutions aiding the visually impaired in mobility as well as aiding other groups of users in travelling in urban environment.

This work was partially supported by the National Centre for Research and Development of Poland under grant no. NR-02 0083-10 in years 2010-2013.

II. System for guidance and public transport assistance

The architecture of the proposed system for guidance and public transport assistance of visually impaired is shown in Fig. 1. The system consists of several subsystems: a mobile user terminal, a network of radio beacons, and application servers. The mobile user terminal can either be a dedicated electronic device or an Android based smartphone. Dedicated terminals, equipped with GSM/UMTS transceivers, GPS receivers, inertial sensors and a camera are used to obtain precise user location information, to provide communication channel to remote assistant of the user, and to present voice messages to the user. The Android smartphone based version of the terminal in general plays the same role, however, its functionality can be easily modified by installing additional applications. In the next section of the article we present a

978-83-60810-53-8/$25.00

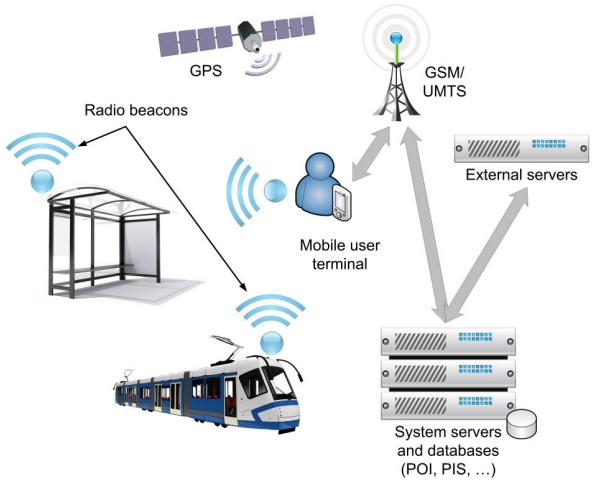


Fig. 1. Architecture of the electronic system for guidance of the visually impaired in urban environment.

set of mobile phone applications aiding in urban navigation and public transport accessibility. The network of low power and low range radio transmitters is used to provide precise information on actual location of the user and to facilitate access to position related data, e.g. a timetable of vehicles arriving at the bus or tram stop [13]. The servers of the system store public transport passenger information data and points of interest databases, as well as expose services enabling communications between all of the system components.

A. Passenger Information System

Currently a lot of cities operate real-time public transport management systems (PTMS). Such systems help transport company to increase the efficiency of running vehicles, reduce travel times and/or improve the punctuality, especially when PTMS is integrated with traffic lights management system. Very often PTMSs include passenger information subsystem (PIS) providing information about estimated arrivals and/or departure times as well as travel times. Usually such data are displayed on dedicated signs located within bus stops. However such approach is very useful for almost all passengers it is completely useless for blind and semi blind people for obvious reasons. Similar system is operated in Lodz, Poland by municipal transportation company MPK. Lodz is mid-size city in central Poland of the area 300 km^2 with more than 700000 inhabitants. MPK operates about 200 trams and 400 buses serving almost 100 routes. Trams and buses fleet is managed by RAPID system delivered by Sigtec, Australia integrated with adaptive traffic system SCATS developed by Road and Maritime Services in Sydney. Currently RAPID controls 14 passenger information signs located on tram stops, in 2013 next 16 signs will be added. Vehicle localization in MPK system is made by on-board GPS receivers. Vehicles send data to the server in 20 seconds intervals. Position messages are also sent when the vehicle enters or exits stops and intersections. This allows the server to estimate travel times and send predicted departure times to stop signs. In case of traffic problems passengers can also see appropriate messages. Unfortunately blind people due to the nature of their disabilities cannot be informed about, vehicles approaching or awaiting at the stops, vehicle delays and temporary changes in route paths introduced in response to serious traffic problems. Visually impaired people may also found problems with stops localization, especially when they do not know city topology or stops were moved to temporary locations due to street construction works. Blind aid system designed in Institute of Electronics is connected with MPK system through VPN tunnel and receives in real time all major information: current timetables, route paths, vehicles allocations, trips cancellations, detours, run-ins and run-outs. The system also incorporates radio tags placed on stops, points of interests as well as vehicles. There are running two services for travel aiding: trip planner and trip assistant. The aim of the trip planner is to find optimal from blind people point of view route from point A to point B. Trip planner optimizes travel time as well as walking distance necessary to reach the destination. Starting and ending trip points may be described as geographic coordinates and/or points of interest including public transport stops. Travel start time interval must also be described. For guided people convenience trip planner allows for maximum one vehicle exchange. Vehicle kind (tram, bus or both) may also be specified. Regardless the form of describing start and end points of the travel trip planner locates up to 10 nearest stops. Even if direct trips were found the system also searches for trips with one exchange. This allows to find several trips serviced with different routes and select the best one for the guided person. In the case where no trips were found system increases default values for the maximum distance between change stops and searches for start and end stops within greater radius. As a result following data for each trip are obtained: trip identifier, route number, direction, vehicle identifier and type (tram or bus), estimated arrive times for first and last stops, distance between starting point and start stop and the distance between ending point and end stop.

Arrive times are calculated by averaging travel times between consecutive stops for given route, type of the day and time of the day. As the type of the day working days, Saturdays and holidays are distinguished. The whole day is divided for 2 hours intervals, so separate calculations are performed for peak and off-peak hours. Trip assistant starts working when the passenger begins journey approaching starting stop. Awaiting passenger may be informed about estimated arrive time of the desired vehicle. Next, when passenger entered the vehicle, he/she may be informed about remaining stops to the destination or the exchange stop as well as about remaining time of the travel.

III. Mobile Applications for the Visually Impaired

To present the information from the public transport Passenger Information System to the users, a set of mobile phone applications was developed. POI Explorer and Public

Transport Explorer applications dedicated for Android based smartphones are used to aid the visually impaired users in urban navigation and travelling. Two other applications use NFC and USSD technologies to access the data. The NFC application can be used with any NFC enabled device. Taking into account low market penetration of NFC capable phones, also USSD messaging was implemented as an option available to almost all the range of mobile phones.

A. Smartphone Based Urban Navigation ij POI Explorer and Public Transport Explorer

Blind users usually use iOS or Android based mobile phones. The reason for that is that both the systems have built-in text to speech modules: Voice Over [14] and TalkBack for iOS and Android respectively. Availability of such system modules allows developers to create their own applications using standard GUI elements which can be easily presented to the visually impaired users. Most of smartphones are equipped with touch screens and have gesture-based screen readers, for example a single tap causes a button's description to be read, requiring a double tap to activate the button's original function. Such an interaction with a smartphone requires the use of both hands and may be especially uncomfortable for blind users who are at the same time using a white cane. That is why we proposed a dedicated electronic device, equipped with Bluetooth module and keyboard which can be used to control selected functions of mobile applications [15]. Moreover, it can be used to read data from a network of radio beacons indicating various point of interest [13], and to pass this information to the phone. Depending on the beacon type, application can present the user information on entering some area, on vehicles approaching a stop, etc. We developed two applications POI Explorer [16] and Public Transport Explorer aiding the visually impaired in travelling with public transport. They are dedicated for the most popular mobile platform: Android. First one uses points of interests to the navigation purposes. They are stored on remote MySQL/PHP server, which allows to keep the database update. Moreover, such a solution allows to provide a universal API for other platforms. To exchange the data between the server and mobile phones XML language was used. Users can also add additional personalized information to the points (text notes, voice records) to enrich the database. POIs are organized into categories and subcategories, which allows to find necessary information easily. After logging in it is possible to add user's private data (paths or points of interests).

Users are navigated along the predefined paths or to the selected point (e.g. selected bus stop) using distance and direction information. Because the electronic device is equipped with an electronic compass, mobile phone can be kept in a pocket and the user can use the device for the orientation purposes. Keyboard allows to select application functions. Text to speech module is used to read messages. Additional feedback is provided by the vibration engine. POI Explorer can be also used without the device, in this case TalkBack screen reader is used for sonification of messages. The POI database can be managed using web application. Such solution is dedicated for sighted persons (for example someone from visually impaired person family) and allows to plan path or to manage private POIs.

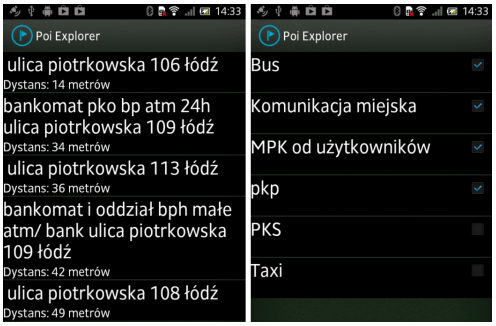


Fig. 2. Screenshots from the POI Explorer application captured by Sony Xperia S Android Phone. The list of the points of interests in the front of the user (direction is calculated based on compass values) are presented in a standard list box. On the right: list of subcategories of the category "Transport".

Public Transport Explorer application uses data served by passenger information system. Users can get timetables, current position of the selected vehicle, they can also plan their travel. As the passenger information system relies on actual GPS based positions of the public transport vehicles, the data presented to the user are up-to-date. It especially important if the user is inside the vehicle, where GPS signal can be very weak. This feature is also of a great value when the stops are moved to new, temporary locations, or when the routes or timetables of public transport vehicles temporarily change. As can be noticed, the graphical user interfaces of POI Explorer and Public Transport Explorer use large, high contrast characters aiding users with moderate visual impairment. System requirements of our applications allow to run them on low cost devices. Both applications have been consulted with the blind users from the Polish Association of the Blind.

B. NFC Enabled Passenger Information System Access

Near Field Communication (NFC) is designed to allow short-distance data exchange. It supports data exchange in two modes: passive mode, where only one device generates electromagnetic field (carrier) while the other device only modulates it. Moreover, modulating device might use power from electromagnetic field generated by another one, thus making one of the devices a transponder. The other mode is an active mode, where both devices generate EM fields alternately. Developed NFC enabled Passenger Information System makes use of the passive NFC tags. The tags are placed close to timetable boards at the bus or tram stops. Every tag stores a code that uniquely identifies a stop and redirects the user to a dynamically generated web page presenting the most up-to-date information on the bus or tram arrival times. The advantage of the electronic timetable is the use of real time data from passenger information system, therefore the web

Fig. 3. Screenshot of an application for NFC enabled passenger information system access.

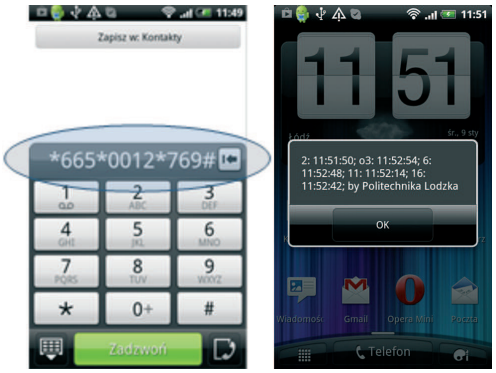


Fig. 4. USSD enabled passenger information system access.

page always displays actual bus or tram departure time as shown in Fig. 3. For this stateless service a dedicated proxy service has been developed. It uses Windows Communication Foundation (WCF) to provide RESTful API for passenger information system.

C. USSD Enabled Passenger Information System Access

Unstructured Supplementary Service Data (USSD) is a protocol used in GSM network for data exchange between mobile phone and Public Land Mobile Network (PLMN) infrastructure. With the use of API exposed in Telco 2.0 model it is possible to use USSD messages for communication between any GSM-enabled device to any Internet service. Proposed access method requires user-initiated USSD session. In the initial message (that is provided in the same manner as number user want call to) user has to provide unique USSD service prefix (assigned by PLMN operator) and a bus or tram stop number. When a user initiates the dialogue, a message is passed through internal PLMN components to Telco 2.0 WebGateway (exposing PLMN features to the Internet) and then to Passenger Information System application server. Information on next bus/tram departure time is then retrieved from the system via aforementioned proxy service and sent back to the user as shown in Fig. 4.

IV. Summary

Although many electronic travel aids have been developed so far, urban spaces and public transport services still remain hardly accessible to the visually impaired. Poor spatial orientation and inability to access textual information makes difficult to locate and identify bus or tram stops and to read public transport timetables or numbers of vehicles arriving at the stops. The system described in the paper is a part of a solution aiming at assisting the visually impaired in travelling with public transport. The passenger information system provides actual data on routes and timetables of the vehicles while proposed mobile applications makes that information accessible to the users equipped only with ordinary mobile phones.

Acknowledgment

This work was partially supported by the National Centre for Research and Development of Poland under grant no. NR-02 0083-10 in years 2010-2013.

References

[1] World Health Organization, http://www.who.int/mediacentre/ factsheets/fs282/en/, Accessed 19 May 2013.
[2] P. Strumiłło, "Electronic Interfaces Aiding the Visually Impaired in Environmental Access, Mobility and Navigation," *in Proc. 3rd International Conference on Human System Interaction,* Rzeszów, Poland, 2010, pp. 17–24.
[3] Talking Signs. http://www.talkingsigns.com/. Accessed 22 February 2013.
[4] S. Bohonos, A. Lee, A. Malik, C. Thai, R. Manduchi, "Universal Real-Time Navigational Assistance (URNA): An Urban Bluetooth Beacon for the Blind," *in Proc. 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environment,* New York, 2007, pp. 83–88.
[5] J. Marski, P. Bajurko, K. Radecki, and T. Buczkowski, "Miniaturowe radiolatarnie i terminale z sygnalizacją RSSI do wspomagania orientacji osób niewidomych," ("Miniature radio beacons and terminals with RSSI signaling to support the orientation of the blind") Telecommunication Review – Telecommunication News (Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne), vol. 6, 2010, pp. 320–323. (in Polish)
[6] PAVIP, http://bones.ch/, Accessed 20 May 2013.
[7] Step-Hear, http://www.step-hear.com/, Accessed 20 May 2013.
[8] P. Barański, M. Polańczyk, P. Strumiłło, "A Remote Guidance System for the Blind," *in Proc. 12th IEEE International Conference on e-Health Networking, Applications and Services HealthCom,* Lyon, France, 2010.
[9] NaviEye (Nawigator), http://www.migraf.pl/, Accessed 20 May 2013.
[10] On the Bus, http://www.onthebus-project.com/, Accessed 20 May 2013.
[11] Ł. Kamiński, K. Bruniecki, "Mobile Navigation System for Visually Impaired Users in the Urban Environment," Metrology and Measurement Systems, vol. XIX (2), pp. 245–256, 2012.
[12] Loadstone GPS - Free GPS Software for Your Mobile Phone, http://www.loadstone-gps.com/, Accessed 21 May 2013.
[13] P. Korbel, P. Skulimowski, and P. Wasilewski, "A Radio Network for Guidance and Public Transport Assistance of the Visually Impaired," *in Proc. 6th International Conference on Human System Interaction HSI 2013,* Sopot, Poland, 2013.
[14] VoiceOver, http://www.apple.com/accessibility/voiceover/, Accessed 21 May 2013.
[15] J. Blumenfeld, P. Poryzała, T. Woźniak, and P. Skulimowski, "Moduł czytnika systemu rozproszonej sieci znaczników radiowych wspomagający osoby niewidome w orientacji przestrzennej i w podróżowaniu w mieście," ("Navigation device for communication with distributed network of radio tags to assist the blind") Electronics – Constructions, Technologies, Applications (Elektronika – Konstrukcje, Technologie, Zastosowania), Vol. 10, 2012, pp. 73–75. (in Polish)
[16] M. Polańczyk, P. Skulimowski, B. Sujecki, and D. Sulmowski, "Personal Navigation System for the Blind based on Points of Interest," *in Proc. II Forum Innowacji Młodych Badaczy 2011,* Łódź, Poland, 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 829–836

Towards networks of the future: SDN paradigm introduction to PON networking for business applications

Paweł Parol
Orange Labs
Obrzezna 7, 02-691 Warsaw, Poland
Warsaw University of Technology
The Faculty of Electronics and Information Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
Email: Pawel.Parol@orange.com

Michał Pawłowski
Orange Labs
Obrzezna 7, 02-691 Warsaw, Poland
Email: Michal.Pawlowski1@orange.com

Abstract—The paper is devoted to consideration of an innovative access network dedicated to B2B (Business To Business) applications. We present a network design based on passive optical LAN architecture utilizing proven GPON technology. The major advantage of the solution is an introduction of SDN paradigm to PON networking. Thanks to such approach network configuration can be easily adapted to business customers' demands and needs that can change dynamically. The proposed solution provides a high level of service flexibility and supports sophisticated methods allowing user traffic forwarding in effective way within the considered architecture.

I. Introduction

IN RECENT years Internet traffic is skyrocketing (traffic growth is exponential) as users are consuming more and more Internet services (e.g. video or cloud based solutions). The problem is often highlighted by telecommunications providers but high growth is also observed by organizations like enterprises, universities, governmental entities. Thus many institutions have to adapt and bolster their traditional IT and network infrastructure in order to handle that phenomenon.

In this chapter the overview of legacy campus networks, typically used by institutions, is given. Also LAN (and WAN access) solutions provided to business customers are described.

A. Office networks overview

Nowadays access to Internet is prevalent among companies. Moreover many enterprises have own intranet. In order to provide connectivity to different devices likes PC, laptops or tablets in-building network infrastructure is needed. It can be composed of a single modem/router but also tens of devices and substantial amount of transmission media (optical fibers, twisted pair cables etc.). In case of big organizations all those components form a campus network– computer network interconnecting LANs (Local Area Networks) within a limited geographical area. The infrastructure is usually owned by campus owner / tenant e.g. enterprise, university, hospital.

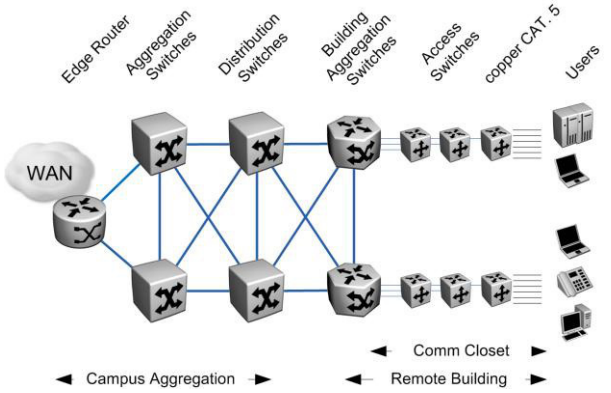


Fig. 1 Campus network hierarchy

Early LANs were large flat networks with peer to peer Layer 2 (L2) communication based on Ethernet [4]. It was a simple approach but with network growth ultimately led to disruptions (e.g. due to broadcast storms). Over the time Layer 3 (L3) has been introduced dividing campus network into smaller segments (allowing avoiding such problems). Additionally numbers of different solutions like VLANs (Virtual LANs [5]), RSTP (Rapid Spanning Tree Protocol [6]) or IP subnets have been developed making campus networks easier to maintain and manage.

Also the topology evolved towards more hierarchical and structured design. 4-tier network architecture (see Fig. 1) has become common ([8], [10]). In that approach access layer provides connectivity to end devices with copper twisted pairs (usually UTP CAT. 5 – Unshielded Twisted Pair Category 5, nowadays CAT. 6 cables are gaining popularity). Fast Ethernet or Gigabit Ethernet (100BASE-TX or 1000BASE-T [4]) are typically used. Access layer contains multiple L2 switches. They are located nearby users, e.g. in communication closets on each floor of the building. At the next level additional switches aggregate traffic for each building (concentrating multiple access layer switches) Interconnection between access layer switches and building aggregation switches can be provided with copper cables, or with fibers. Building aggregation switches are connected to

978-83-60810-53-8/$25.00

campus aggregation and distribution switches which then connect to router (being a gateway to external networks). As a transmission medium for interconnecting building aggregation with campus aggregation segments fiber optic cables are often used due to higher bandwidth requirements (i.e. 10 GbE interfaces) and distance.

Tiered network design gives flexibility in terms of supporting numerous functions and end devices (for example growth of client population can be accommodated by adding access layer switches, but that approach is costly). The logical division for different layers does not need to be done with physical tiers; access and aggregation can be provided on the same equipment. It can be especially useful in case of smaller campus simplifying management of reduced number of devices [9].

Important to note is fact that legacy Ethernet-based campus networks have significant drawbacks. Maximum length of copper Ethernet cables is limited to 100 meters. In fact 4-tier topology with switches on each floor is an answer to that limitation. Ethernet LAN requires a cable connection to every single user port. This means significant number of access layer switches and wires (copper cables) and at the end results in high costs. High-frequency signals (used for Fast and Gigabit Ethernet) require more sophisticated copper cable constructions which are physically larger than for lower frequencies (necessary to avoid signal disturbances). In consequence the space required for racks, communication closets is large. Crucial amount of heat is produced, power consumption is high. Management of high number of active devices is not easy.

For those reasons legacy Ethernet LAN is not always the best answer for campus network requirements. That is why an important issue is to find a more effective approach for office networks infrastructure.

B. Scenarios for B2B services

B2B (Business to Business) telecommunications services' landscape is diverse. It includes services like Internet access, POTS (Plain Old Telephony Service), VoIP (Voice over IP), dedicated links, VPN (Virtual Private Network), etc. One can distinguish large (Enterprise), medium (SME – Small and Medium Enterprises) and small (SOHO – Small Office Home Office) market segments. However service overlapping (the same services) is possible, but often there are special offers for different segments.

Services' requirements largely depend on type of customer. Big entity owning campus network (and considerable number of network equipment) has other needs than company with small branches scattered around the country (and with lack of its own interconnection) and than small company located in single office building.

For entity with campus network usually telecommunications operator provides its services to location where campus edge router is placed, further propagation is the responsibility of the entity itself (compare Fig. 1). In the second case (several branches) it is important to provide interconnection among branches.

For office building, in which many companies are located, there are two most common infrastructure scenarios (see Fig. 2). First one is based on existing copper CAT. 3 cables which reach customers' desk / office and can be reused by telcos. Modem or router is the termination point of the services (Fig. 2 Scenario A).

In the second scenario office building has infrastructure based on active Ethernet LAN with copper cables CAT. 5 (Fig. 2 Scenario B). Telecommunications operators need to provide its interconnecting cables up to building's technology room. Separation of services / between different operators can be provided on logical level e.g. by means of VLANs.

For both scenarios the only responsibility of telco is to somehow access business customers. Herein, one could think of a new role for operators targeting office buildings environment: what added values are possible to be identified

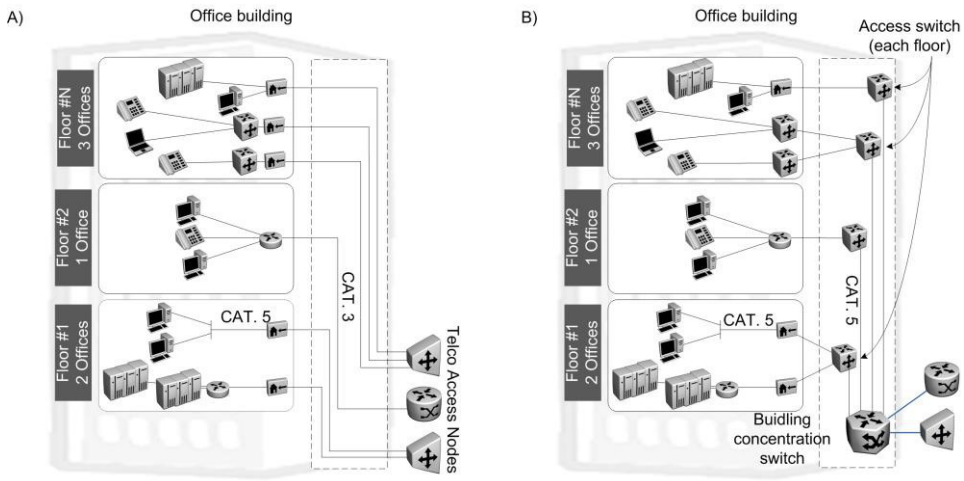


Fig. 2 Legacy infrastructure in office buildings. Scenario A: Telecommunication operators' cables CAT. 3 up to the office. Scenario B: In-building infrastructure based on Active Ethernet LAN (copper cables CAT. 5)

if telcos take the responsibility of building and administrating the entire in-building office network?

II. OPTICAL LAN

Optical LAN is a new approach for office networks infrastructure and an answer to limitations of legacy Ethernet LANs. All-fiber LAN interconnecting existing Ethernet end devices allows reducing costs and making the network more reliable.

Proposed solution is based on GPON (Gigabit Passive Optical Network) [1]. It is standardized, well known and widely adopted telecommunications access technology, used by many operators worldwide. GPON uses point-to-multipoint topology and employ fiber optics as a transmission medium. As a real passive solution – no active equipment is used in-between GPON Access Node: OLT (Optical Line Termination) and line termination at customer side: ONT (Optical Network Termination). In campus network based on Optical LAN number of active equipment is significantly reduced comparing to traditional LAN scenario. From OLT GPON port a single strand of fiber goes out to a passive optical splitter(s) which splits the signal onto fibers terminating at up to 64 (or even 128) ONTs (see Fig. 3). All the fibers, splitters connected to one GPON port on OLT form a GPON tree. ONT device terminates GPON transmission and provides 10/100/1000-BaseT Ethernet connectivity to desktop equipment such as PC computers, laptops, voice over IP phones, and video phones using regular copper patchcords (or by 802.11 WiFi). ONT can be located on customer's desk (ONT per desk) or in office closet (ONT per office). Those two options are called respectively: Fiber-to-the-Desktop (FTTD) or a Fiber-to-the-Communications (FTTC) room. High flexibility of Optical LAN solution allows reusing existing copper infrastructure in buildings (for example GPON access is terminated on ONT located in the floor communication closet, from where existing copper cables are used up to customer's desk, see Fig. 3 – Floor #2).

Thanks to fiber optics-based transmission Optical LAN is a long reach access solution – maximum reach is equal to 20 km in a standard mode. It is a tremendous improvement comparing to traditional copper Ethernet (100 m.). It allows placing OLT in distant locations, giving high flexibility in network design (in case of campus network OLT no longer need to be installed in the same building in which customers reside).

GPON technology assures 2.488 Gbps of downstream bandwidth and 1.244 Gbps of upstream bandwidth. Bandwidth is shared among customers connected to the same GPON tree. Advanced GPON QoS mechanisms assure appropriate bandwidth distribution among many users and different applications.

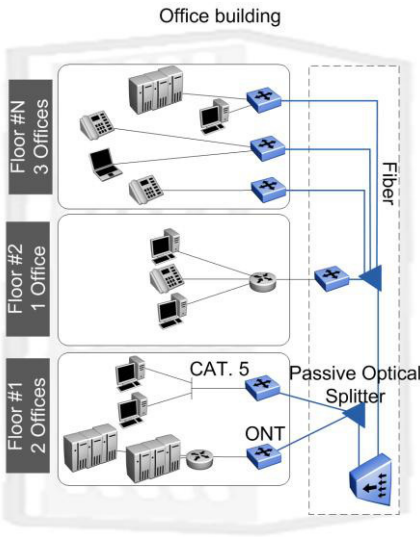


Fig. 3 Office building infrastructure based on Optical LAN

Optical LAN solutions are present in portfolio of several vendors (e.g. Motorola [11], Tellabs [12], Zhone [13]). According to vendors estimations introducing of Optical LAN will reduce power consumption by up to 65%, space requirements by up to 90%, capital costs related to network elements by up to 74% [13]. Optical LAN is seen as a new paradigm in campus networking allowing optimization of investments and at the same time improving overall efficiency of the network.

III. A NOVEL APPROACH TO B2B

In this chapter we formulate three postulates, which are, from our perspective, crucial for deploying future proof access networks for business applications:

1. Applying Optical LAN concept

Currently Optical LAN vendors target big entities with large campus networks. In typical deployment Optical LAN is used by only one organization – the owner and the administrator of the campus network. Office buildings with many tenants, each of them having its own LAN network (at least up to some point) are not yet addressed.

In this paper we propose a solution to that deficiency. It is based on concept known from telecommunications world where many customers are connected to the same Access Node (different users served on the same equipment). In our proposition enterprises no longer need to operate any active network equipment or to build networks itself. LAN becomes a service, provided in similar fashion as e.g. Internet access. LAN service provider is responsible for service creation, administration and adjustment according to needs of customers (enterprises using LAN). That also means that network infrastructure is built for offices by LAN service provider. In fact such network is similar to GPON access networks used by telecommunications operators to provide services to its customers. For B2B scenario different customers are also served by the same GPON OLT unit.

2. A new role for telecommunications operators

Telecommunications operators are well positioned to play the role of Optical LAN service providers. Usually they have necessary experience with GPON technology, operational resources and existing access network. Telcos are able to deploy optical fiber LAN in office buildings and to provide flexibility in management, service creation and administration.

Such approach has many advantages in terms of optimal usage of network resources. Single OLT can be used for several buildings, even if they are located in distant areas (due to long reach offered by GPON technology which capabilities in terms of maximum physical reach are not fully used in current optical LAN implementations). Also interconnection of distant branches becomes easier (in specific cases they can be served by the same OLT). Additionally a new type of services can be introduced called Office LAN services: e.g. on-demand LAN connections between companies located in the same building, access to in-building monitoring system, etc.

This novel approach also creates a new business model for telecommunications companies who become Optical LAN operator (builder and administrator). This opportunity to find new B2B market seems to be a good argument in convincing telco players to work on such solutions.

3. Business-user-oriented access network design

Another assumption for the presented approach is that it is based on user-oriented access network design. Service portfolio dedicated to business customers is typically more complex than the one for residential users. For business applications customized services need to be taken into account. Moreover, customer demands can change dynamically over short periods of time. That is why a challenge for networks deployed in business environments is to provide a high level of service flexibility and to forward user traffic in effective way. To meet those requirements we present in this paper an access network architecture based on SDN (Software-Defined Networking) paradigm which assumes data plane and control plane abstractions separation ([17]). Thanks to such approach network devices become programmable units. In practice it means that network configuration can be easily adapted to the fast-changing needs.

IV. SDN-BASED GPON SOLUTION FOR BUSINESS APPLICATIONS

In order to introduce SDN paradigm to GPONs area one can propose different methods to accomplish that. One of the possible ways would be to develop a brand new protocol allowing GPON devices to become programmable units. Such approach is supposed to be an appropriate one for designing an optimal logical architecture of OLT and ONT in the scope of data processing and forwarding. However, development of generic SDN-based protocol for GPON would require a lot of standardization efforts and probably it would take a few years to obtain a solution being ready for deployment. Moreover, it would be limited to GPON technology and thus it could not be applied for other network types and applications.

In this paper we present another approach. We propose a solution based on OpenFlow ([16]) which is the most widely deployed SDN-based protocol. OpenFlow Switch architecture consists of at least three parts ([15]) – see Fig. 4:

- Flow Table(s) – a structure within switch implementation with an associated actions with each flow entry; the Flow Tables define the ways of how the traffic flows have to be processed by the switch
- Controller – an external unit running a remote control process that manages the switch via the OpenFlow protocol; the Controller can add, remove and update flow entries from the Flow Table(s)
- Secure Channel (also called OpenFlow Channel) – a channel which enables a communication (i.e. sending packets and commands) between the Controller and the switch

For a more detailed description of OpenFlow-specific logical components and functions please refer to [15].

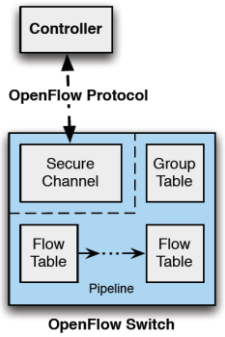


Fig. 4 OpenFlow Switch logical scheme (source: [17])

OpenFlow was originally designed for L2/L3 switches (or routers) equipped with native Ethernet-based physical interfaces. That is why it is important to notice that it is useless to implement pure OpenFlow in GPON OLTs and ONTs. The reason for that is simple: although GPON effectively carries Ethernet frames, in practice it operates at Layer 1 (according the OSI model) with its own dedicated framing and GEM (GPON Encapsulation Method) protocol used for encapsulation higher-layer Protocol Data Units (e.g. Ethernet frames) into GTC (GPON Transmission Convergence) layer. The current specification of Open Flow protocol does not support such kind of non-Ethernet-based physical interfaces. That is why some additional GPON-related functions have to be introduced to OpenFlow.

A. SDN-based protocol for GPON

A single logical connection within the GPON system is called GEM Port and it is identified by GEM Port-ID. A GEM Port can be considered as a channel within GTC layer and is capable to transport one or more traffic flows. In the

upstream direction GPON system also utilizes T-CONTs (Transmission Containers) corresponding to allocated timeslots within TDMA multiplexing existing in GPON. Each T-CONT represents a group of logical connections (GEM Ports) that appear as a single entity for the purpose of upstream bandwidth assignment on the PON (see Fig. 5 – GPON-specific traffic entities identifiers are pointed in brackets).

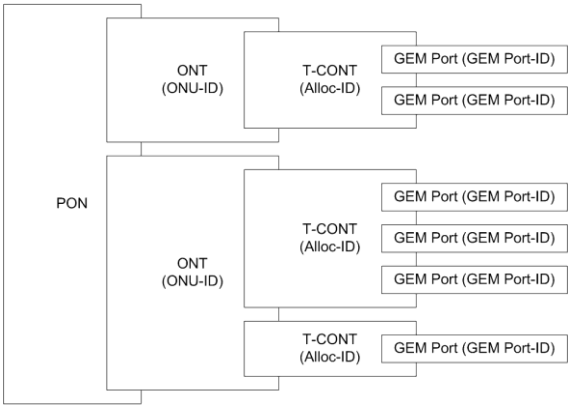


Fig. 5 Upstream multiplexing in GPON system

Each T-CONT can be seen as an instance of upstream queue with a certain bandwidth profile (a set of bandwidth parameters). The bandwidth assignment model applied in GPON system effectively introduces a strict priority hierarchy of the assigned bandwidth components ([2]):

- fixed bandwidth: with highest priority
- assured bandwidth
- non-assured bandwidth
- best-effort bandwidth: with lowest priority

Five T-CONT types are defined by [2]. Depending on the traffic type (latency-sensitive traffic, data transmission, etc.) the most appropriate T-CONT type should be selected to carry considered traffic flows.

Upstream user traffic (Ethernet frames) is encapsulated into GEM Ports and then into T-CONTs. Each GPON ONT uses its own set of T-CONTs and GEM ports, a unique one within a GPON tree which ONT belongs to. A single GEM Port can be encapsulated into only one T-CONT, however a single T-CONT may encapsulate multiple GEM ports. In downstream direction only GEM Ports are used to carry traffic flows since no TDMA multiplexing exists there and thus the notion of T-CONT is not relevant for GPON downstream transmission. For a more detailed explanation please refer to [2].

One of the key aspects of GPON-based network applications is to ensure effective traffic forwarding on the GTC layer. In order to do that it is important to define appropriate rules (consistent and unambiguous ones) allowing to map traffic flows incoming from users to appropriate GEM Ports. In most of commercial implementations mapping rules “built-in” GPON ONTs are mono-criterion- i.e. mapping is based on only one of the following criteria like: VLAN ID (Virtual LAN identifier), p-bit ([5]) or UNI (user port number on ONT). For some cases also double-criterion combinations of aforementioned parameters are available (e.g. VLAN ID + UNI) for the mapping purpose. Since GPON was originally designed for B2C (Business to Customer) market segment for which only Triple-Play (Internet, ToIP and IPTV) services are considered such approach was sufficient. For business applications where not only service portfolio is more complex but also customized services are taken into account, much more sophisticated methods (i.e. mapping rules) are required in order to ensure effective traffic forwarding through the system ([14]). In most scenarios currently deployed GPON ONT with limited set of hardcoded mapping and forwarding functions would not be able to address such needs. In such cases software upgrade is needed but it leads to higher operational costs - especially if business customer demands changes dynamically and it is possible that new set of functions is required. For such a scenario multiple software upgrades have to be taken into account.

The solution for the issue is SDN-based protocol for GPON allowing OLT and ONT to become programmable units. In this paper we propose OpenFlow-based solution. As mentioned before the current specification of OpenFlow protocol does not support GPON natively. That is why our vision is to introduce GPON-related functions to the specification in order to develop a protocol extension which we called OpenFlowPLUS.

The main assumption for the OpenFlowPLUS is that it inherits all the functionality, architecture and capabilities of original OpenFlow. The essential improvement is an introduction of GPON-related functions to the protocol in terms of traffic forwarding in order to make the solution relevant also for GPON technology.

According to OpenFlow Switch architectural assumptions each device (OLT, ONT) within considered GPON tree contains Flow Table(s) and communicates over a Secure Channel with remote Controller via OpenFlowPLUS protocol (see Fig. 6).

For that purpose OLT and ONTs are supposed to have IP address configured. Since Controller and OLT are assumed to be connected to IP/Ethernet network they can establish L3 connection. ONTs are accessible by Controller only via OLT. One could take advantage of that and for the purpose of OpenFlowPLUS messages exchange between ONTs and Controller make use of GPON-specific mechanisms defined by [3]. In such a scenario protocol messages are transported through the PON via a dedicated OMCI (ONT Management and Control Interface) channel towards OLT and then they are sent directly to the Controller using OLT’s Secure Channel. Obviously, employing OMCI by OpenFlowPLUS for some new applications does not mean that the protocol takes the control over the entire GPON system. All functions which are out of the scope of traffic mapping and forwarding

(e.g. ONT discovery and provisioning-related functions, Dynamic Bandwidth Allocation mechanism, optical layer supervision, alarms and performance monitoring etc.) are assumed to be realized in traditional way, i.e. in line with recommendations defined in [2] and [3]. That is why an optimal approach seems to be adding OpenFlowPLUS controller as a functional module to the standard EMS (Element Management System) managing the system.

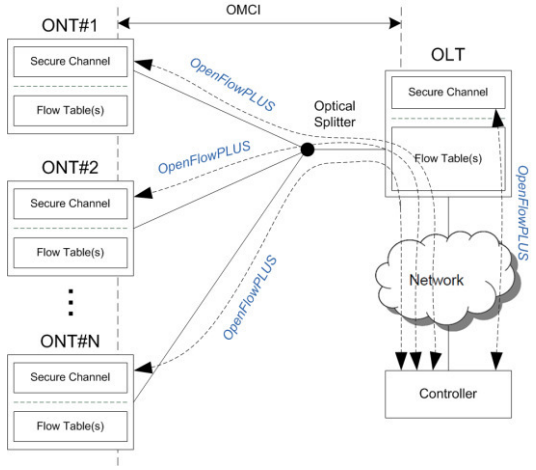


Fig. 6 OpenFlowPLUS–based GPON solution overview

As mentioned before the idea of OpenFlowsPLUS is to provide GPON-related functions to the protocol in terms of traffic mapping and forwarding. Similarly to original OpenFlow, OpenFlowPLUS is assumed to use Flow Table(s) which perform packet lookups, modification and forwarding. Each Flow Table contains multiple flow entries. Each flow entry contains:

- match fields – to match against packets; match fields include packet header fields (e.g. VLAN ID, MPLS label, IP destination address, TCP source port, etc.), an ingress port and metadata that pass information between tables; flow entries match packets in priority order, with the first matching entry in each table being used ([15])
- counters – which can be maintained for each port, table, flow, etc.
- instructions – operations which are executed when a packet matches a flow entry

Instructions define the ways of how single action is processed. Actions represent operations of packet modification or forwarding to the specified port. Actions are grouped by different action types, for instance pop action type (e.g. pop VLAN header action), set action type (e.g. set MPLS traffic class action), etc. Instructions executed during OpenFlowPLUS pipeline processing can either add appropriate actions to the current action set (a set of actions that are accumulated when the packet is processed by the tables and that are executed after exiting the processing pipeline by the packet), or force some actions to be applied immediately. OpenFlowPLUS defines new actions which are relevant to GPON technology. The considered functions are presented in Table I.

TABLE I.
MAIN GPON-RELATED FORWARDING FUNCTIONS PROVIDED BY *OPENFLOWPLUS*

GPON unit	Action type /Action	Remarks
ONT, OLT	gpon: Map to GEM Port	introduction of a new action to the original OpenFlow action set function: mapping Ethernet frames to particular GEM Port instance
ONT	gpon; Map to T-CONT	introduction of a new action to the original OpenFlow action set function: mapping GEM Ports to particular T-CONT instance
ONT, OLT	output	action modification when executed for GPON interfaces new function: GTC framing before forwarding the packet on the GPON port

OpenFlowPLUS introduces a brand new action type called gpon related to GPON-specific mapping methods. The considered action type provides two actions: Map to GEM Port action which represents an operation of mapping Ethernet frames to particular GEM Port instance and Map to T-CONT action which represents an operation of mapping GEM Ports to particular T-CONT instance. Additionally the new functionality for original OpenFlow output action is supposed to be supported: when a packet is destined to be forwarded to the GPON port, GTC framing is performed for the packet before exiting the interface. The aforementioned protocol improvements are the main GPON-related forwarding and mapping functions provided by OpenFlowPLUS.

B. Use case

In this section we present a possible application for SDN-based GPON concept which is proposed in the paper. The following assumptions are made for the considered use case:

- the solution is dedicated to business customers who reside in office buildings
- operator acts not only as a service provider but is responsible also for administration of in-building network
- in-building network is based on Optical LAN solution
- different service types can be offered: Internet access, Metro Ethernet (corporate connections), cloud computing-based services, office LAN services; the considered traffic flows are listed in Table II
- access network architecture is based on OpenFlowPLUS GPON solution (see Fig. 7)

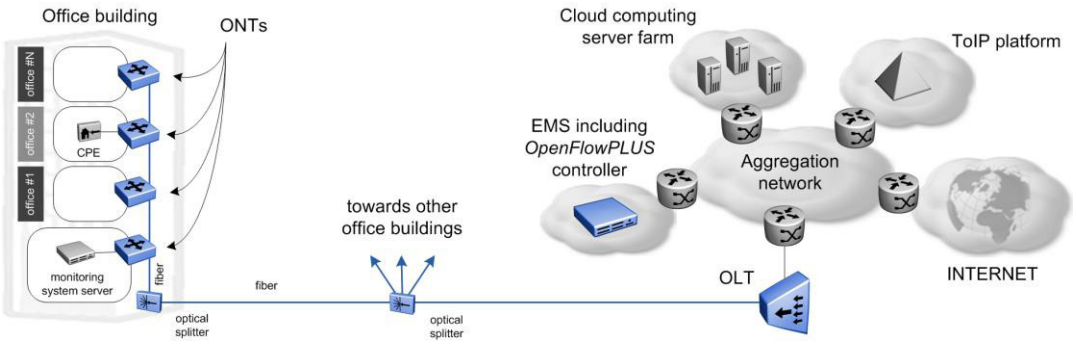


Fig. 7 OpenFlowPLUS-based architecture for business applications overview

TABLE II.
SERVICES AND TRAFFIC FLOWS OVERVIEW

Flow ID	Traffic flow/service	Remarks
F#1	Internet Access: HTTP, FTP, etc.	standard Internet services
F#2	Internet Access: web-based application hosting	connections from Internet are established using HTTPS (SSL + HTTP) protocol (TCP port 443)
F#3	Internet Access: remote access to intelligent installation system controller	connections from Internet to intelligent installation system controller physically located in the office are based on KNXnet/IP protocol (port 3671); IP address of the controller: IP@1.3
F#4	Metro Ethernet: connections to remote company branch	remote company branch is supposed to use IP@2.X address pool;
F#5	ToIP (telephony over IP)	IP phones used in the office are assumed to mark IP ToS field with DSCP "EF" value; IP address of ToIP platform: IP@4.1
F#6	Office LAN: on-demand connections to different companies located in the same building	connections allowed for a designated sub-pools of addresses from IP@1.X and IP@priv (office) and IP@3.X (different company office)
F#7	Office LAN: access to in-building monitoring systems	in-building monitoring system server is assumed to be connected to a dedicated ONT with IP address: IP@5.1
F#8	Cloud computing: remote storage, backups	IP address of cloud computing server: IP@6.1

Each office in the building is connected to the optical network via ONT. Copper cables terminated with RJ-45 sockets are deployed in office rooms. Each user device (PC, IP phone, application server, etc.) is connected to one of multiple Ethernet LAN ports which ONT is equipped with (see Fig. 8). ONT aggregates the entire traffic incoming from user terminals (this traffic contains no VLAN tags) and provides the functionality of L3 gateway. Public IPv4 addresses are assigned to ONT (IP@1.1) and to some selected user devices: web-based application server – IP@1.2 and intelligent installation system controller – IP@1.3. For other devices private addressing is used (IP@priv). ONT is assumed to act as an internal DHCP server for that purpose. Any kind of additional CPE (Customer Premise Equipment) is not required in the considered network.

Thanks to applying OpenFlowPLUS sophisticated mapping rules are supported in order to ensure effective traffic forwarding through the GPON. As an example we show how flow entries match fields with corresponding instructions can be defined within ONT Flow Table for traffic flows transmitted in upstream direction (see Table III).

Based on the user traffic to GPON-specific instances mapping methods (using limited set of parameters like VLAN ID, pbit, UNI) which are currently supported in typical commercial implementations it would be very difficult or even impossible to follow the traffic forwarding model presented in considered use case. For instance, in traditional approach it would be impossible to map traffic flows F#1 and F#3 to different GEM Ports if they originated from the same end-device. The proposed solution is very flexible and convenient from customer perspective. Since OpenFlowPLUS-based GPON is a programmable system its configuration can be easily adapted to support new services and business needs when they appear.

Presented access network model supports also openness for alternative operators what is typically required by country-specific regulations. Each customer served by another operator connects a dedicated CPE to the ONT which is configured as a bridge that passes assigned VLAN(s) through the system up to the first alternative operator's switch or router (see "office #2" in Fig. 7).

V. CONCLUSION

In the paper we presented a novel approach to deploy optical access networks addressed to B2B market segment where we defined a new role for telcos. The major advantage of our solution is its flexibility thanks to introduction of SDN paradigm to GPON-based networking. We believe our work

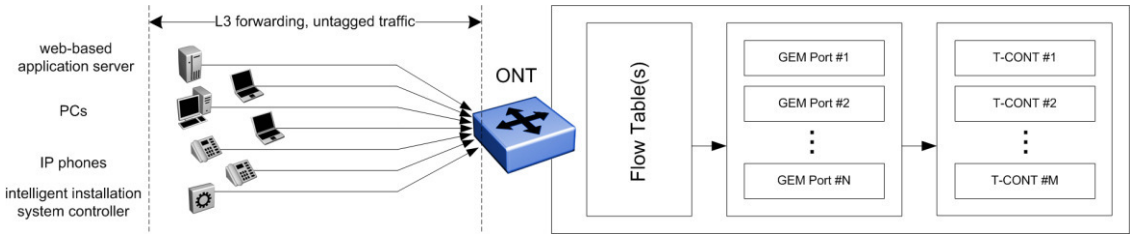


Fig. 8 OpenFlowPLUS-based ONT logical scheme

can be considered as a conceptual framework for further analysis and solution development.

TABLE III.
MATCH FIELDS AND INSTRUCTIONS FOR FLOWS INCOMING TO ONT

match fields of flow entries (in priority order)	Matched flow	Instructions
IPv4 dst = IP@4.1 AND IPv4 ToS bits = EF	F#5	Apply-Actions { Map to GEM Port: 1 Map to T-CONT: 1 (type 1)}
IPv4 dst = IP@2.X	F#4	Apply-Actions { Push VLAN header Set VLAN ID: 1001 Set VLAN priority: 3 Map to GEM Port: 2 Map to T-CONT: 2 (type 2)}
IPv4 dst = IP@3.X	F#6	Apply-Actions { Push VLAN header Set VLAN ID: 301 Set VLAN priority: 3 Map to GEM Port: 3 Map to T-CONT: 3 (type 2)}
IPv4 src = IP@1.2 AND TCP src port = 443	F#2	Apply-Actions { Set IPv4 ToS bits = CS2 Map to GEM Port: 4 Map to T-CONT: 4 (type 3)}
IPv4 src = IP@1.3 AND TCP src port = 3671	F#3	Apply-Actions { Set IPv4 ToS bits = CS1 Map to GEM Port: 5 Map to T-CONT: 4 (type 3)}
IPv4 dst = IP@5.1	F#7	Apply-Actions { Push VLAN header Set VLAN ID: 200 Set VLAN priority: 1 Map to GEM Port: 6 Map to T-CONT: 5 (type 2)}
IPv4 dst = IP@6.1	F#8	Apply-Actions { Set IPv4 ToS bits = CS2 Map to GEM Port: 7 Map to T-CONT: 6 (type 2)}
IPv4 dst != {IP@1.X, IP@priv, IP@2.X, IP@3.X, IP@4.1, IP@5.1, IP@6.1}	F#1	Apply-Actions { Set IPv4 ToS bits = none Map to GEM Port: 8 Map to T-CONT: 4 (type 3)}

REFERENCES

[1] ITU-T Gigabit-capable Passive Optical Networks (GPON) : General characteristic, ITU-T G.984.1, 2008.

[2] ITU-T Gigabit-capable Passive Optical Networks (GPON) : Transmission convergence layer specification, ITU-T G.984.3, 2008.

[3] ITU-T Gigabit-capable Passive Optical Networks (GPON) : ONT management and control interface specification, ITU-T G.984.4, 2008.

[4] IEEE Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specification, IEEE Standard 802.3-2008, 2008.

[5] IEEE Standard for Local and metropolitan area networks – Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks, IEEE Standard 802.1Q-2011, 2011.

[6] IEEE Standard for Local and metropolitan area networks – Media Access Control (MAC) Bridges, IEEE Standard 802.1D-2004, 2004.

[7] FTTH Council Europe, FTTH Handbook, Edition 5, 2012.

[8] Cisco Systems. (1999). White Paper, Gigabit Campus Network Design – Principles and Architecture [Online]. Available: http://www.cisco.com/warp/public/cc/so/neso/lnso/cpso/gcnd_wp.pdf

[9] Brocade. Designing a Robust and Cost-Effective Campus Network [Online]. Available: http://www.brocade.com/downloads/documents/design-guides/robust-cost-effective-lan.pdf

[10] S. T. Karris, Networks Design and Management. 2nd ed. Orchard Publications, 2009.

[11] Motorola. (2012). White Paper, Creating Simple, Secure, Scalable Enterprise Networks using Passive Optical LAN [Online]. Available: http://moto.arrisi.com/staticfiles/Video-Solutions/Solutions/Enterprise/Passive-Optical-LAN/_Documents/_Staticfiles/WP_POL_CreatingNetworks_365-095-20298-x.1.pdf

[12] Tellabs. (2011). How Enterprises Are Solving Evolving Network Challenges with Optical LAN [Online]. Available: http://www.tellabs.com/solutions/opticallan/tlab_solve-net-challenges-with-optical-lan_an.pdf

[13] Zhone. FiberLAN Optical LAN Solution [Online]. Available: www.zhone.com/solutions/docs/zhone_fiberlan_solution.pdf

[14] P.Parol and M.Pawlowski, "How to build a flexible and cost-effective high-speed access network based on FTTB+LAN architecture," in Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.655,662, 9-12 Sept. 2012

[15] The OpenFlow Switch Specification. [Online]. Available: http://OpenFlowSwitch.org.

[16] N.McKeown, T.Anderson, H.Balakrishnan, G.Parulkar, L.Peterson, J.Rexford, S.Shenker, and J.Turner, "OpenFlow: enabling innovation in campus networks". SIGCOMM Comput. Commun. Rev. 38, 2, pp. 69-74, March 2008

[17] Open Networking Foundation. (2012). White Paper, Software-Defined Networking: The New Norm for Networks [Online]. Available: https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 837–844

Are Graphical Authentication Mechanisms As Strong As Passwords?

Karen Renaud*, Peter Mayer†, Melanie Volkamer† and Joseph Maguire*
*School of Computing Science, University of Glasgow
†Center for Advanced Security Research Darmstadt, Technische Universität Darmstadt
E-mail: karen.renaud@glasgow.ac.uk, {peter.mayer, melanie.volkamer}@cased.de

***Abstract*—The fact that users struggle to keep up with all their (textual) passwords is no secret. Thus, one could argue that the textual password needs to be replaced. One alternative is graphical authentication. A wide range of graphical mechanisms have been proposed in the research literature. Yet, the industry has not embraced these alternatives. We use nowadays (textual) passwords several times a day to mediate access to protected resources and to ensure that accountability is facilitated. Consequently, the main aspect of interest to decision-makers is the strength of an authentication mechanism to resist intrusion attempts. Yet, researchers proposing alternative mechanisms have primarily focused on the users' need for superior usability while the strength of the mechanisms often remains unknown to the decision makers. In this paper we describe a range of graphical authentication mechanisms and consider how much strength they exhibit, in comparison to the textual password. As basic criteria for this comparison, we use the standard guessability, observability and recordability metrics proposed by De Angeli *et al.* in 2005. The intention of this paper is to provide a better understanding of the potential for graphical mechanisms to be equal to, or superior to, the password in terms of meeting its most basic requirement namely resisting intrusion attempts.**

I. Introduction

ONE OF the most basic, everyday tasks of computer usage is authentication. Every user will, sooner or later, have to authenticate themselves. Their ability to do this effectively will impact on their ability to do their daily jobs and on their personal lives. The failure of the mechanism to resist intrusions will potentially have an impact on the user personally (e.g., in terms of ID theft or financial losses) or in professional environments on the organisation he or she works for.

Nowadays, the most widely used authentication mechanism is the textual password. However, it is well known, that most users are frustrated by their experiences with these traditional passwords in general [1]. Even if they want to behave securely, they often do not understand what constitutes a "secure" password since guidelines for the creation of secure passwords are seldom adequate [2]. Even with good guidelines in place, human nature will lead users to prefer the path of least resistance e.g. choosing weak passwords, writing them down, storing them in plain text on their mobile phones or reusing them [3], [4]. This is understandable considering the findings of Ives, Walsh and Schneider [5]: users are expected to recall an average of 15 different passwords on a daily basis. Due to human cognitive limitations, four or five is normally the maximum a typical user can handle [3].

Password managers can help users to manage an unlimited number of passwords. However, they constitute a single point of failure and systems cannot be easily accessed from a device that does not have the manager installed. Thus, password managers are no substitution for a secure and usable authentication solution [6]. The same holds for single sign on solutions.

To address the human inability to deal with large amounts of passwords, a new type of authentication system was conceived. The graphical password, first proposed by Blonder [7], required the person to verify their identity by clicking on positions within a picture. This is called a *locimetric* system. Other common types are *searchmetric* (pick a picture from a grid of images) and *drawmetric* (draw your secret) [8]. The most important motivator behind the use of a graphical authentication mechanism is that their memorability is superior to that of textual passwords. In the first place, there is what is called a "picture superiority effect", as described by Paivio [9]. Paivio explained that pictures were encoded using a dual mechanism. So, a password, being textual has only one route whereby the human can reach it. If that route decays, and is forgotten, the password cannot be accessed. If the memory item is visual, there will potentially be multiple routes to access it, and the decay of one access route does not render the item unreachable.

Numerous studies regarding the usability of graphical authentication schemes have been conducted. Yet, many of these sweep security concerns aside or deal with them in a desultory fashion [10]. However, very few graphical mechanisms are used in practice. Notable exceptions are the Windows 8 picture password and the Android lock-screen pattern. A number of reasons could be advanced. In this paper we will consider the elephant in the room: do these mechanisms provide the basic requirement namely resisting intrusion attempts at an equal or higher level than textual passwords do? In order to answer this question, we need to be clear about exactly what variation of the amorphous password we consider, since this impacts the resulting security-level. Hence, we consider textual passwords with a length of at least 8 characters and which are used in a system with a three-times-lockout technique.

We will use the different categories of attacks proposed by De Angeli, Coventry, Johnson and Renaud [11] in 2005 as a starting point. In each case the mechanism will be compared to textual passwords. The remainder of this paper is structured as follows. First, we present the three types of attacks we

978-83-60810-53-8/$25.00

use to evaluate whether the same security level as textual passwords are provided, and show how the most common attacks fit into this framework. Next, we describe the general ideas behind the different classes of graphical authentication mechanisms. Then, we report on efforts that have been made to strengthen these mechanisms. We then compare the security properties of graphical authentication schemes to the security offered by textual passwords. Finally, we summarise, discuss and conclude.

II. Vulnerabilities

To evaluate the security of different authentication mechanisms, the resistance against certain attacks is reviewed. Table I classifies the set of attacks proposed in [12] into the evaluation categories which we use throughout this paper, namely Guessability, Observability, Recordability, and Memorability; which are explained in the following paragraphs.

TABLE I
Common Attack Types

Vulnerability	Attack Type	Tool
Guessability	Brute Force Attack	Offline
	Dictionary Attack	On & Offline
Observability	Shoulder Surfing	Human Observer
	Spyware	Technology
Recordability	Social Engineering	Deception
	Theft	Unsecured Record
Memorability	Forgetting	Coping Techniques

Guessability: Brute force and dictionary attacks are the two types that have to be considered in this category. For dictionary attacks to be possible, passwords have to be predictable so that an attacker can create a dictionary with the most widely used passwords. Obviously having personal knowledge about the user can make it even more predictable. These attacks can in particular be carried out if the database or password file is obtained (without in our setting the account will be blocked after three trials). In this case the attacker can perform an offline attack to test all possible combinations of the password-space and thus has no limitations regarding aspects like a lockout policy etc. One way to resist this kind of attack is to impose password strength requirements when a password is chosen. Another way is for the system to issue strong passwords. This ensures that passwords are unpredictable and that all possible elements of the password space are evenly likely to occur. To be resistant against offline attacks, the password space has to be sufficiently large, where values of 2^{20} to 2^{28} seem to be commonly used on the Internet [13]. To be secure, the password space should be greater than 2^{80}, which is the lowest security strength NIST allows for government applications [14].

Observability: Observing the user while authenticating can be performed either by a human (shoulder surfing), by a human with technical equipment (filming the action of authentication) or using technical means (e.g. spyware). The goal is to collect information that allows an attacker to reproduce the authentication with an as high as possible probability. Shoulder surfing is performed by observing the user while he is authenticating himself with the given implementation. Lately, the ubiquity of mobile phones means that a camera could be used to capture the user's authentication secret without their knowledge. Spyware or malware in general can be installed on the user's system by the attacker. It can monitor input peripherals or obtain screenshots during authentication.

Resisting this kind of attack is challenging. One can resist this to a certain extent by making it necessary for an attacker to capture multiple authentication attempts before they obtain the full authentication secret [15].

Recordability: Attacks exploiting the recordability of an authentication mechanism are always performed through the human factor. The first way a password can leak to an adversary is that the user records their password in some way and it is then stolen (theft). The second way is for the person to be fooled into disclosing their secret to another person (social engineering). Social engineering includes all attacks, which do not target the system as such, but the user.

Both attacks rely on the relative ease with which a user can record his/her authentication secret. An implementation is resistant if it is hard for the secret to be recorded or disclosed. In the era of ubiquitous mobile phones with built-in cameras it is very difficult to resist this kind of vulnerability.

Memorability: While this is not an attack type, it can be a vulnerability if attackers exploit the consequences of a user's coping strategies [3]; e.g. writing down has an impact on observability as someone might not observe the user authenticating but the note with the information about the password. Similar re-usage has an impact on the dictionary attack. Note, with respect to memorability it is important to consider also situations where one might only authenticate once in a while, like once in a semester for students to register or once in a year to file a tax return.

III. Graphical Authentication

In general, graphical authentication works like any other knowledge-based authentication mechanism. The user has to verify knowledge of a secret he or she shares with the system. Contrary to the abstract nature of textual passwords, graphical authentication relies on visual memory. In both cases he or she has to access that secret in stored memory. Memory can be accessed in three ways, as described below, using the password as an example in each case to explain approaches for each way.

Recall: Information is extracted from memory when requested. This is the paradigm adopted by the traditional textual password authentication. Recall is a cognitively difficult task. Therefore, users tend to resort to coping strategies. Graphical passwords that rely on this kind of memory are the drawmetric based schemes like Android screen unlock and the searchmetric grIDsure [16].

Cued-Recall: Information is extracted from memory when cued. One can also ask for a password framed as a response task, similar to Zviran & Haga's associative passwords [17]. In their scheme users provide a number of associations

at enrolment which they are prompted for at authentication. Most of the graphical authentication schemes relying on this memory are locimetric, but exceptions such as the drawmetric BDAS exist [18].

Recognition: Information is presented and the individual is able to identify the correct item. One could conceivably display a number of passwords on the screen and ask a user to identify theirs, but this scheme is so obviously weak that it has not been trialled. Many graphical authentication mechanisms do rely on this mechanism, since it is the least cognitively demanding and particularly suitable for use with images. These are the searchmetric mechanisms. They display a succession of challenge sets, with one "target" image and a number of distractor images. The user identifies the target image in each challenge set by clicking on it.

The following sections provide examples of graphical authentication mechanisms that rely on each of these memory types. The sections are intended to be illustrative and examples of mechanisms have been chosen because they were the first of their kind. The inclusion of a particular mechanism does not suggest that it is in any way superior to others which are not mentioned. There is, unfortunately, no room for an exhaustive review of all mechanisms.

A. Recall

Draw-a-secret or DAS, proposed by [19], is a recall-based graphical authentication mechanism. The approach expects an individual to draw their authentication secret to access an application. These should be more memorable than passwords because they rely on visual, lexical and kinaesthetic memory [8]. DAS does not rely on drawings from a semantic perspective but on the underlying grid sectors.

Thorpe and van Oorschott [20] postulated that *symmetrical graphical secrets* are a real concern, as symmetrical drawings have superior recall. They argue that an attacker could craft a dictionary of symmetrical secrets and use it to compromise DAS which would take only 6 days to crack if the password is symmetrical. Nali and Thorpe conducted an informal user-study with 16 individuals to determine if user-generated DAS secrets exhibited any patterns [21]. Their participants' drawings did exhibit patterns: 45% created symmetrical drawings, 56% of the drawings were centred and 80% of the drawings used fewer than 3 strokes. The authors argue if these results are symptomatic of a larger user-base, then DAS has a much smaller, practical password-space.

The latest drawmetric mechanism is the common Android lock-screen pattern authentication. Such pattern based authentication mechanisms are also vulnerable to attacks based on observations of smudges on the device touchscreen [22].

Another recall-based approach is grIDsure, an authentication scheme that relies on knowledge of a secret pattern [16]. It requires an individual to create a sequenced pattern on a 5x5 grid. The user is presented the same 5x5 grid during authentication, except each one of the 25 cells contains a random value between 0 and 9. The values are randomly generated for each authentication attempt and are not unique to a cell. The secret pattern, generated by the user, is applied to the grid to generate the authentication secret, i.e. a 4-digit PIN. Bond identified some severe security flaws in this scheme [23]. He was able to identify the user's secret using only two forged authentication grids.

B. Cued-Recall

There is some concern that users forget their drawings with recall-based mechanisms such as DAS, or at least the stroke order [24]. To address this, Dunphy and Yan proposed a grid superimposed over a background image Background DAS (BDAS) to act as a cueing mechanism to improve memorability. Unfortunately users still created weak passwords [18].

Building on the ideas of Blonder, Wiedenbeck [25] proposed a mechanism called PassPoints. In PassPoints the user is expected to select five click points on an image. The sequence of click points is the authentication secret. Each position has a small tolerance radius as perfect replication is not expected. The password-space of PassPoints is vast, even with the addition of the tolerance radius, as a single image can contain a large amount of possible click points. The image can be selected from a library or provided by the user, the only requirement being that the image is complex enough to inspire users and protect the secret.

The apparent strength of the PassPoints approach is the large theoretical password-space afforded by the pixel-rich images. Thorpe and Van Oorshot argue the practical password space of PassPoints is reduced because of 'hot-spots', i.e. popular click points, as well as patterns within secret generation [26]. They investigated both human-based attacks and automated attacks. They investigated two highly-detailed images for popular positions. They discovered that 5 points in both images proved popular with individuals, between 24-31% for the first image and 20-24% for the second. Similarly, Dirik, Memon and Birget [27] developed a model that they claim can identify popular regions for points. They cautiously report that they were able to extract 70 - 80% of points. Furthermore, Thorpe and Van Oorshot [26] also suggest that predictable patterns exist in sequence selection.

C. Recognition

Dhamija and Perrig [1] propose Déjà Vu, a recognition-based graphical authentication mechanism. Each image is abstract in nature and the collection is generated using a mathematical formula, the output depends on an initial seed. The beauty of this design is that the actual images do not need to be stored, just the small initial seed. Déjà Vu performed well against competing recall-based approaches such as passwords and PINs. Indeed, Dhamija and Perrig reported that more individuals were unable to recall their username than were unable to recognise the images within their secret sets. Individuals using Déjà Vu felt that it was overall easier to use but at the expense of time and security. However, they reported an interesting insight in regards to the image-type used in Déjà Vu. When using semantic images, i.e. photographic scenes, some individuals selected the same images. One specific image

was selected by 9 out of 20 individuals. Furthermore, these images are far easier to explain and describe, thus, as a consequence an individual's secret set of images is easier to convey to someone else. For example, the aforementioned popular semantic image contained the Golden Gate bridge. Conversely, abstract images rarely overlapped and descriptions of them rarely, if ever, matched. In theory this strengthened the practical password space of the approach as there was no real pattern or popular images. Naturally, further investigation will be required but Dhamija and Perrig highlight the impact the image type can have on the memorability of images.

PassFaces is one of the few commercial graphical authentication mechanisms. The authentication approach assigns an individual a collection of faces as their authentication secret. The user is then presented a series of challenge stages that are each comprised of a nine image grid. In each grid the user has to identify one image from his or her password (*target*) among eight *distractors*. One property which might severely impact guessability (offline attacks) of searchmetric schemes, is that they usually need to store some password information in the clear [28].

IV. Strengthening The Graphical Password

A couple of improvements and also combinations of different approaches have been proposed in order to overcome weaknesses with respect to one or several of the vulnerabilities mentioned in Section II. These are proposed and discussed in this section.

A. Guessability Resistance

Cued-Recall: PassPoints authentication exhibits popular positions or hot-spots which are problematical in terms of guessability. Chiasson, van Oorschot and Biddle [29] propose Cued Click Points or CCP. The approach is a variation on PassPoints, in the sense that an individual selects a position from an image. However, the main difference is that an individual is required to repeat this action over several images. Therefore, the secret is a sequence of click points selected from a series of images with one click point on each image. The images are intended as cues.

There are concerns about the predictability of CCP, primarily those inherited from PassPoints, such as popular click points. Chiasson, Forget, Biddle and van Oorschot tackle this specific problem with Persuasive Cued Click Points or PCCP [30]. This approach uses CCP with the addition of a *persuasive viewport*. During the registration phase a viewport is randomly positioned over the image. The viewport is emphasised by reducing the brightness of the rest of the image. The individual is only allowed to select a click point from within the viewport, which they can shuffle if they do not like the position. Their lab-based experiments showed that the viewport is successful in reducing hot-spots and increases the spread of click points and their web-based trials were equally positive [30]. However, a more longitudinal evaluation is required to confirm whether the findings are replicated in the wild.

Recognition: Graphical authentication relying on recognition requires people to use images. There are three ways of associating people with images: (1) let them supply the images themselves, or (2) allow them to choose from a range of images or (3) assign images randomly to them. Unfortunately, the first two options can cause severe security concerns. If users are allowed to supply their own images they tend to choose predictable images [31]. The same holds, if users select their images from a set of supplied ones. As long as users have a choice they will behave predictably. Davis, Monrose and Reiter [32] discovered that individuals make predictable choices when they are required to select images for use in graphical authentication utilising facial images. Individuals are influenced by attraction, race and familiarity [33]. Even with everyday representational images, humans tend to make predictable choices [34]. To minimise guessability, images should be assigned to users randomly. In this case the theoretical and the practical password-space are the same.

B. Observation-Resistance

The first reaction to graphical authentication is often that it will be too easy for a human observer to gain knowledge of the authentication secret. Thus, a variety of attempts have been made to make it more difficult for observers to do this. Essentially, variants focussing on the observation resistance have been proposed for all three approaches to graphical authentication.

Recall: A DAS secret is easily exposed to onlookers. If an attacker is able to observe entry of a DAS secret, he or she may be able to authenticate using the same drawing (this holds for the cued-recall BDAS as well). Lin, Dunphy, Olivier and Yan proposed Qualitative DAS (QDAS) to tackle the problem of observation [35]. Chakrabarti, Landon and Singhal argued that rotating the canvas which the user draws on could improve the resiliency of DAS to observation [36]. Yet, neither of these has been tested in the wild so only lab-based results about the impact of the scheme on ease of observation have been reported.

Cued-Recall: While originally intended to be used with recognition-based authentication systems, the approach of Dunphy, Heiner and Asokan proposed in [15] can easily be translated to the CCP and the PCCP scheme. They propose to use a portfolio of target images of which only a different random subset is needed for the authentication process each time. Likewise CCP and PCCP could be implemented using a click point portfolio with only a random subset of the images (and the according click point) being needed to authenticate. To the authors' knowledge this has not yet been attempted, but might prove an interesting subject for future research.

Recognition: A range of observation resilient systems have been proposed. Dunphy, Heiner and Asokan [15] tested redundancy with users of a searchmetric system with 8 distractors and one target image. Each user had a portfolio of 6 images, of which only 4 were used at each authentication attempt. Attackers needed 7.5 observations, on average, in order to be able to reconstruct the password.

TABLE II
COMPARISON OF GRAPHICAL PASSWORD SCHEMES TO TEXTUAL PASSWORDS. G=GUESSABILITY, O=OBSERVABILITY, R=RECORDABILITY, M=MEMORABILITY, ?=NO DATA AVAILABLE.

	Scheme	Guessability	Observability	Recordability	Memorability
Recall	QDAS [35]	G ?	O ?	R ?	M ?
	BDAS [18]	$G\downarrow$	O ?	$R=$	M ?
Cued-recall	PassPoints [37]	$G=$ [26], [38]	O ?	$R=$ [39]	$M\uparrow$ [40], [6], [41]
	PCCP [30]	$G\uparrow$ [30]	O ?	$R=$ [39]	$M=$ [30]
Recognition	Passfaces [42]	$G\downarrow$	$O\uparrow$ [43], [15]	$R=$ [39]	$M\uparrow$ [44]
	Passfaces (system issued passwords)	$G\uparrow$	$O\uparrow$ [43], [15]	$R=$ [39]	M ?
	Dynahand [45]	$G\downarrow$ [45]	$O=$ [45]	R ?	$M=$

Wiedenbeck [46] proposes an authentication approach, framed as a game with a cognitive trapdoor, that relies on users generating a convex hull. The cognitive trapdoor is knowledge of 5 icons or pass-icons. The game comprises of a series of challenges, in each challenge the user needs to click within a specific region to 'win'. The specific region is revealed by uncovering a convex hull. During authentication, the user is presented a canvas containing several icons, including *at least* 3 pass-icons. The user is required to envisage the convex hull spanned by the pass-icons and needs to click within the convex hull to complete the challenge. Regardless of whether users click within the correct region or not, they progress to the next challenge when they click on the canvas. Wiedenbeck states that authentication times for the proposed Convex Hull Click approach are lengthy, they are the necessary expense of increased resilience to observation. Moreover, the author argues that energy has been spent to delight and excite the user to maintain interest with the approach.

Searchmetric mechanisms can be morphed into *limited disclosure* searchmetric mechanisms to foil shoulder surfing and key-logging software, since they rely on the use of arrow keys or a mouse to manipulate sets of pictures and the user does not click on their actual image. Most limited disclosure searchmetric mechanisms have some redundancy built in so that the observer is not able to deduce the key from casual observation but has either to observe a number of authentications or carry out an error-prone deduction of the key based on a few observations.

Tetrad [47] was proposed by Renaud and Maguire. Tetrad displays a grid (9x5) of facial images. Users line up their secret images by manipulating rows and columns instead of clicking on the images themselves. This introduces a level of indirection which means that casual observation is less profitable to an attacker.

C. Recordability Resistance

When graphical passwords were originally released, one of the their most touted strengths was the fact that they would be harder to record than passwords. This was naïve, in hindsight. The ubiquity of mobile phones with cameras makes it trivial to record anyone doing anything, including using the mouse to enter the graphical password. Even without the use of additional electronic gadgets incorporating cameras, the computer user can record the graphical password easily using the ever-available screenshot facility. Storing such an image on the hard drive or printing it out is equivalent to an unencrypted textfile or the infamous post-it for the textual password.

V. COMPARISON

Before providing the comparison and discussing it, we discuss some of the requirements of textual passwords relevant for the comparison.

A. Some information about textual passwords

The following information has been considered for the comparison in the next subsection.

Guessability: The susceptibility of textual passwords to guessing attacks has been shown again [2] and again [48]. User-chosen textual passwords are not uniformly distributed in the password space. Malone and Maher showed that Zipf's Law is a relatively good model for password distributions [49]. Consequently, adversaries can easily create a dictionary of the most commonly used passwords. Weir, Aggarwal, Collins and Stern used such dictionaries to conduct an analysis of large sets of revealed textual passwords [2]. They were able to crack at least about one fifth of 8 character passwords in those sets and only slightly less of the passwords with a length of 9 and 10 characters.

Observability: As described above in Section II it is important to differentiate between human and technical observers. Using technical means it is possible to reconstruct textual passwords from video footage or even from the sound of the keyboard input [50]. Human observers have to rely on visual input. Tari, Ozok and Holden [51] discovered that when users type long and obscure passwords, entry is more easily observed by shoulder-surfers than when typing simple and familiar words. Unfortunately, textual authentication secrets generated by users to protect bank accounts and tax records are likely to exhibit exactly these characteristics, so efforts spent by a user to be "secure" might actually backfire. Even so, most users are fairly confident that observers cannot guess their password with any degree of accuracy [52] even though such confidence is probably misplaced.

Recordability: Passwords are trivial to record and users can and will write down their passwords when the burden of recovering a lost password is too high [53]. Security professionals realise that this is an inherent vulnerability of

textual passwords, so they deal with it by instructing people not to take this action. However, their instructions are mostly ignored [3].

B. Overview of the comparison

Table II presents a comparison between the textual password and some graphical authentication schemes. To give a better idea of how secure an authentication scheme is, we introduce the following four levels.

1) A scheme is considered equal to textual passwords in our setting (=), when it offers roughly the same resistance to common attacks as does the password.
2) A scheme is considered worse than textual passwords in our setting (↓), when it offers even less resistance to common attacks than the password.
3) A scheme is considered better than textual passwords in our setting (↑), when it is better than the password in this particular respect.
4) A scheme might not allow a rating (?) if no data regarding the aspect is available or the available data does only allow a very rough estimation instead of a real assessment (e.g. very small sample).

These levels are based on the literature which often reports findings that are extremely difficult to compare, so the comparison should not be considered definitive, but rather based on an understanding of whether the approach is prone to show vulnerabilities. Moreover, it becomes apparent that there are many aspects that do not allow a rating due to missing data or data that only allows a very rough estimation instead of a real assessment.

In terms of *guessability*, graphical mechanisms generally can be as strong as textual passwords in practise are. Most schemes have a variable password space and can therefore easily be adopted to be resistant to pure brute force attacks. This, however, requires special configurations to strengthen them, which have not yet been tested in the wild. Dictionary attacks remain a severe concern for many schemes, but examples such as PCCP show that guiding the user during password choice using persuasive technology can mediate this issue. Recognition-based schemes in which the password is issued by the system can even avoid this problem entirely. *Observability* is a problem across the board and attempts to introduce redundancy or indirection into the process tend to increase authentication times unacceptably (eg. up to 180 seconds [54]). However, it must be acknowledged that comparisons between textual passwords and graphical schemes regarding the vulnerability to shoulder surfing are hard to find, even among those schemes specifically designed to be observation resistant. It could be that they are equally resistant or even better as sometimes appraised by their proponents, but in the absence of hard data we cannot judge. Also, whether the approach of Dunphy, Heiner and Asokan can be successfully applied to a wider range of schemes than initially proposed might be an interesting topic for future research. In terms of *recordability* there does not seem to be much difference between graphical and textual passwords. Recognition-based secrets are mostly more *memorable*, but the other types can display the same memorability as textual passwords.

C. Discussion

The previous section has shown that many of the proposed graphical authentication schemes exhibit advantages and disadvantages in one area or another. Most have advantages regarding their memorability. This is no surprise, as the intention behind graphical passwords was to relieve users of the cognitive burdens of textual passwords.

The predictability of the users' drawings and the complete recreation of the secret during authentication in the drawmetric approaches (both recall and cued-recall based) cause severe concern and it is unclear or open for future research whether they perform good enough with respect to their memorability if the system would set the drawing for the user. However, the iterative processes the locimetric appraoches (i.e. the remaining cued-recall based approaches) and the searchmetric/recognitaion-based approaches have gone through has resulted in a more robust set of mechanisms with respect to guessability. For example, PCCP is very guessability resistant in particular compared to the textual passwords in our setting. The same holds for recognition-based schemes with system issued passwords.

While, the focus of graphical password research was on improving memorability they raised new usability issues: It is clear from the literature that it takes at least as long [6] for users to authenticate using these mechanisms and mostly even longer [30], [55], [1]. Thus, depending on how regularly one needs to authenticate some of the graphical alternatives might not be an alternative although security wise at least as secure as textual passwords. At creation time this makes them inconvenient but also extends the window for observation. Whether the timings achieved by systems more resilient to observation or predictability are acceptable might be up for debate, as the difference decreases the longer the period between two logins with the same password [1].

However, when considering the time it takes to login one should also consider the time and effort it takes to reset passwords. Thus, if it is much less often needed to replace a secret for a particular type of graphical password due to its superior memorability, a longer login time might become acceptable. The resulting benefits in time expenditure and convenience might well be worth the offset. To make a decision here it is very important to have detailed knowledge about the situation and application for which an alternative is considered.

Graphical authentication has its strengths and its weaknesses. Where authentication timings are of the essence other solutions might be a better choice. However, when retention times are high (consider e.g. a task that has to be carried out a few times a year), graphical passwords, with their superior memorability, can mediate.

Another important aspect is that one can never look at authentication in isolation. The context of deployment has to be considered. Different devices impose different constraints

on the authentication process. The two most prominent are desktop computers and mobile devices such as smartphones or tablets. Desktop computers and laptops normally offer high resolution screens and diverse input devices. The variety of techniques that can be applied is thus larger. For recognition-based schemes, however, the available data suggest, that the chosen bitstrength has a severe influence on the efficiency of the systems. Traditionally, mobile devices offer fewer resources such as smaller screens and especially a limited physical keyboard (if at all). Hence, keyboard-based solutions are mostly infeasible in the mobile environment. The most important factor here is eavesdropping through shoulder-surfing, as usage of such devices often occurs in public places [15]. As on mobile devices PIN-equivalent security is the de facto standard, recognition-based authentication schemes using shoulder-surfing resistant variants, as e.g. proposed by Dunphy, Heiner and Asokan, come to mind. They offer an alternative to textual passwords comprising both, high usability and security equal to the PINs they ought to replace.

VI. Conclusion

The base question behind this work was whether graphical authentication can be as strong as textual authentication. Based on our analysis the answer must be: "it can be, if you design it properly." So, what constitutes a proper design? Regarding the three metrics guessability, observability and recordability the following three considerations can serve as a first quick assessment:

(1) Unguided user choice translates to predictable choice. A resistant scheme should specifically encourage or even better force users to choose random passwords or have the system issue passwords.

(2) Obfuscation techniques, such as the asterisks routinely used to obscure textual password entry or portfolio-based approaches are examples of observation resistance.

(3) The scheme should generate secrets that are hard to describe or record.

These three considerations should serve only as a starting point for an evaluation. In some situations not all of the three aspects above might be important. For example, the ubiquitous textual password does not conform to statement (3). Whether a scheme is appropriate for a certain situation depends on the context of use, the risk associated with the asset the authentication mechanism controls access to, the time constraints, the device being used and the frequency of use. If the mechanism is low risk and used infrequently, graphical authentication might well be better than textual authentication.

Graphical schemes have the potential to be as secure as textual systems. Yet, the jungle of diverse graphical authentication schemes easily explains decision-makers' reluctance to adopt graphical authentication. They are rightly sceptical about the strengths of the mechanisms and also still in a one-authenticator-for-everything mindset. If we want this mindset to change, we, as researchers, will have to provide a way for decision makers to start deploying a wider range of mechanisms, in a more nuanced and discriminating fashion than a one-size-fits-all password. What we should be striving for is more diversity of authentication mechanism usage, deploying the wide variety of mechanisms that have been trialled in situations where we can match them to the risk mitigation requirement and deployment context. This way, users would finally benefit from superior memorability of available alternatives and organisations from less secret reuse across systems.

Therefore, we want to encourage the research community to try and fill the gaps in the knowledge about the promising graphical authentication schemes already in existence, rather than proposing even more. Unification of testing methodologies and comparability of approaches should be the goal. This can facilitate decisions to integrate one of the schemes in a particular application or environment and could give insights important not only for the domain of graphical authentication, but for authentication as a whole.

References

[1] R. Dhamija and A. Perrig, "Déjà vu: a user study using images for authentication," in *Proc. SSYM '00*, 2000, pp. 4–4.

[2] M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing metrics for password creation policies by attacking large sets of revealed passwords," in *Proc. CCS '10*, 2010, pp. 162–175.

[3] A. Adams and M. A. Sasse, "Users are not the enemy," *Comm. of the ACM*, pp. 40–46, 1999.

[4] N. Ben-Asher, N. Kirschnick, H. Sieger, J. Meyer, A. Ben-Oved, and S. Möller, "On the need for different security methods on mobile phones," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '11, 2011, pp. 465–473.

[5] B. Ives, K. R. Walsh, and H. Schneider, "The domino effect of password reuse," *Comm. of the ACM*, vol. 47, pp. 75–78, 2004.

[6] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle, "Multiple password interference in text passwords and click-based graphical passwords," in *Proc. CCS '09*, 2009, pp. 500–511.

[7] G. Blonder, "Graphical password," Sep. 24 1996, uS Patent 5,559,961.

[8] K. Renaud and A. De Angeli, "Visual passwords: cure-all or snake-oil?" *Communications of the ACM*, vol. 52, no. 12, pp. 135–140, 2009.

[9] A. Paivio, *Mental representations. A dual coding approach*. New York: Oxford University Press, 1986.

[10] R. English and R. Poet, "Towards a metric for recognition-based graphical password security," in *Network and System Security (NSS), 2011 5th International Conference on*. IEEE, 2011, pp. 239–243.

[11] A. De Angeli, L. Coventry, G. Johnson, and K. Renaud, "Is a picture really worth a thousand words? exploring the feasibility of graphical authentication systems," *International Journal of Human-Computer Studies*, vol. 63, no. 1, pp. 128–152, 2005.

[12] X. Suo, Y. Zhu, and G. S. Owen, "Graphical passwords: A survey," in *Computer Security Applications Conference, 21st Annual*. IEEE, 2005, pp. 10–19.

[13] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, "Of passwords and people: measuring the effect of password-composition policies," in *Proc. CHI '11*, 2011, pp. 2595–2604.

[14] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid, "Recommendation for key management - part 1: General," in *NIST Special Publication 800-57, NIST*, 2005.

[15] P. Dunphy, A. P. Heiner, and N. Asokan, "A closer look at recognition-based graphical passwords on mobile devices," in *Proc. SOUPS '10*, 2010, pp. 3:1–3:12.

[16] S. Brostoff, P. Inglesant, and M. Sasse, "Evaluating the usability and security of a graphical one-time pin system," in *Proceedings of the 24th BCS Interaction Specialist Group Conference*. British Computer Society, 2010, pp. 88–97.

[17] M. Zviran and W. J. Haga, "Cognitive passwords: the key to easy access control," *Computers & Security*, vol. 9, no. 8, pp. 723–736, 1990.

[18] P. Dunphy and J. Yan, "Do Background Images Improve "Draw a Secret" Graphical Passwords?" in *Proceedings of the 14th ACM Conference on Computer and Communications Security*. ACM, October 2007, pp. 36–47.

[19] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin, "The Design and Analysis of Graphical Passwords," in *Proceedings of the 8th USENIX Security Symposium*. Washington DC, 23-26 August 1999, pp. 1–14.

[20] J. Thorpe and P. Van Oorschot, "Graphical dictionaries and the memorable space of graphical passwords," in *13th USENIX Security Symposium*, 2004, pp. 135–150.

[21] D. Nali and J. Thorpe, "Analyzing user choice in graphical passwords," *School of Computer Science, Carleton University, Tech. Rep. TR-04-01*, 2004.

[22] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proceedings of the 4th USENIX conference on Offensive technologies*. USENIX Association, 2010, pp. 1–7.

[23] M. Bond, "Comments on gridsure authentication," 2008, http://www.cl.cam.ac.uk/ mkb23/research/GridsureComments.pdf.

[24] J. Goldberg, J. Hagman, and V. Sazawal, "Doodling our way to better authentication," in *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 2002, pp. 868–869.

[25] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, "PassPoints: Design and Longitudinal Evaluation of a Graphical Password System," *International Journal of Human-Computer Studies*, vol. 63, no. 1, pp. 102–127, 2005.

[26] J. Thorpe and P. van Oorschot, "Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords," in *Proceedings of the 16th USENIX Security Symposium*. USENIX Association, 06-10 August 2007, p. 8.

[27] A. Dirik, N. Memon, and J. Birget, "Modeling user choice in the PassPoints graphical password scheme," in *Proceedings of the 3rd Symposium on Usable Privacy and Security*. ACM, 18-20 July 2007, pp. 20–28.

[28] R. Biddle, S. Chiasson, and P. van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys 44(4)*, 2011.

[29] S. Chiasson, P. van Oorschot, and R. Biddle, "Graphical Password Authentication Using Cued Click Points," *European Symposium on Research in Computer Security*, vol. 4734, pp. 359–374, September 2007.

[30] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. C. Van Oorschot, "Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism," Carleton University, Ottawa, Canada, Tech. Rep., 2011.

[31] K. Renaud, "On user involvement in production of images used in visual authentication," *Journal of Visual Languages & Computing*, vol. 20, no. 1, pp. 1–15, 2009.

[32] D. Davis, F. Monrose, and M. Reiter, "On User Choice in Graphical Password Schemes," in *Proceedings of the 13th USENIX Security Symposium*, 2004, pp. 151–164.

[33] J. N. Maguire, "An ecologically valid evaluation of an observation-resilient graphical authentication mechanism," Ph.D. dissertation, Computing Science, 2013.

[34] R. English and R. Poet, "Measuring the revised guessability of graphical passwords," in *Network and System Security (NSS), 2011 5th International Conference on*. IEEE, 2011, pp. 364–368.

[35] D. Lin, P. Dunphy, P. Olivier, and J. Yan, "Graphical Passwords & Qualitative Spatial Relations," in *Proceedings of the 3rd Symposium on Usable Privacy and Security*. ACM, 18-20 July 2007, pp. 161–162.

[36] S. Chakrabarti, G. Landon, and M. Singhal, "Graphical passwords: drawing a secret with rotation as a new degree of freedom," in *The Fourth IASTED Asian Conference on Communication Systems and Networks (AsiaCSN 2007)*. Citeseer, 2007, pp. 561–173.

[37] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, "Passpoints: design and longitudinal evaluation of a graphical password system," *Int. J. Hum.-Comput. Stud.*, vol. 63, pp. 102–127, 2005.

[38] P. C. Van Oorschot, A. Salehi-Abari, and J. Thorpe, "Purely automated attacks on passpoints-style graphical passwords," *Trans. Info. For. Sec.*, vol. 5, pp. 393–405, 2010.

[39] P. Dunphy, J. Nicholson, and P. Olivier, "Securing passfaces for description," in *SOUPS '08: Proceedings of the 4th symposium on Usable privacy and security*. ACM, Jul. 2008.

[40] S. Chiasson, R. Biddle, and P. van Oorschot, "A Second Look at the Usability of Click-based Graphical Passwords," in *Proceedings of the 3rd Symposium on Usable Privacy and Security*. ACM, 2007, pp. 1–12.

[41] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: Basic results," in *Proc. HCI International '05*, 2005.

[42] Real User Corporation, "The science behind passfaces," 2004. [Online]. Available: http://www.realuser.com

[43] F. Tari, A. A. Ozok, and S. H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *Proc. SOUPS '06*, 2006, pp. 56–66.

[44] S. Brostoff and M. A. Sasse, "Are passfaces more usable than passwords? a field trial investigation," in *Proc. HCI '00*, 2000. [Online]. Available: http://www.cs.ucl.ac.uk/staff/S.Brostoff/index_files/brostoff_sasse_hci2000.pdf

[45] K. Renaud and E. Olsen, "Dynahand: Observation-resistant recognition-based web authentication," *Technology and Society Magazine, IEEE*, vol. 26, no. 2, pp. 22–31, 2007.

[46] S. Wiedenbeck, J. Waters, L. Sobrado, and J. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2006, pp. 177–184.

[47] K. Renaud and J. Maguire, "Armchair authentication," in *BCS-HCI '09: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*. British Computer Society, Sep.

[48] C. Herley and P. van Oorschot, "A Research Agenda Acknowledging the Persistence of Passwords," *Security & Privacy, IEEE*, vol. 10, no. 1, pp. 28–36, 2012.

[49] D. Malone and K. Maher, "Investigating the distribution of password choices," in *WWW '12: Proceedings of the 21st international conference on World Wide Web*. ACM, Apr. 2012.

[50] L. Zhuang, F. Zhou, and J. D. Tygar, "Keyboard acoustic emanations revisited," in *Proceedings of the 12th ACM conference on Computer and communications security*, ser. CCS '05, 2005, pp. 373–382.

[51] F. Tari, A. Ozok, and S. Holden, "A Comparison of Perceived and Real Shoulder-surfing Risks between Alphanumeric and Graphical passwords," in *Proceedings of the 2nd Symposium on Usable Privacy and Security*. ACM, 12-14 July 2006, pp. 56–66.

[52] D. Weirich and M. Sasse, "Pretty Good Persuasion: A First Step towards Effective Password Security in the Real World," in *Proceedings of the 2001 Workshop on New Security Paradigms*. ACM, 2001, pp. 137–143.

[53] P. G. Inglesant and M. A. Sasse, "The true cost of unusable password policies: password use in the wild," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10, 2010, pp. 383–392.

[54] D. Weinshall, "Cognitive authentication schemes safe against spyware," in *Security and Privacy, 2006 IEEE Symposium on*, 2006, pp. 6 pp.–300.

[55] K. M. Everitt, T. Bragin, J. Fogarty, and T. Kohno, "A comprehensive study of frequency, interference, and training of multiple graphical passwords," in *Proc. CHI '09*, 2009, pp. 889–898.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 845–849

Tests of Smartphone Localization Accuracy Using W3C API and Cell-Id

Grzegorz Sabak
Orange Polska
ul. Św. Barbary 10, 82-300 Warszawa, Poland
e-mail: grzegorz.sabak@orange.com

***Abstract*—Location based services (LBS) are considered very relevant for the users of mobile networks. All local events and facts related to area nearby seem to be more important that others which happen in remote places. Localization data is used in all types of services: weather, traffic, tourist info, etc. One of its most important (and regulated by law) applications is providing persons location in case of emergency.**

This paper presents results of field tests related to assessment of accuracy of two most commonly used localization methods: Cell-Id and W3C Geolocalization. The tests were conducted in the form of test drives along some of the most important roads in Poland. Position of the test vehicle obtained using analyzed methods was logged and compred to localization obtained from the Global Positioning System (GPS).

Data collected during test drives was processed and statistical information about localization accuracy was calculated. Results obtained for different methods were compared and conclusions about localization quality are provided.

The paper also describes test environment and data model which were used during work being reported.

I. Introduction

RECENT surge in number of smartphones used worldwide causes increased interest in development of applications dedicated for advanced mobile phones. There are many types of applications available which can be either bought or downloaded for free. Depending on user's current needs he or she can select from company's application shop (e.g. Google Play, Samsung Apps, iTunes) what suits him best. Typically, after having used application for some time user is expected to assess the application and share his/her opinion for the benefit of its developers and future users.

Smartphones are considered to be personal devices which are almost always carried by their owners and used in all kinds of places and situations. This means that *portability* is one of their key features and all services and content which are based on their localization are more relevant than generally available, non-localization dependent, information.

Valuable and popular services are available not only in the form of applications. To provide a complete offer, application developer very often prepares a version which can be used in a web browser run on a mobile phone. Such services are typically written using HTML5 and Javascript which gives great flexibility in preparing user interface which can be very similar to interface prepared for native application.

This paper focuses on assessment of localization algorithms available to developers of browser based services.

II. Localization Methods

There are different sources of device localization available for the application and service developers. The best accuracy can be obtained using positioning systems based on signals emitted by satellites (e.g. GPS, GLONASS, Galileo) [10], [4]. However, it is not always possible to use this method of localization (due to lack of satellite signal, limitations of device battery capacity or simply lack of required signal receiver in the device). In such cases localization can be obtained from a cellular network or through algorithms provided by operating system vendor.

A. W3C Gelocalization API

This API [9] was proposed by World Wide Web Consortium (W3C) [2] as the uniform way to access mobile device location from the Web browser. It is currently implemented in all popular browsers. The API defines programmatical access to localization data. Taking available information as input data, dedicated algorithms are able to calculate position of the mobile device and asses accuracy of such calculations. The most commonly used data sources include: Wi-Fi connection parameters, device's IP address used for mobile communication, list of sensed GSM/CDMA cells, radio communication signal strength.

Location providers continuously collect data from mobile devices being used worldwide and improve quality of localization accuracy. However, because they do not control configuration of the infrastructure which is used for mobile communication, any major change in it may cause drop in the quality of information obtained through the API.

B. Cell-Id based localization

Localization based on Cell-Id is one of the most commonly methods used by land mobile networks. Its popularity comes from the fact that it relies on the mechanisms already in place which are required for basic voice and data communication.

Implementation of Cell-Id localization requires relatively low investment in network infrastructure. Usually, deployment of a Gateway Mobile Location Center (GMLC) is the key part of projects aiming to launch such capabilities in the mobile network.

In this method, a geographical location (a pair of coordinates) is assigned to every cell in the network. Location error

978-83-60810-53-8/$25.00

depends on the size of a cell and can vary from tens of meters to c.a. 20-30 kilometers.

Localization accuracy is related to the size of the cell which serves the mobile device. Previous work [5], [8], [12], [11] showed that cell sizes depend on the type of the area. Smaller cells are found in city centers which is in accordance with network capacity requirements. In the rural areas, bigger cell sizes are used to ensure network coverage.

Other, more accurate, localization methods are also available: Time of Arrival, Enhanced Observed Time Difference [7], [6]. They rely on information coming from more than one network towers (base stations). Using information about signal strength and/or time parameters of communication between the device and the base station, it is possible to calculate user location with smaller error. This, however, requires additional infrastructure which is not available in many networks.

III. Data Model

Let *a localization* of an object be defined as a pair $p = \langle x, y\rangle \in \Re^2$, where x and y are called *coordinates* of object's location. Distance between two objects will be denoted as $d : \Re^2 \times \Re^2 \mapsto \Re^+ \cup \{0\}$.

Many different coordinate systems are used worldwide. Humanity since centuries needed the ways to represent parts of the Earth surface (modeled as an ellipsoid) as a subsets of a plane (the maps). Nowadays, the projection which is used in GPS (known as WGS84) is a de facto standard in the Internet. In WGS84 *geographic* coordinates i.e. latitude and longitude of a point are provided. Because of that, in order to avoid systematic error in calculations of the objects' distance, coordinates have to be translated to a Cartesian coordinate system. PUWG2000 which is a Polish standard of geodetic coordinates was used during calculations reported in this paper.

Let localization event l be a triple $l = \langle t, p, a\rangle$, where $t \in \Re^+ \cup \{0\}$ is a timestamp of the event, p is object's location at time t and $a \in \Re^+$ is a measure of localization accuracy.

Let $L = \{l_1, l_2, \ldots, l_N\}$ denote *localization event stream* defined as a finite sequence of N localization events ordered according to timestamp values.

L models information about object's location at some points in time. Based on L it is possible to estimate object's location $p'(t, L)$ at any point in time. During calculation of $p'(t, L)$ it is assumed that:

- the object was located at p_1 for all $t < t_1$,
- for $t_k, t_{k+1} \in \langle t_1, t_N\rangle$ the object was moving with constant velocity along line segment $\overline{p_k p_{k+1}}$,
- the object was located at p_N for all $t > t_N$.

During the tests the following localization data was collected:

- L^{GPS}- vehicle track logged by GPS device (Garmin eTrex Vista),
- L^{Orange}- geolocalization using Cell-Id in Orange Polska,
- L^{iPhone}- sequence of localization events of iPhone device,
- L^{Android}- W3C Geolocalization API events obtained for Android device.

The goal of the test was to compare different localization methods. The basic measure of localization quality is a localization error $e \in \Re^+ \cup \{0\}$. It is defined as the distance between point being a result of analyzed localization method and a point $p^{\text{GPS}} = p'(t, L^{\text{GPS}})$ i.e. object's localization according to GPS receiver.

Let $E = \{e_1, e_2, \ldots, e_N\}$ be a localization error stream defined as a finite sequence of localization errors. For each L a localization error stream can be calculated, providing that a reference localization event stream $L^{\text{GPS}}$is available.

In order to compare different localization methods, the basic statistics were calculated for E^{Orange}, E^{iPhone}, E^{Android}which denote localization error streams calculated for L^{Orange}, L^{iPhone}, L^{Android}respectively. The results of these calculations are presented in the following sections of this report.

IV. Test Environment Configuration

The tests were performed in the form of test-drives during which localization of the vehicle was monitored using methods which were to be compared. Additionally a GPS receiver was put onboard of the test vehicle. It was used for logging of the vehicle location with maximum available accuracy and frequency.

During the tests the following devices were used: iPhone 3GS (iOS), ZTE SanFrancisco (Android), and Garmin eTrex Vista GPS receiver. Smartphones were equipped with subscription of Orange Polska mobile network.

The Fig. 1 shows main components of the test environment and main communication channels between them. iPhone and Android smartphones communicate with W3C API servers to receive device localization. When localization is calculated and returned to the device it sends a request to log it in the test application. Devices' localization is concurrently monitored by mobile network through Gateway Mobile Location Centre (GMLC). Position of the vehicle is logged by GPS device.

For the purpose of the test a dedicated web page was designed, developed, and exposed in the Internet. Its role was to trigger calls to W3C Gelocalization API on any device which opened it in the Web browser. Result of the localization API calls was logged in a database which was located on a Web server.

V. Test results

The test drive took place in January 2013. Its route is shown in Fig. 2. The tests covered some of the most popular roads in Poland. Test drive started in Warsaw and then proceeded through Czestochowa, Krakow, Nowy Targ, Kielce, Radom and ended in Warsaw.

A. Maps

After completion of the test-drive, data were processed and visualized using open source Quantum GIS [3] software. Sample maps showing data collected during tests-drives are presented in the Fig. 3, Fig. 4, and Fig. 5.

Places which are results of localization procedures are marked with dots. Dashed line segments connect locations obtained through different methods and actual device's position.

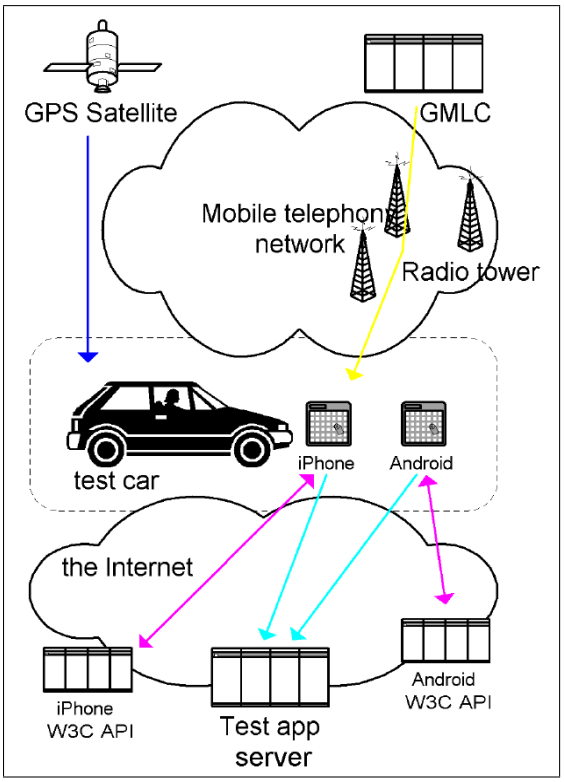


Fig. 1. Test data collection environment

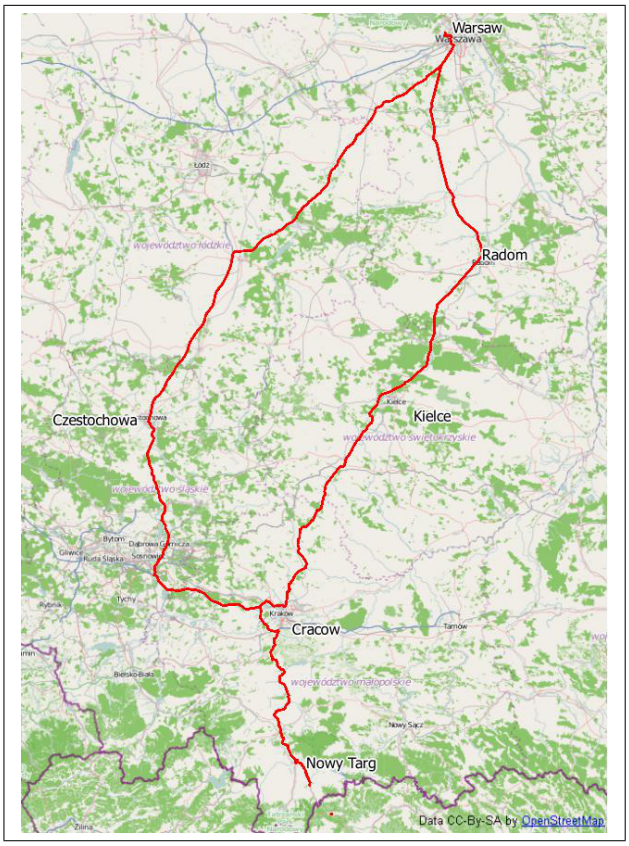


Fig. 2. Test drive routes

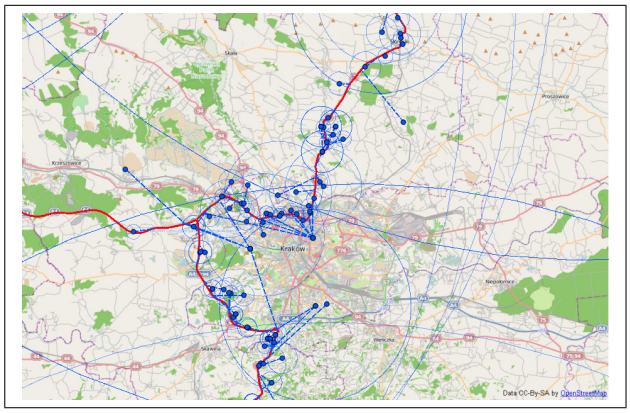


Fig. 3. iPhone location events in Cracow area

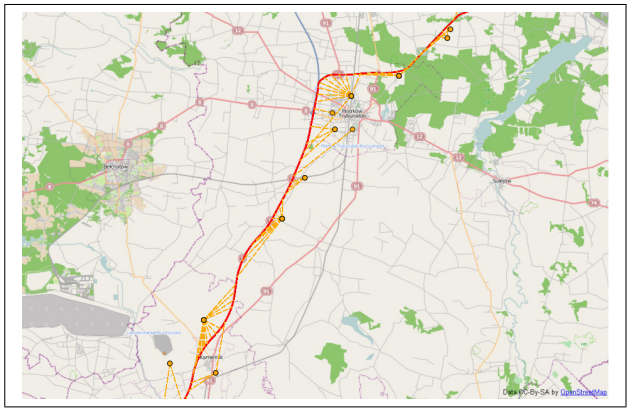


Fig. 4. Cell-Id (Orange Polska) localization events

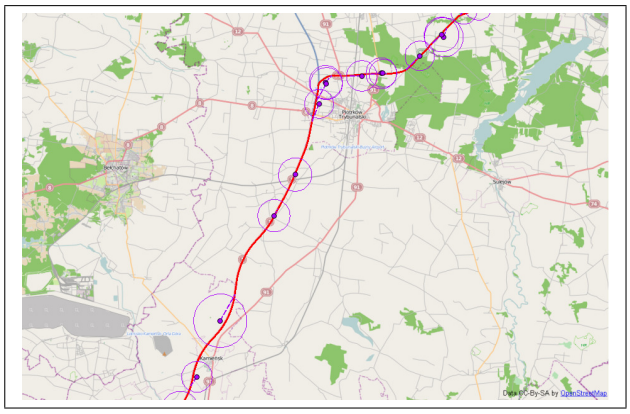


Fig. 5. Android localizations near Piotrkow Trybunalski

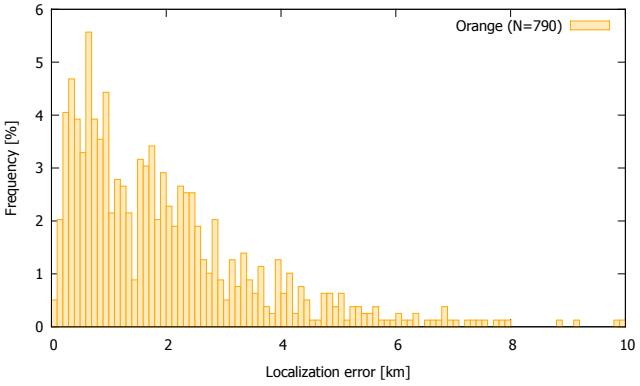


Fig. 6. Cell-Id location error distribution

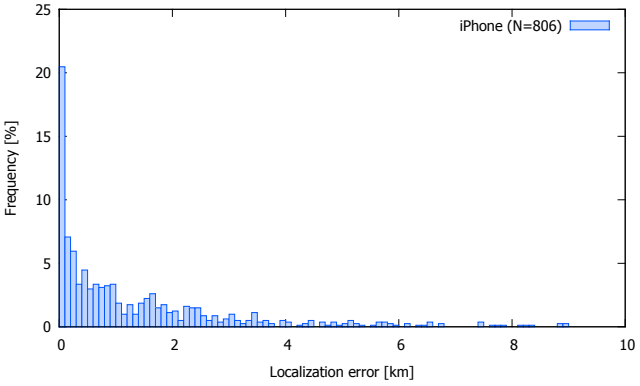


Fig. 8. iPhone location error distribution

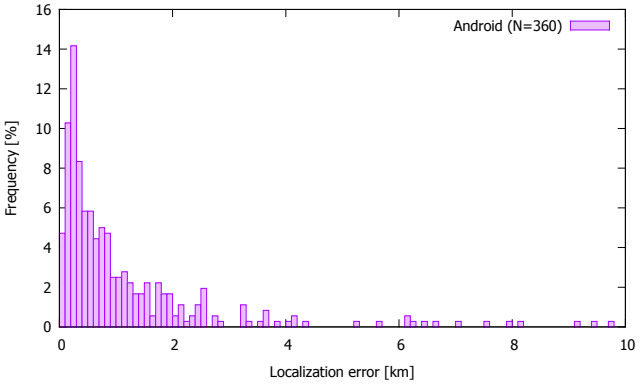


Fig. 7. Android location error distribution

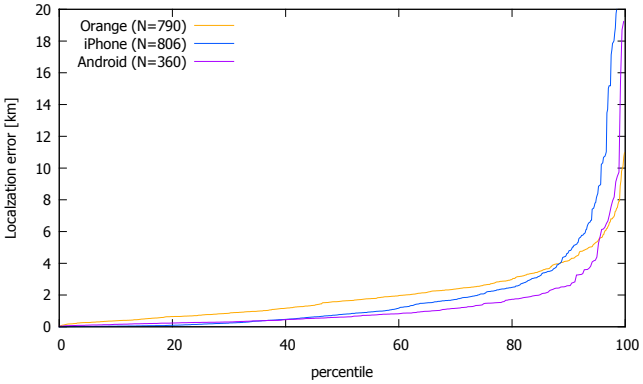


Fig. 9. Error value - cumulative distribution

Circles represent accuracy of localization as returned by W3C Geolocation API.

From the analysis of graphical data it can be concluded that in most of the cases W3C API localization is much more accurate than one based on Cell-Id. However, for some events W3C API localization error (examples can be seen in Fig. 3), is several times higher and is comparable to localization errors found in Cell-Id method.

B. Localization error distribution

In Fig. 6 location error density for Cell-Id is shown in the form of a normalized histogram. The maximum number 5.6% of location events fall in to the bin representing range $\langle 0.6\text{km}, 0.7\text{km}\rangle$ of error value.

Similar histograms visualizing data related to W3C API implemented in iPhone and Android mobile phones are shown in Fig. 7 and Fig. 8.

The bins with maximum percentage of events are: $\langle 0.2\text{km}, 0.3\text{km}\rangle$ for Android phone and $\langle 0.0\text{km}, 0.1\text{km}\rangle$ for iPhone. They account for 14.2% and 20.4% of localization events respectively.

C. Comparison

Localization quality of analyzed methods is compared on chart in Fig. 9 and Fig. 10. It presents cumulative distribution of localization error for all localization methods. It can be noted that for the range from 0m until about the median, iPhone localization is much better than other two. Taking into account about 90% of measurements Cell-Id localization is worse than other two. However, when all results are analyzed, the maximum error values significantly are smaller in case of Cell-Id than in methods based on W3C API.

Summary of basic statistical information is presented in

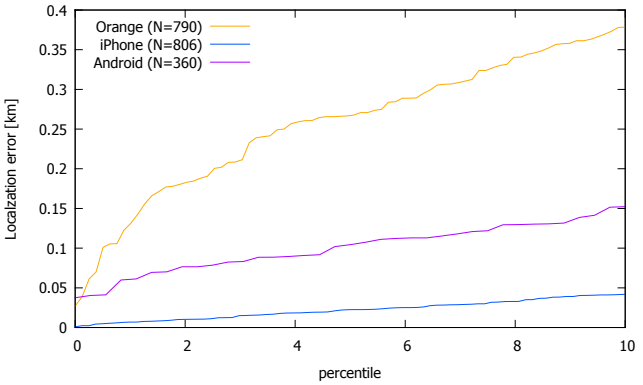


Fig. 10. Error value - cumulative distribution 0-10th percentile

TABLE I
COMPARISON OF LOCALIZATION ERROR STATISTICS. VALUES IN KM

Method	N	e_{avg}	σ	Q_1	Q_2	Q_3	e_{max}
Cell-Id	760	1.99	1.77	0.73	1.62	2.62	10.96
Android	360	1.28	2.16	0.28	0.61	1.40	19.24
iPhone	806	2.18	5.38	0.16	0.78	2.07	80.30

Table I. Taking into account average error value and standard deviation, Cell-Id seems to be the best localization method. However, there is big discrepancy in values of the first, second, and third quartiles in favor of W3C API methods.

VI. CONCLUSIONS

The set of tests showed that both W3C Geolocation API and Cell-Id methods are very good sources of location information for application developers. The quality of the W3C API implementations is very impressive, keeping in mind that API providers do not have any a priori information about spatial configuration of GSM and Wi-Fi networks.

Tests were performed with a kind of black box approach. Probably some results could be explained if geolocation algorithms used by Android and iPhone APIs were known.

REFERENCES

[1] ETSI TS 101 723 V7.3.0. (2001-04). Digital cellular telecommunications system (phase 2+) (GSM); Location Services (LCS); Service description, Stage 1 (3GPPTS 02.71 version 7.3.0 Release 1998, 2001

[2] `http://www.w3.org/`, homepage of World Wide Web Consortium, retrieved 8 May 2013

[3] `http://qgis.org/`, homepage of Quantum GIS project, retrieved 8 May 2013

[4] B. Harvey, *The rebirth of the Russian space program: 50 years after Sputnik, new frontiers*, Springer, Berlin, 2007.

[5] R. Klukas, G. Lachapelle, and M. Fattouche, "Field tests of a cellular telephone positioning system", In *Proc. IEEE 47th Vehicular Technology Conference*, Phoenix, AZ 4-7 May 1997

[6] S. A. Kyriazakos and G. T. Karetsos, "Architectures for the provision of position Location Services in cellular networking environments", in *Telecommunications and IT Convergence Towards Service Evolution*, Lecture Notes in Computer Science. Springer, Berlin, 2000.

[7] W. Michalski, *Przeglad metod okreslania lokalizacji abonentow w ruchomych publicznych sieciach komorkowych GSM/UMTS z uwzglednieniem dokladnosci dostarczanej informacji, technicznych mozliwosci wdrozenia oraz czynnikow ekonomicznych i prawnych. etap 1: Charakterystyka metod sluzacych do okreslania lokalizacji abonentow w sieciach GSM i UMTS* (in Polish) , National Institute of Telecommunications, Warszawa, 2007

[8] B. Ludden, A. Pickford, J. Medland, H. Johnson, F. Brandon, L. E. Axelsson, K. Viddal-Ervik, B. Dorgelo, E. Boroski, and J. Malenstein, *CGALIES final report, Report on implementation issues related to access to location information by emergency services (E112) in the European Union* Technical report, 2002.

[9] A. Popescu (editor), *Geolocation API Specification, W3C Proposed Recommendation 10 May 2012* , W3C, retrieved 10 May 20013

[10] R. Prasad and M. Ruggieri, *Applied satellite navigation using GPS, GALILEO, and augmentation systems. Mobile Communication Series* Artech House, Boston, MA, 2005

[11] G. Sabak, "Cell-ID based vehicle location accuracy - field tests report", *Logistyka*, nr 4/2010, Poznan, 2010.

[12] E. Trevisani and A. Vitaletti, "Cell-ID location technique, limits and benefits: an experimental study" In *Proc. Sixth IEEE Workshop on Mobile Computing Systems and Applications WMCSA 2004*, Low Wood, 2-3 Dec 2004.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 851–858

Integration of context information from different sources: Unified Communication, Telco 2.0 and M2M

Grzegorz Siewruk[1,2]
Marek Średniawa[2]
[1] Orange
ul. Skargi 56
03-516 Warsaw, Poland
Email: Grzegorz.Siewruk@orange.com
[2]Warsaw University of Technology Faculty of
Electronics and Information Technology
ul. Nowowiejska 15/19
00-665 Warsaw, Poland
Email: mareks@tele.pw.edu.pl

Sebastian Grabowski
Jarosław Legierski
Orange Labs
ul. Obrzeżna 7,
02-691 Warsaw, Poland
Email:
sebastian.grabowski@orange.com
jaroslaw.legierski@orange.com

***Abstract*—The paper presents an idea of a context-aware application, which collects context data from many different sources, stores them in a dedicated database and makes use of it to support flexible scenarios for end users. Using open APIs it integrates different types of context information provided by: Unified Communication system, APIs exposed by communication service providers and information from Machine to Machine (M2M) framework. Methods for recording and unifying different types of context data are proposed and their performance is compared with results for the most popular database structures. A context-aware contact list application for a mobile phone user is presented as an example illustrating the main ideas of the paper.**

I. Introduction

NOWADAYS, value of information about context of communication is growing and having more impact on next generation of context aware services. Companies such as Google, Twitter or Facebook which are in possession of large databases which contain mostly context information [1] became major players in the ICT market. Information about users and their activities is the base of social networks' owners' business model and has started to be monetized, e.g. mainly in personalized focused advertising.

Availability and usage of precise context information has a very big potential for enterprises and constitutes a catalyst for innovative ICT systems and applications [2]. From a mobile phone user's point of view, participation in social networks implies a change of a lifestyle. People want to "stay in touch" which means that they have to be reachable via phone independently of their location. Most of the businessmen (but not only) can't imagine working without multiple phone numbers: a personal number to contact family, an office phone number to contact coworkers and a business number to communicate with customers.

The idea of easy contact for business users is one of the principles of Unified Communications services – a very popular communication trend in enterprise area. Typically it is implemented with:

- One number service – functionality allows the user to define and manage preferred communication devices via voice and video. Despite of using different terminals, subscriber is identified by one universal telephone number.
- Integrated application – single application supporting in a unified way many different communication channels: voice, video, e-mail, voice mail, Instant Messaging and web collaboration.

Efficient usage of Unified Communication systems and their features requires interaction with the end system user. Usually a user has to pay attention to and consciously set his or her status in UC application (ready, out of office, on holiday etc.), set and activate auto response in corporate e-mail system or set an active phone (office phone, mobile phone, softphone etc.).

From the end user the perspective, these processes preferably should be automated to ease him/her of manual setting of an active phone. All information required for Unified Communication system can be acquired by sensors or software components taking advantage of context information.

The idea of automatic information exchange between devices and sensors is similar to the smart home automation concept, and M2M systems are a very promising source of context information e.g. for potential commercial systems.

The main aim of this paper is propose method to store in one database context information characterized of various data structure from different mentioned above sources (UC, Telco and M2M systems).

978-83-60810-53-8/$25.00

II. BACKGROUND

A. Existing solutions

Many context-aware computing paradigms have been developed in the recent years. For example, Cisco context-aware Healthcare solution, which is a tool for monitoring and simplifying business operations in hospitals [3]. However the most popular context-aware solution is currently the Google Ads Gadget, which displays context based personalized commercials to users browsing a specific website [4].

Cisco solution [3] uses RFID [5] (Radio Frequency Identification) technology for real time tracking of patients, medical personnel and hospital assets, what is crucial for providing the best quality health care services and optimization of workflow and medical staff management. The uniqueness of the approach stems from the fact that Cisco health care solution treats equipment and machines, such as X-ray or wheelchairs, as objects which are considered as the source of the context information. Context aware systems also bring new features into the survey conducting area [6], [7], [8], [9]. In this field, gathering context data and taking advantage of data collected by mobile devices, such as smart phones or tablets, allows developers to invent useful mechanisms which in turn allow to specify the exact target group of a survey. A solution presented in *A survey on context-aware systems* [6] uses a variety of many flexible sources, including widgets, wireless sensors and middleware infrastructure, which is somewhat similar to the mechanism presented in this paper.

B. Unified Communication

Unified Communication [10], [11] systems constitute an important element of private communication networks dedicated to enterprises. Many vendors (e.g.: Avaya, Cisco, IBM, Microsoft, Oracle and Siemens) offer comprehensive UC solutions. An example of such system developed by Siemens Enterprise Communications is OpenScape UC application.

UC system offers many features, e.g. standardization of communications services, support for user mobility, communication medium neutrality and support for any type of equipment (type of terminal). Possibility of virtualization of resources and services and their exposition by using LAN, WAN or Internet networks is of utmost importance for business users. Openscape UC (Fig. 1.) offers VoIP (Voice over IP) services (implemented in the OpenScape Voice subsystem), video communication between employees (OpenScape Video) and integration of e-mail and voice in a single message box (OpenScape Messaging). Integration of external applications and systems with OpenScape system is possible due to availability of Application Programming Interfaces (APIs), implemented in the UC system. Its APIs are exposed as Web Services and are offered using a Service-oriented architecture model (SOA) and SOAP Protocol (Simple Object Application Protocol). By taking advantage of UC open APIs, external developers can create their own specific application and UC system extensions described in [11] and [12].

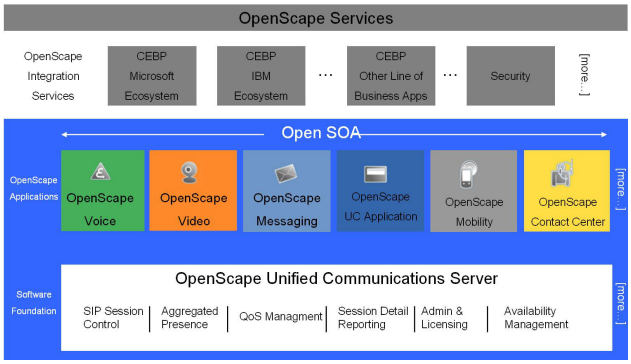


Fig 1.OpenScape UC system architecture [11]

C. Telco 2.0

Telco 2.0 [13] is an idea based on exposition of Communication Service Providers APIs for external developers and third party companies on the Web. From the technical point of view, telecom operator's APIs are exposed using a dedicated entity – a Service Delivery Platform (SDP). An SDP is located between the operating network and the Web. Its southbound interfaces are connected to the operating network and integrate it with enablers offering telecommunications capabilities (e.g.: SMS, MMS, USSD, voice call, etc.). The northbound interface exposes open APIs as Web Services on the Web. The SDP is also connected to an Operations Support Systems responsible for: maintenance, inventory, provisioning and fault management as well as Business Support Systems (responsible for order handling, billing, payments etc.). There are two ways of exposing Telco 2.0 APIs – either via SOAP [15] interfaces or by using REST architectural style [14]. Telecommunications service delivery platforms (SDP) usually allow access to their capabilities in the service-oriented architecture (Parlay X specification [16]) or resource oriented Web Services (RESTful) model compliant with the OneAPI specification [17]. By using open Telco 2.0 APIs, it is possible to develop new applications and services merging IT and the Internet world with telecommunication area [12], [18], [19].

D. M2M approach

The Machine to machine (M2M) concept refers to technologies that allow both the wireless and wireline systems to communicate with other devices with the same capabilities. M2M uses devices, such as e.g. sensors or meters, to capture events or current state indicators (temperature, inventory level, etc.), which are then relayed via a network (wireless, wireline or hybrid) to an application (software

program), that translates them into meaningful information (for example, items need to be restocked) [20]. Such communication was originally implemented using a remote network of machines which relayed information back to a central hub for analysis, which would then be rerouted into a system like a personal computer.

III. System Architecture

Our solution integrates three systems as a context data source (Fig 2). The first data source is the Unified Communication system. Using the SDK (Software Development Kit) provided by Siemens, it is possible to implement a service which captures events reflecting subscriber's activity, e.g. phone status and change of user status. The Openscape UC API enables implementation of the click to call feature as a dedicated button at the website. Web Services exposed by Communication Service Provider in the Telco 2.0 model allow collecting context data from mobile phones (e.g. using terminal status and terminal location APIs). The third source of the context information are events generated by M2M devices. This last area is a comprehensive source, which can send information of all type (presence, alarms, user activities etc.).

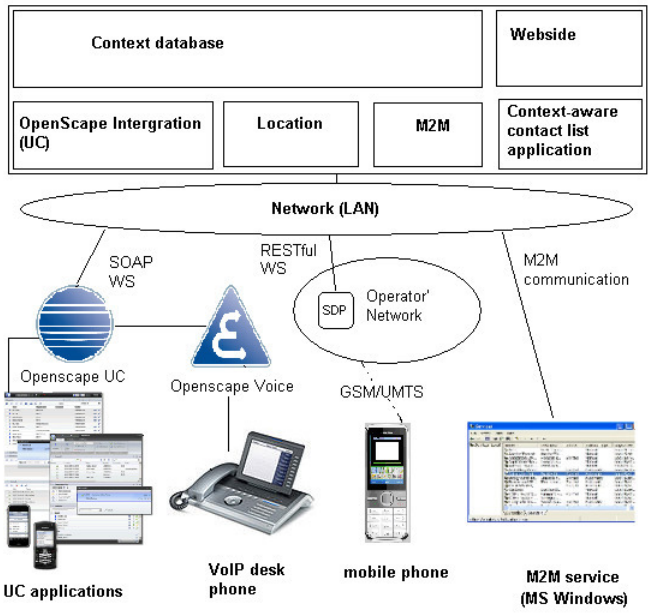


Fig 2. Architecture of system

This chapter presents the main modules comprising the system collecting context information from different sources.

A. Database

The data base is the most important part of our solution. It plays the role of the data repository collecting context records from three mentioned above sources. The data base has to be: scalable, easy to access and fast. Meeting these requirements is easy in case of small databases but must be relaxed in case of large systems handling thousands of records.

The implemented MySQL database uses a relational data model and specifies four main entities (Fig 3.):

- *User* – entity stores information about the user: login information for OpenScape management portal, name, surname, Openscape VoIP terminal and mobile terminal numbers as well as the home address.
- *Locations* – entity stores known addresses of the user for example home address and work address.
- *Context_data* – stores context information: timestamp and context data set which is JSON form of context information (for example latitude and longitude).
- *Context_type* (Fig 4.) – a table which contains information about the context data source.
- *Param_type* – a table which assists in detection of the type of parameters coded in context data set (JSON).

The presented database schema guaranties to be scalable – e.g. there is no need to add new table when the system is integrated with a new context information source in the future. Data sets are easy to access because of the usage of JSON [21] format, which is supported by many programming environments.

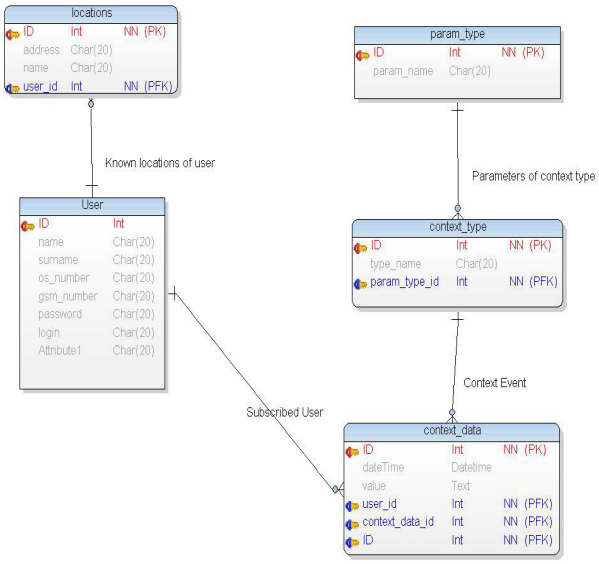


Fig 3. ER diagram of database

id	name	paramtype_id
1	openstageMediaStat	4
2	openstagestat	5
3	gsmlocation	1
6	systemlogin	6

Fig 4. Context type entity

B. Location module

The location module is responsible for collection of information about a mobile phone's geolocation. The location data is collected using terminal location API calls, based on those exposed in the Web by Orange using a dedicated Service Delivery Platform (SDP). The response received from *Terminal Location API* contains approximated latitude and longitude of the mobile terminal. The location information is stored in the database using pooling method. The measurements are repeated every minute. Unfortunately, the pooling method is not very efficient in terms of performance, but currently there aren't any other methods available for small developers and external companies that could be exposed on the Web by Communication Service Provider in the Telco 2.0 model.

C. OpenScape integration module

Siemens provides a high level Software Development Kit for programmers, which allows implementing a method responsible for integration with a UC system. Its main role considered here is implementation of a method, which captures events in relation to the change of media (terminal) status or user status in UC system. There are three possible states of media status: 0. *UNKNOWN* (when the observing user is not authorized to see the recipient status) 1. *AVALIABLE* (when there is no call on the phone line) and 2. *UNAVAILABLE* (when the user is already busy in a call). The events, like e.g. change of the status are captured and stored in the database with a timestamp and an appropriate JSON value.

The second parameter correlated with the user activity by the presence of API is the user status. There are 6 possible states of user status: 1 - *Do not disturb* 2 - *Be right back* 3 - *Unavailable* 4 - *Busy* 5 - *In meeting* 10 – *Available*. This parameter is stored in the database – when an event (triggered by the change of the status) appears. The user status information is stored (in the JSON format) with a timestamp in the database.

Another method implemented in this module is the *click to call* feature. This method requires user login and password information to recognize an active phone from which the call originated.

D. M2M module

This application was implemented in Windows Service form which starts with the user's login and ends with the system user's logout. The module collects information about the user logged on to the workstation. The name of the user is sent to the main system and is stored in the database, using timestamp in JSON format.

E. Module Website

Module Website – the graphical user interface was created using ASP.NET framework. The module contains 4 subpages: *login.aspx*, *contactlist.aspx*, *adduser.aspx* and *about.aspx*. Only a successfully authorized user is able to use the portal. This module implements login and password; the ones required by portal are the login and password from OpenScape UC system. If the logged user is also the administrator, he or she is able to manage the system's administration panel, e.g. add or remove users to the system, (Fig. 4) define work address and home address.

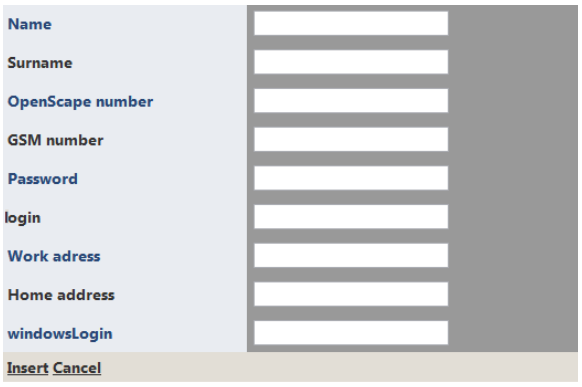


Fig. 4. Module Website - add user form

To add a user, all fields are required, with the condition of setting a correct login and password (same as OpenScape UC). The ending user application of the website is *contactlist.aspx* which contains all contacts, context data and the click to call button (Fig. 5.).

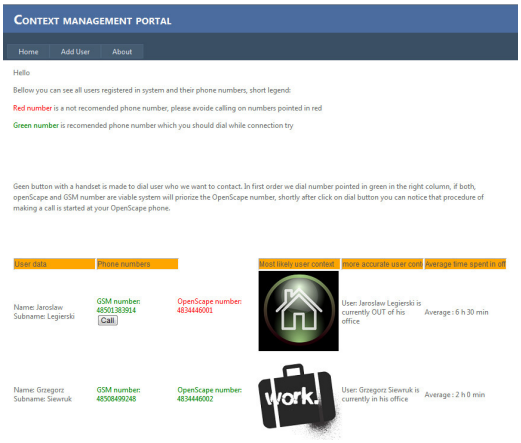


Fig. 5. Context-aware contact list application

F. Context-aware contact list application

The contact list application contains user information, such as: name, surname, mobile terminal number and OpenScape UC number (VoIP office phone). The database also contains location information, such as the user's home address and work address. The application based on the context data can recognize the actual location of the user (using Telco 2.0 and UC data) and can present more specific context data e.g. if the user is at work, it shows the average time spent in the office (M2M based information). The application can even suggest the number of users recommended to contact them. The website also has

a "call" button which allows dialing the recommended phone number using the OpenScape UC system.

IV. Simple Context Based Decision Algorithm

This chapter, based on the contact list application features, presents a simple context-based decision algorithm. The first column of this application (Fig. 5) contains contact data, such as the name and surname from the database. The next column: "Most likely user context" is a field based on information acquired from the Telco 2.0 location API.

The application uses Google Map API to recognize the home address and work address based on geolocation (longitude and latitude), which is recognized by using the distance between the user's current location (the latest database record) – and simultaneously between the current location and home location. If the distances calculated above are less than 1000 m, the algorithm determines that the user is at the defined location (work or home).

The next column - "More accurate user context" - contains information from data collected by the M2M application. If the user is already logged on to the workstation, the algorithm decides that the user is in office and vice versa: if the user is not logged on to the system, the result is interpreted as "out of office".

The column "Average time spent in office" is based on data received from the M2M application. The application measures the time from the timestamp between login and logout, and calculates an average value.

The last column - "Phone numbers" contains all phone numbers defined for a specific user. If the phone number is underlined in red – the number is not recommended to dial, when it is in green– the number is recommended for use.

There are several conditions that must be met for the number to be underlined in red: for the context information from OpenScape UC – *AggregatedMediaStatus* attribute must have value *Unavailable*; *userStatus – In Meeting*, *Unavailable*, *Busy* or *Do not disturb*; however, if *AggregatedMediaStatus* is Available but the user is out of office, the number will also be underlined in red.

Finally, by clicking the "call" button, the user can dial the number which is underlined in green. When all numbers are highlighted in green, the number allocated to Openscape system has higher priority.

V. Information Integrity and Security

A. Context Conflicts

There is a possibility that pieces of the context information which are received from different sources may be inconsistent or conflicting. Situation like this can happen when the subscribed user by mistake leaves his mobile phone at home. In this case M2M module has discovered user as logged in and present in his office however Telco 2.0 module is indicating that the user is at home.

It is extremely important to deal with this kind of situations properly because any inconsistency has impact on reliability of the presented solution. For example in the scenario described above, a source which can authenticate user is selected as the stronger one and other conflicting information will be ignored (however the user will be informed that there is something wrong with his devices). In order to be able to use M2M module, the user has to be logged in to his computer by providing a correct password or plugging a USB token with a certificate.

Submitted solution is efficient enough but can be improved by using semantic algorithms.

B. Information Security

As it is described in the previous chapters presented system deals with sensitive data and implementation of methods preventing data loss is of utmost importance.

There are several mechanisms to perform such tasks both at the application side and at the server side. A list of basic requirements to provide security for this kind of environment is as follows.

Application:

- enforce users to use strong passwords (Uppercase letters, special signs, numbers),
- enforce users to change their passwords every 30 days,
- user roles which give the certain user possibility to perform only allowed operations.

Server:

- up to date antivirus software,
- configured firewall,
- database calls can be made only from the specific IP addresses (website).

In Context-aware contact list application requirements for strong passwords and regular change of password is possible to achieve only if administrator of OpenScape system implement such policy in OpenScape panel (password for application and OpenScape is the same).

Application is immune to the most common SQL-injection and Cross Site Scripting (XSS) attacks.

VI. Measurements

A. Performance tests result

In order to verify the SQL database structure presented in chapter III, some performance tests had been run. The proposed database scheme was compared with a classical database structure based on 5 independent tables (user, *gsmlocationcontext*, *openscapeaggmediacontext*, *openscapestatuscontext*, *m2mcontext*) and filled with context information.

The first test presents time needed to load all the context datasets to the memory. In order to get all the information about a specific user, 4 SELECT statements were executed, whereas getting all the possible information in the presented solution required only one SELECT statement. The results are presented in Table 1.

TABLE 1
. SQL QUERY RESPONSE TIME (4 CONTEXT SOURCES)

Select all context			
Context-aware contactlist database		Standard context database	
Table length [bytes]	SELECT duration [s]	Table length [bytes]	SELECT duration [s]
500	0,00501125	500	0,047842
1000	0,0101115	1000	0,055079
1500	0,01409725	1500	0,206002
2000	0,0230885	2000	0,179314
2500	0,02743775	2500	0,275252

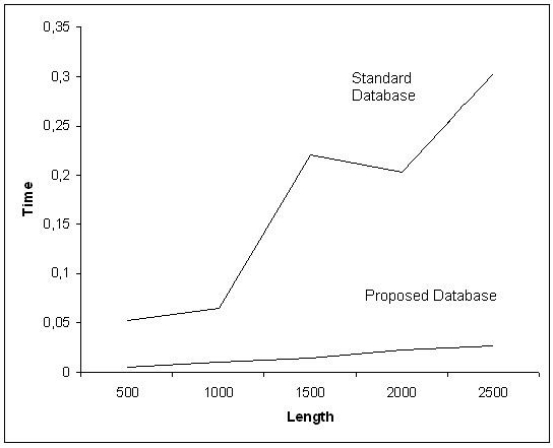


Fig. 6. SQL query response time. (4 context sources)

TABLE 2.
SQL QUERY EXECUTED FOR THE PROPOSED DB SCHEME (4 CONTEXT SOURCES)

```
SELECT * from contextdata where user_id =1
```

TABLE 3
SQL QUERY EXECUTED BASED ON STANDARD SOLUTION. (4 CONTEXT SOURCES)

```
SELECT * from gsmlocation where user_id=1; SELECT * from openscapestatus where user_id=1; SELECT * from openscapeaggmediastatus where user_id=1; SELECT * from m2mcontext where user_id=1
```

The results presented above show dependency of the execution time on the table length. The proposed database structure is more efficient because it decreases the number of SQL Select statements. The execution of one SELECT statement is faster than the execution of four SQL Select statements.

In the next test, the performance of a database based on the information from only one context source was measured. The SQL queries used during the test are presented in Tables 4 and 5 respectively.

TABLE 4.
SQL QUERY EXECUTED FOR THE PROPOSED DB SCHEME (1 CONTEXT SOURCE)

```
SELECT * from contextdata where user_id =1 and cotnexttype=1
```

TABLE 5.
SQL QUERY EXECUTED FOR THE STANDARD SOLUTION (1 CONTEXT SOURCE)

```
SELECT * from gsmlocation where user_id=1
```

The results of the measurement are presented in Table 6.

TABLE 6.
SQL QUERY EXECUTION TIME (1 CONTEXT DATA SOURCE)

Select one context			
Context-aware contact list database		Standard context database	
Table length [bytes]	SELECT duration [s]	Table length [bytes]	SELECT duration [s]
500	0,011354	500	0,019467
1000	0,020775	1000	0,02147375
1500	0,02485975	1500	0,04349075
2000	0,0322245	2000	0,05073
2500	0,04947975	2500	0,05930075

Using the proposed context notation in the JSON format results in the execution time of SQL SELECT function which is shorter than in case of the standard notation. It should be emphasized that in the presented solution, the table has always 4 columns and this value is independent from the number of parameters represented in the context information (context data source).

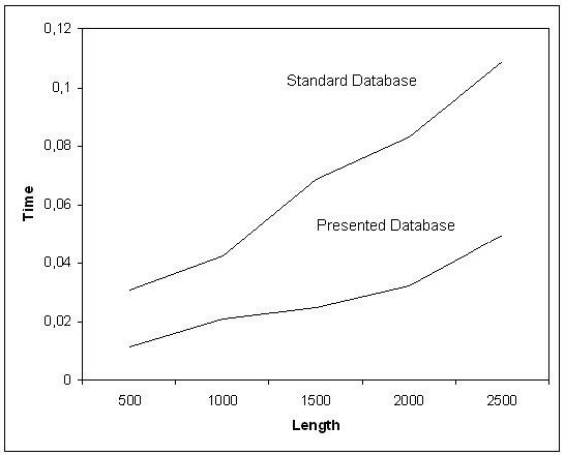


Fig. 7. SQL query response time. (1 context data source)

B. Functionality test results

Figures 8 and 9 present the results of the functional tests. The maps present the location of selected points where the service was invoked, and show the location received by using the Telco 2.0 location API as well as the potential location based on the user's activities in the UC system.

Blue arrow - real location of the user.

Green arrow - location based on the information from UC system.

Red arrow - location based on the mobile network.

For the presentation of the test results, Google maps application was employed. Figure 8 shows the case in which the user forgot to change his or her UC status, and left the office. Based on the mobile terminal location services, the presented context aware system can, in this case, change the status of the user in the UC system automatically. Figure 9 presents the approximate range within which the user is considered to be in his or her office.

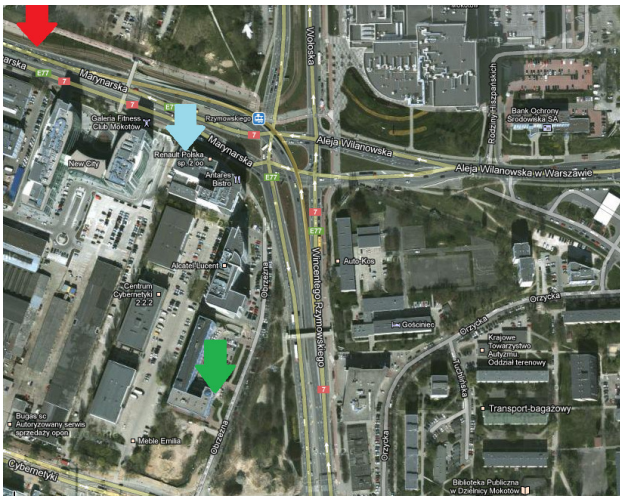


Fig. 8. Sample result – "user out of office"

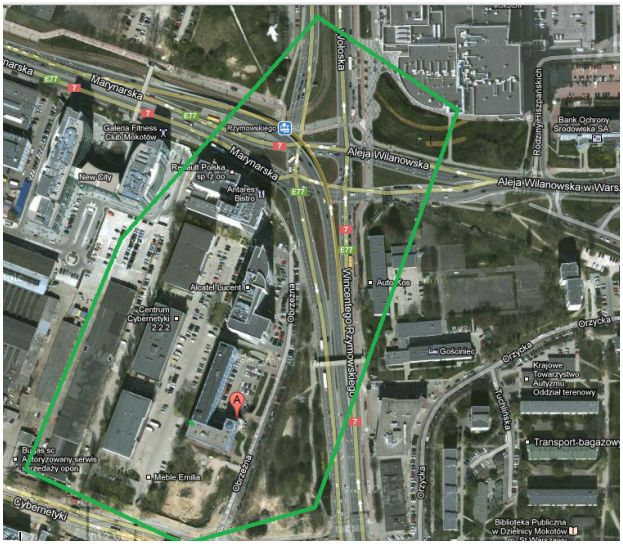


Fig. 9. "user in the office area " - approximate range

When the user passes the border line highlighted in green, he or she can be recognized by the context application as being out of office.

VII. Summary

The architecture of the context aware application presented in the paper features easy access to the context data sets stored in the database. The context information is stored using the JSON format, which is very popular in the IT world.

The main challenge was development of the structure of the database which had to meet many criteria described in the system architecture (chapter II). Further work on the described system should be focused on the usage of NoSQL databases as a data repository.

The presented solution is, first of all, flexible, ready and easy to extend. The most important aspect is the possibility to add another source of context just by inserting a record into the database with a new context data definition.

The functional test results presented in chapter IV show the location error in mobile networks using Terminal Location API and their impact on the context aware application functionality. This error can be minimized using more accurate location algorithms by communication service providers or the implementation of other location methods e.g. based on GPS.

The prototype of the context aware application has been developed under the Orange Labs Open Middleware 2.0 Community program [22] as a part of Grzegorz Siewruk' B.Sc. Thesis.

References

[1] A.C, Weaver, B. B. Morrison, Social Networking, Computer, Volume: 41 , Issue: 2, 2008

[2] C. Pettey and H. Stevens, "Gartner Says Context-Aware Computing Will Provide Significant Competitive Advantage," Gartner Press Releases, 2010. [Online]. Available: http://www.gartner.com/it/page.jsp?id=1190313 [20.05.2013]

[3] The Cisco Location-Aware Healthcare Solution , http://www.cisco.-com/web/strategy/docs/healthcare/CLA_HealthcareSolution.pdf . Cisco Systems, Inc. 2007 [20.05.2013]

[4] Magnus Yang, Introducing Google Gadget Ads, Google Inside AdSense Blog http://adsense.blogspot.com/2007/09/introducing-google-gadget-ads.html [20.05.2013]

[5] Stephen A. Weis, RFID (Radio Frequency Identification): Principles and Applications, MIT CSAIL http://www.eecs.harvard.edu/cs199r/readings/rfid-article.pdf [20.05.2013]

[6] Matthias Baldauf, Int. J. Ad Hoc and Ubiquitous Computing, A survey on context-aware systems, Vol. 2, No. 4, 2007 pp.263-277 ,

[7] B.Chien, H. Tsai, Y.Hsueh, " CADBA: A Context-aware Architecture based on Context Database for Mobile Computing" Autonomic and Trusted Computing, 2009. UIC-ATC '09

[8] A.Srinivasan, G. Irwin, "Communication the Message: Translating Task Into Queries in a Database Context", IEEE transactions on professional communication vol 49, no.2 june 2006

[9] Bauer, M.; Kovacs, E.; Schülke, A.; Ito, N.; Criminisi, C.; Goix, L.-W.; Valla, M., "The Context API in the OMA Next Generation Ser-

vice Interface," Intelligence in Next Generation Networks (ICIN), 2010 14th International Conference on , vol., no., pp.1,5, 11-14 Oct. 2010
[10] Siemens to Reveal New LifeWorks Vision and Showcase Advanced IPBased applications and Services, SUPERCOMM 2003
[11] D. Bogusz, P. Korbel, J. Legierski, Integracja systemów Unified Communications z platformami usługowymi operatorów, KSTiT 2011 conference materials, Przegląd Telekomunikacyjny 8-9/2011
[12] Bogusz, D.; Legierski, J.; Podziewski, A.; Litwiniuk, K., "Telco 2.0 for UC — An example of integration telecommunications service provider's SDP with enterprise UC system," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.603,606, 9-12 Sept. 2012
[13] PORTAL STL PARTNERS: http://www.telco2.net [20.05.2013]
[14] L. Richardson, Sam Ruby, David Heinemeier Hansson, RESTful Web Services, O'Reilly, 2007
[15] SOAP specification W3C, http://www.w3.org/TR/soap/ [5.03.2013]
[16] Open Service Access (OSA); Parlay X web services, 3 GPP, http://www.3gpp.org/ftp/Specs/html-info/29199-01.htm, [5.03.2013]
[17] GSMA OneAPI http://www.gsma.com/oneapi/ [5.03.2013]
[18] Litwiniuk, K.; Czarnecki, T.; Grabowski, S.; Legierski, J., "BusStop — Telco 2.0 application supporting public transport in agglomerations," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.649,653, 9-12 Sept. 2012
[19] Podziewski, A.; Litwiniuk, K.; Legierski, J., "Emergency button — A Telco 2.0 application in the e-health environment," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.663,677, 9-12 Sept. 2012
[20] M2M: The Internet of 50 Billion Devices" , WinWin Magazine, January 2010
[21] Crockford, Douglas (May 28, 2009). [21] "Introducing JSON" . json.org. Retrieved July 3, 2009.
[22] Open Middleware 2.0 Community portal – http://www.openmiddleware.pl [20.05.2013]

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 859–864

Mobile Payment System – Telco 2.0 application dedicated for payments

Piotr Trusiewicz
Warsaw University of Technology
Faculty of Mathematics and
Information Science
ul. Koszykowa 75
00-665 Warsaw, Poland
Email: trusiewicz.piotr@gmail.com

Maciej Witan
Warsaw University of Technology
Faculty of Mathematics and
Information Science
ul. Koszykowa 75
00-665 Warsaw, Poland
Email: maciej.witan@gmail.com

Marcin Kuzia
Orange Labs
ul. Obrzeżna 7
02-691 Warsaw, Poland
Email: marcin.kuzia@orange.com

Abstract — **The following paper presents Mobile Payment System. The system is a prototype of an innovative method of paying for services using the mobile phone. The method is quite straight-forward, basically the user wanting to access some online service supplies his/her cellular phone number to the web form and receives a token via USSD message (Unstructured Supplementary Service Data). Then, introducing the token into the web form gives the user an access to the desired content. At this exact moment the charging is done. The due amount of money is simply added to the monthly bill, in case of postpaid phones, or subtracted from available credits, in case of prepaid mobile phones. The functionality of sending USSD messages from the system to the subscriber mobile phone was achieved by using Telco 2.0 Web Services provided by Orange Labs.**

I. Introduction

As it is observed, nowadays in the global web there are more and more paid services. There are a numerous portals which require registration fees, an access to some scientific articles is restricted until it is paid for. There is of course possibility of buying ticket for a concert or other cultural event and a lot of other services not mentioned here. Some of the services have a great number of users and other not so many. Also as it is commonly known, everybody wants get the access in the shortest time and the easiest way possible. The need therefore arises: quick, easy mobile payment method with no scalability problems.

A. Existing solutions

At the moment there already exist some payment methods like [11]:

- premium-rated SMS
- credit cards
- pay-pal (and alike systems)

Unfortunately each of described above system has some limitations. Premium-rated SMS requires a huge number of users per month to be cost-effective. The problem with credit cards is such that not everybody has one, another is that for some people it may seem risky to send sensitive credit card data over the internet which will discourage them from using this method. Pay-pal and pay-pal-like systems require registration and uses dedicated account which is time consuming and can act like a discouraging factor.

B. Migration from Telco 1.0 to Telco 2.0

Telco 1.0 is about value-added services (VAS) [1] which are all services that are beyond voice calls or fax transmissions. They, as the name indicates, add value to the standard services. Historically, SMS, MMS or data access were considered VAS, however nowadays they are standard services. In that approach only the operator could create new added telecommunication services. The reason for that was the fact that such services used intelligent network platforms (telecommunication platform with centralized management), which were accessible only by the operator. Since a few years migration from Telco 1.0 to Telco 2.0 [2][2] can be observed.

In Telco 2.0, an additional layer in the Intelligent Network Platform architecture has been created. It has been added to make a creation of other VAS possible by parties other than the operator alone. In practice, the operator offers an access to the basic telecommunication functionalities (like sending and receiving USSD, SMS, MMS messages, managing phone calls, locating terminals etc.) in form of easy to use web services. The communication with those services is possible using the SOAP and REST [3][3] protocols. The conceptual difference between Telco 1.0 and Telco 2.0 approaches is

978-83-60810-53-8/$25.00

visualized in the Fig.1. The architecture of the Telco 2.0 solution is presented in Fig. 2.

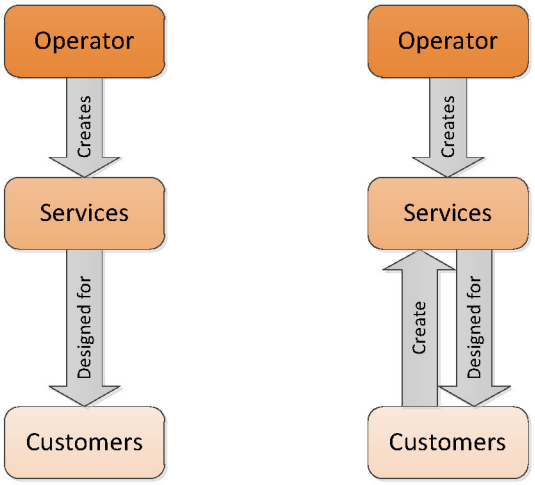


Fig 1. Telco 1.0 vs. Telco 2.0 concepts[7][7]

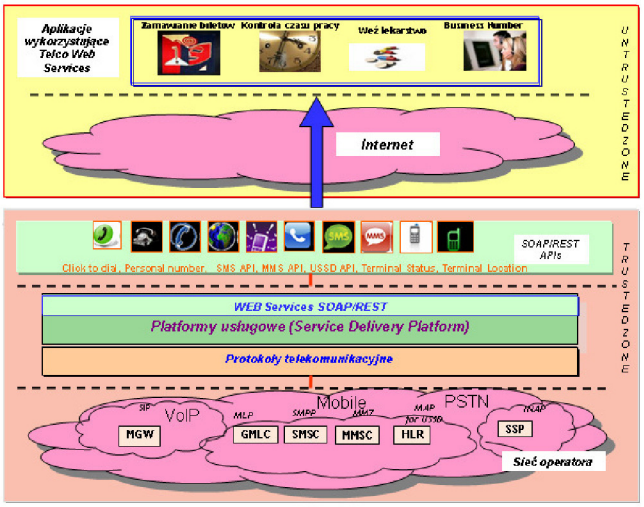


Fig 2. Telco Web Services architecture [10][10]

C. Profits of using Telco 2.0 to solve the presented problem

The great advantage of Telco 2.0 is its the range of usage. It is not restricted to the users of smartphones only but it is accessible by all mobile phone users within coverage area which is over 95% of the inhabitants of Republic of Poland. It implies that the service can be launched from any location covered by the radio access network of the native mobile telecommunication operator. During recent years there were developed a number of applications based on operators networks assets. Some examples are described in literature [6][6], [7][7], [8][8], [9][9], [10][10].

II. THE PAYMENT SYSTEM

Mobile Payment System is a prototype of an innovative method of paying for services using the mobile phone. The system presented in this paper addresses all the arisen needs: it is fast, easy and has no scalability problems.

Implemented solution of the method is just a prototype that should be treated as a proof of concept rather than ready to deploy product. System functionalities are limited to granting access to specific URLs. Apart from that the management panels has been implemented in form of web application.

Fig 3. The schematic of service usage[7]

The method is fairly simple, the user wanting to view restricted content supplies his/her cellular phone number to the web form and receives a token via USSD message. Then, introducing the token into the web form gives the user an access to the desired content. Also at this moment the charging is done. The due amount of money is added to the monthly bill, in case of post paid phones, or subtracted from available credits, in case of pre-paid mobile phones. This is, of course, just the main idea of how the system works. The system itself is more complex. It contains also user account creation, generation and transmitting the account password to the client via SMS, storing the history of subscription, administration panel (managing of 'non-admin' users). Also, the service is protected from unauthorized use thanks to basic authentication mechanism. The functionality of sending SMS and USSD messages from the system to the subscriber mobile phone was achieved by using Telco 2.0 Web Services provided by Orange Labs. These are ready to use services that provide above mentioned functionalities. The whole system was implemented in .NET technology using C# programming language. Entity Framework was used to facilitate all necessary database operations, technology for database is MySQL. The communication with Web Services is

possible with SOAP protocol. The implementation of user interface uses ASP.NET MVC 3 Framework.

III. SYSTEM ARCHITECTURE

A. Functional schematic of the solution

The functionality of the solution is presented in Fig. 4.

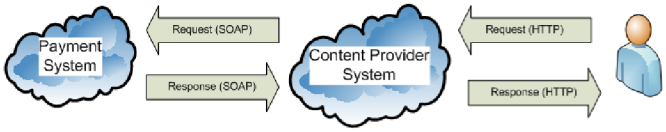


Fig. 4 Functional Schematic of the solution

The end user makes HTTP GET request to Content Provider System if the user is not authorized to view the content SOAP request is made to Payment System where the user is charged. Response is sent to the end user through Content Provider System.

B. Used Telco 2.0 interfaces

The system is based on the functionalities offered by mobile phone networks. SMS API, USSD API and GetOperator API have been used. GetOperator API is used to check if the number provided by the user belongs to Orange operator. SMS API and USSD API provide a possibility to send a message and check the status of the message. SMS API is used in the system to send the generated password to the user. USSD API is used to send generated access token to the user's mobile phone and send notification about charging.

C. Structural architecture of Mobile Payment System

The solution presented in this paper consists of 2 separated systems: Payment System simulating payments and Content System providing resources.

The Payment System (Telco) is responsible for exposing SOAP Web Service providing Payment API functionality. It consists of three parts based on the performed functions shown in Fig 5.

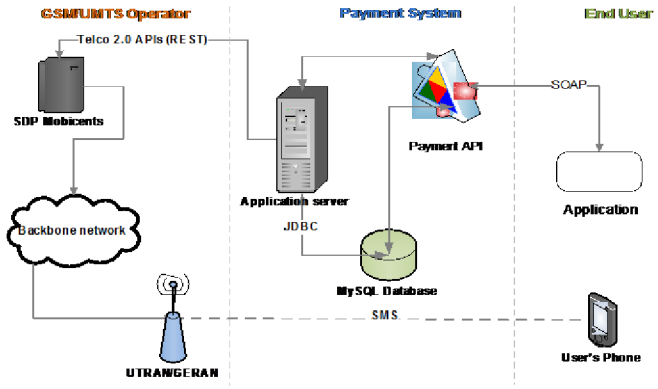


Fig 5. Structure of developed Payment System

• End User – communicates with the system via UTRAN/GERAN access network using mobile phone or SOAP web service.

• Payment System – the logic of the system consists of two applications. First one is the application implementing Web Service and the other is Web application which provides the user interface for database administration. The communication between the system and database is done using Entity Framework. The database stores information about accounts, transactions and web service users. The application uses Telco 2.0 interfaces to send passwords to the user's mobile phones.

• GSM/UMTS Operator – enables the communication with the developed system through exposed Telco 2.0 interfaces.

The second module - content system is responsible for providing paid content to the end user. It consists of three parts based on performed functions shown in Fig. 6.

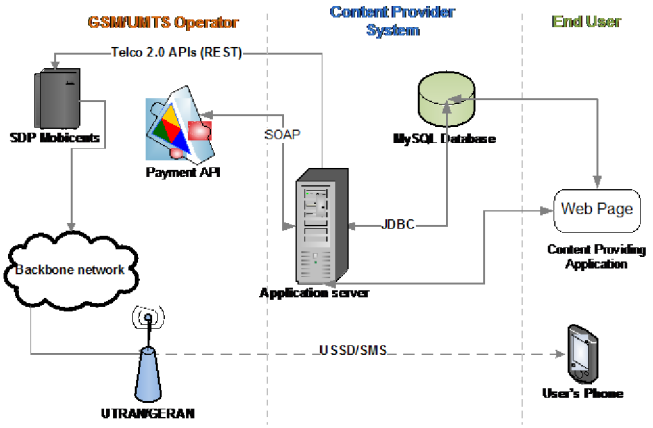


Fig 6. Structure of developed Content System

• End User – communicates with the system via UTRAN/GERAN access network using mobile phone or Web Browser.

• Content Provider System – the logic of the system is implemented as a web application, which provides the user interface for the Content Provider application. The communication between the system and database is done with use of Entity Framework. The database stores information about users, URLs and subscriptions. The application uses Telco 2.0 interfaces to send token to the user wanting to access chosen URL.

• GSM/UMTS Operator – enables communication with the developed system through exposed Telco 2.0 interfaces.

D. Class Diagrams

Presented solution was developed in Model-View Controller (MVC) style. It consists of Controller, Model, ORM and Helper classes. All interesting classes are presented below. Views are simple dynamically generated HTML. Classes in the system:

- Controller classes – responsible for handling HTTP requests. GET and POST requests are mapped to appropriate method called action from the class. Diagram is presented in Fig. 7,8
- Model classes – represents the data objects send within system. Each class consists of properties and the setter/getter methods for each.
- ORM classes – represents the database tables mapped into C# objects. Entities are generated by Entity Framework.
- Helper classes – are responsible for database operations, handling AJAX, sending response to mobile phone end user. Presented in Fig. 9, 10

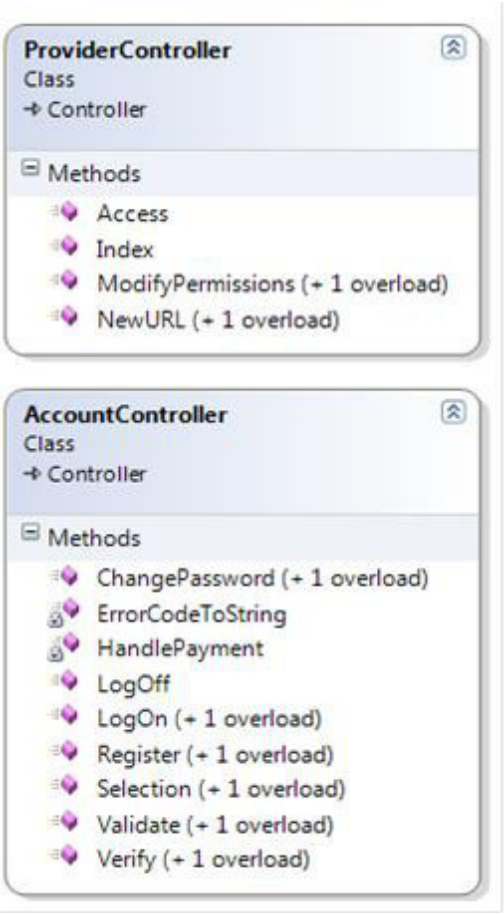


Fig 7. Controller classes

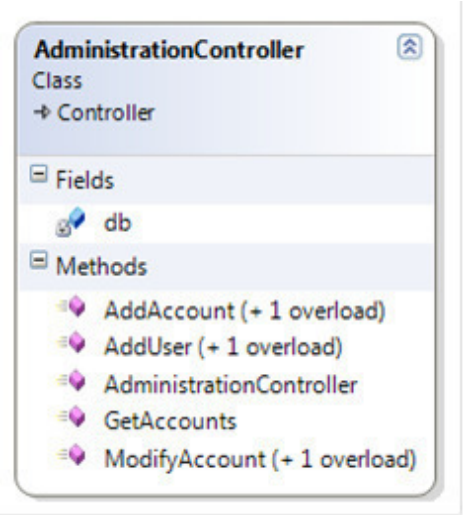


Fig 8. Controller class

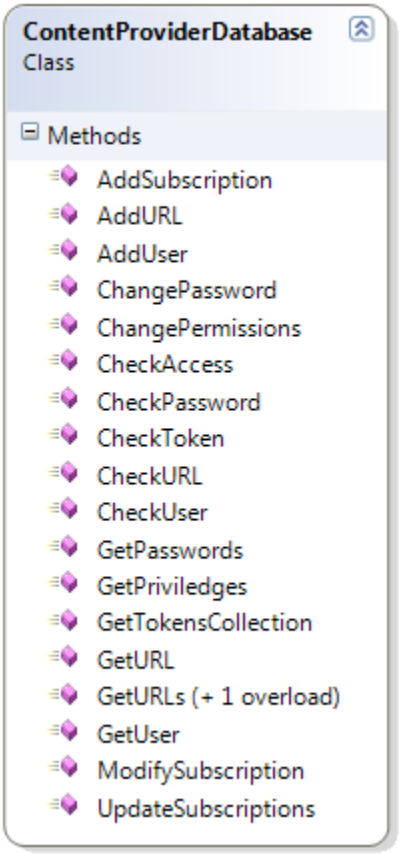


Fig 9. Helper class

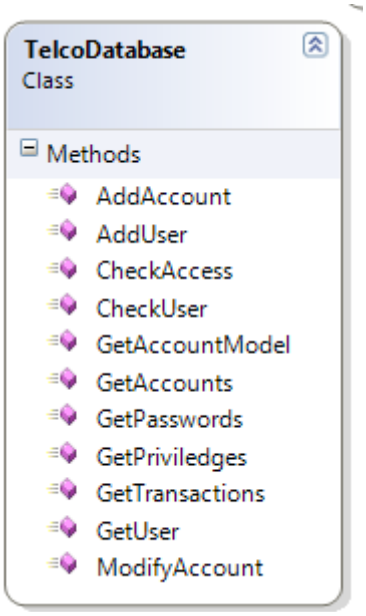


Fig 10. Helper class

IV. USER INTERFACE

The Mobile Payment System was developed using C# programming language and ASP.NET MVC 3 framework. The system has a simple, intuitive web interface. JavaScript was used to create the user interface, Microsoft Internet Information Server was used as an application container. Some interesting views, that user may find in the system, are presented below. Implemented solution supports both types of resources: local and external.

List of URLs

wp.pl

onet.pl

mini.pw.edu.pl

google.pl

allegro.pl

localhost:8080/ContentProvider/Resources/house.jpg

localhost:8080/ContentProvider/Resources/koala.jpg

localhost:8080/ContentProvider/Resources/penguins.jpg

Fig 11. List of supported resources

Presented in Fig. 11 shows the list of supported resources. It can be easily extended by administrator through appropriate web form. User can check the list of his/her subscriptions Fig. 12.

List of subscriptions

URL	Expiration Time
localhost:8080/ContentProvider/Resources/house.jpg	2013-05-16 10:57:04
google.pl	2013-05-12 10:59:55
en.wikipedia.org/wiki/Speech_recognition	2013-05-12 11:06:40

Fig. 12 List of subscriptions

Web application allows to check the transactions made through the system. The exemplary transactions list is presented in Fig 13.

List of user transactions

ID	Amount	Reference Code	State	Time
1	2,3	7fa77935-882d-4792-a03f-dd8a3ec193c4	CHARGED	2013-01-07 13:21:59
2	2,3	6b537d50-3071-4e91-880f-28d7658bbab0	REFUNDED	2013-01-07 13:24:49
3	1	c9e61089-8858-43f9-b9ed-afcb77f398c9	CHARGED	2013-01-07 14:37:39
4	10	76ef3560-5b48-4dae-9865-1aafc80a43e2	CHARGED	2013-01-07 15:47:30
5	1	5b40592d-b901-4d89-a419-2e6ef4f128a1	CHARGED	2013-01-07 15:50:04

Fig 13. User transactions

V. CHALLENGES

The system that has been developed is just a prototype. The most important part of it is the Payment service. Due to the prototype version, the service provides only the basic functionality and has been secured merely with HTTP Basic Authentication without SSL. Such a solution, in fact, is not secure at all. Hence, in future, the thing of utmost importance will be upgrading the security level of the system. Since the service is responsible for financial operations it must be protected by a top-level security system. The functionalities might also be extended. The current version allows only to: charge account, refund account, check transaction status and get transactions list. The possible directions of development are to extend an existing reservation or release a reservation. Payment service is an XML Web Service supporting exclusively

SOAP protocol. Hence, another enhancement could be handling RESTful architecture style with JSON data serialization support. The advantage of REST protocol is that it is very lightweight and uses normal HTTP methods instead of heavy XML format. Both parts of the system: Telco and Content Provider have very simple user interfaces, designed just to present the system capabilities. This is however, another possibility of enhancement. At last, the mobile version of the system may be created as well.

VI. CONCLUSION

Presented in this paper the prototype of the innovative Mobile Payment System shows the possibility of usage of telecommunication open APIs in very wide mobile payment area. The main goal of presented research was to create the payment service and simulate it using Internet. All requirements were considered when designing system's architecture. The created user interface is simplistic and should be treated rather as a proof-of-concept. Tests showed, that all functional requirements are satisfied. The End User tests and Orange Labs experts comments provoked some changes that made the system more user-friendly.

Prototype of the System service was developed under the program Open Middleware 2.0 Community [5] as a part of Piotr Trusiewicz and Maciej Witan B.Sc, thesis.

REFERENCES

[1] M. Średniawa, Telecommunications Reinvented, conference materials XIV Poznań Telecommunications Workshop, Poznań 2010

[2] STL Partners portal http://www.telco2.net/

[3] P. Korbel, J. Legierski, Telco 2.0 – examples of practical use of telecommunication service platforms' interfaces KSTiT 2011 Conference materials, Telecommunications Review 8-9/2011

[4] President of the Office of Electronic Communications, State of the telecommunication market in Poland in 2010 Report, Warsaw 2011

[5] Open Middleware 2.0 Community by Orange Labs www.openmiddleware.pl

[6] Bogusz, D.; Legierski, J.; Podziewski, A.; Litwiniuk, K., "Telco 2.0 for UC — An example of integration telecommunications service provider's SDP with enterprise UC system," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.603,606, 9-12 Sept. 2012

[7] Litwiniuk, K.; Czarnecki, T.; Grabowski, S.; Legierski, J., "BusStop — Telco 2.0 application supporting public transport in agglomerations," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.649,653, 9-12 Sept. 2012

[8] Podziewski, A.; Litwiniuk, K.; Legierski, J., "Emergency button — A Telco 2.0 application in the e-health environment," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.663,677, 9-12 Sept. 2012

[9] Kalitska, S.; Kukielka, P.; Jonczyk, M.; Legierski, J.; Szczekocka, E., "Forecasting of threatening situations in Smart Space," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.641,647, 9-12 Sept. 2012

[10] Legierski J. Tomaszewski T. Udostępnianie interfejsów programistycznych do usług telekomunikacyjnych w Internecie, Software Developer's Journal nr 10 wrzesień 2011

[11] Singh, S. Emergence of payment systems in the age of electronic commerce: The state of art, Internet, 2009, AH-ICI 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 865–869

Parking Reservation – application dedicated for car users based on telecommunications APIs

Piotr Trusiewicz
Warsaw University of Technology Faculty of
Mathematics and Information Science
ul. Koszykowa 75
00-665 Warsaw, Poland
Email: trusiewicz.piotr@gmail.com

Jarosław Legierski
Orange Labs
ul. Obrzeżna 7,
02-691 Warsaw, Poland
Email: jaroslaw.legierski@orange.com

***Abstract*—The main objective of this paper is proposition simple, easy to implementation and low cost solution dedicated for parking lots reservation. The presented application uses Unstructured Supplementary Service Data (USSD) as an communication channel between driver and parking system. USSD communication proposed in this paper is more efficient and comfortable for the end user in comparison with SMS used in many existing parking solutions. System is integrated with communication service provider infrastructure using Service Delivery Platform exposed APIs for telecommunication network in Internet. Application can be launched on every phone and does not require Internet access on mobile phone side.**

I. Introduction

THE parking place is costly and sometimes very limited resource in the cities. Every day thousands of car drivers spend a lot of the time to find an empty parking space. The result of this situation is the air pollution in urban areas, increasing traffic congestion and frustration of drivers. In large cities the traffic generated by drivers searching free parking places can achieve about 40 % of total traffic [1]. In order to solve this problem, the implementation of dedicated reservation based parking system in cities for managing parking places is mandatory.

A. Existing solutions

In last years many researchers proposed architecture of advanced parking systems supporting citizen in free parking spaces allocation. In this chapter are described some of them.

The first solution is Smart Parking Reservation System proposed by researches from University Teknologi PETRONAS in Malaysia [2]. Using this system car driver can reserve parking lot using Short Message Services (SMS). SMS messages are read and interpreted by GSM modem installed in micro-RTU (Remote Terminal Unit). Micro-RTU also sends to the car driver information about reserved lot number and password which is dedicated for opening barrier gate.

Another solution Automated Parking Slot Allocation System [3] proposes using RFID technology for allocation free parking slot. In this system the driver is informed about free parking place using SMS communication channel. The driver can use this channel to reserve his parking slot as well.

Another solution SmartParking described in [4] is dedicated for NOTICE. It is a secure and privacy-aware architecture for the notification of traffic incidents. In this system car driver uses dedicated mobile application for PDAs, smartphones, vehicle display and laptops which can read the information from SmartParking based on Internet access (data) to the system.

Smart Parking System developed by University of Nebraska-Lincoln [5] uses Internet (by Wi-Fi or GSM) for communication with end user using Web Application.

Another solution, Wireless Mobile-based Shopping Mall Car Parking System (WMCPS) [6] uses SMS for interaction with the driver. End user of the system can request for reservation car parking spaces using their mobile phone. WMCPS have got implemented GSM modem for integration with mobile network.

B. Description of the problem

Presented above parking systems uses dedicated hardware (modems) for communication with end users. This solution generates additional costs (hardware) and can be not effective in large scale usage (due to performance issues of GSM/UMTS modems). Proposed and implemented SMS communication results need to send an SMS with the specific content and potential mistakes in SMS content results errors in application usage. Another solution for car drivers dedicated or web application for mobile devices uses data connection and generates costs for user. This paper proposes usage of another communication channel available in mobile network - Unstructured Supplementary Service Data and provides an alternative to existing solutions.

978-83-60810-53-8/$25.00

C. Telecommunication APIs and Service Delivery Platforms

In last few years we can observe the process of opening communication service provider networks for external developers. For many years the network operators were closed to external companies and programmers and only operator was able to develop innovative telecommunication services.

Telecom operators seeing changes in Internet and competing with Internet companies (Over The Top – OTT players) such as Google, Facebook or Skype, were implementing business models based on API (Application Programming Interfaces) exposure.

Using API telecommunication service providers can expose large sets of functionalities in Internet for external developers. Telecommunication functions such as call management, SMS and MMS communication, USSD, payment or locating terminals can be offered third parties as Web Services. Based of them is possible to create new innovative applications connected Internet assets, telecommunication area and IT functionalities such as [7], [8], [9]. From technical point of view – presented above enablers are exposed in Internet using dedicated system- Service Delivery Platform (SDP). SDP is additional layer between Internet and communication network. South interfaces of SDP are connected with network elements such as SMS Center, MMS Center or GMLC (Gateway Mobile Location Centre) using binary telecommunication protocols based on Signaling No 7 stack (SS7). North interfaces are connected to the Internet and expose APIs using SOA model or RESTful architectural style. The architecture of the API exposition using SDP is presented in Fig. 1.

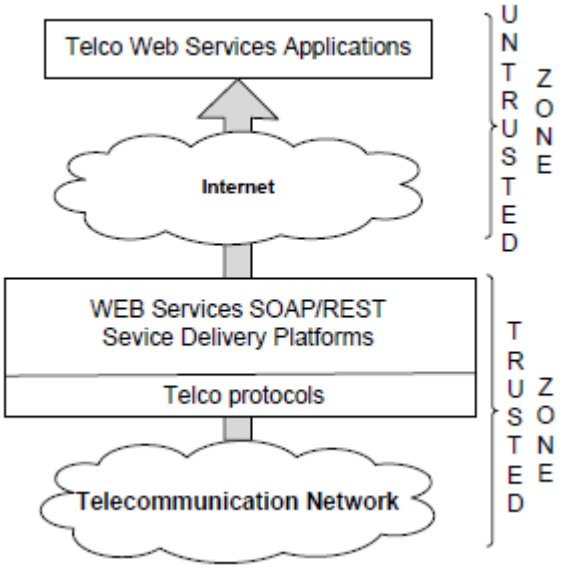


Fig 1. API exposition architecture [7]

Web services were implemented in SDP were standardized using two models: Parlay X specification [10] defined by ETSI and the Parlay Group - based on Service Oriented Architecture [11] and SOAP protocol [12]. Second standard OneAPI [13] was defined by GSMA and the newest version of this specification is based on Representational State Transfer architectural style (RESTful) [14] de facto standard in Web 2.0 and Social Media world.

II. The Parking Reservation System

The Parking Reservation System is an application prototype based on API for operator's network. Presented in this paper system is dedicated for supporting and managing parking places reservation process. The system allows to make reservation or if reservation has been made to cancel it. Application recognizes two types of end users. One is the end mobile user - making the request, the other one is parking security - handling the request using Web based user interface. The mobile users can store in their mobile phone address book two records. The first record is responsible for parking reservation and contains ***665*0015*0#** USSD request**,** the second address book position cancel reservation and is coded as ***665*0015*1#**. To make or cancel reservation the end user must call stored in book number (USSD code) and therefore send specific USSD request.

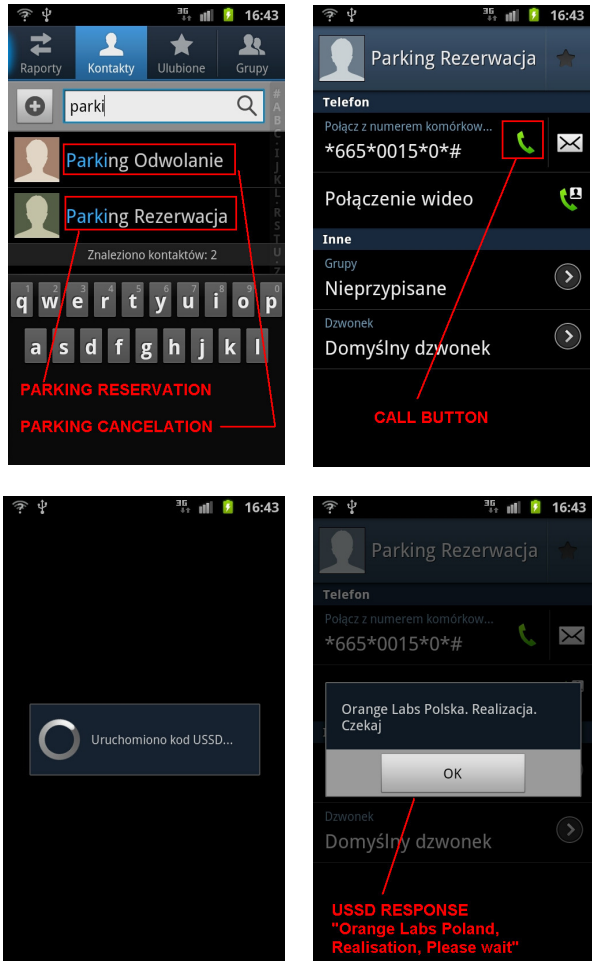


Fig 2. Parking system mobile end user interface

In current version system recognizes only these two described above USSD messages. The request from mobile user in web application is presented to the person responsible for parking place reservation. To handle the request the parking security user selects in web application free parking lot and sends mobile user response using one of the method: USSD or SMS. In a similar way (using another USSD code) is realized parking place reservation cancelation. In case of no free parking places the parking security user can send to the mobile user dedicated message predefined in application (Fig. 3).

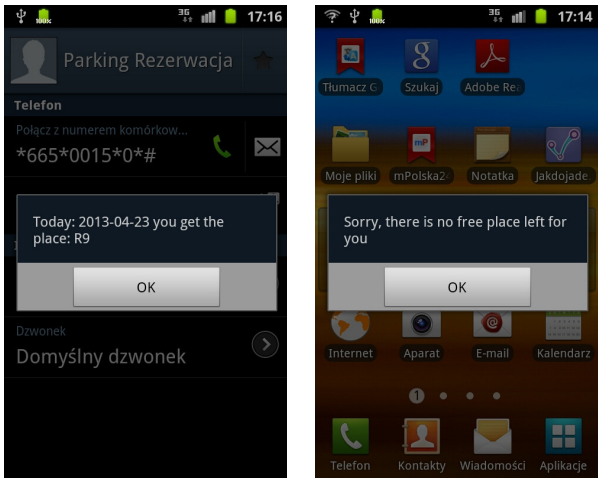


Fig 3. Parking system - example messages

III. System Architecture

A. Functionality of the solution

Using their mobile phone mobile end users sends USSD request to the system. The parking system notifies parking security end user in web application using AJAX request. In the next step the parking security end user using web application (Fig 4.) sends response to the system and application forward it to the mobile phone end user through appropriate SDP's Web Service. The functionality of the system is shown in Fig.4 and Fig 5.

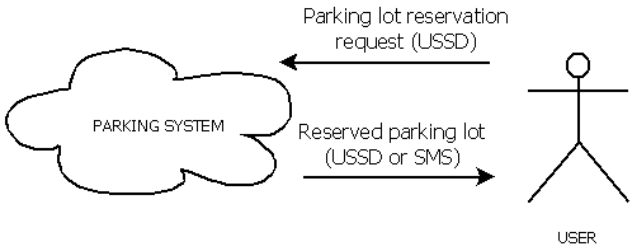


Fig 4. Functionality of the Parking System

B. Used API interfaces

The Parking System functionality is based on API provided by the cellular phone networks. The system is invoked with the USSD request. The response can be send using two methods: USSD or SMS message. Both functionalities are realized by Orange Service Delivery Platform \using Web Services implemented in RESTful architectural style as Receive USSD, Send SMS and Send USSD APIs.

C. Structural architecture of Parking System

The Parking System service consists of three main functional parts based on the performed functions shown in Fig. 5.

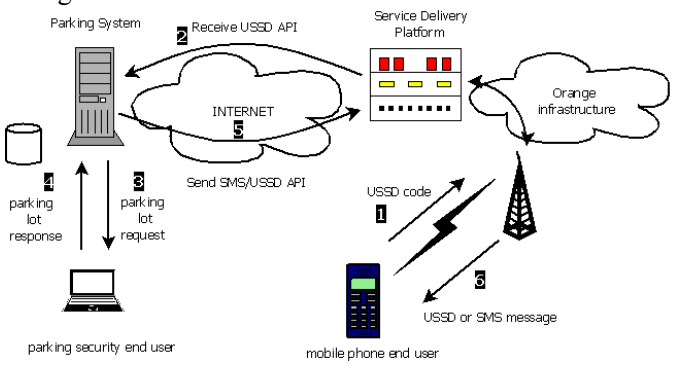


Fig 5. Structure of the developed service

- End Users – there are two types of end users: one communicates with the parking system using mobile phone through UTRAN/GERAN (mobile end user) and the second (parking security end user), which using Web Application maintains parking resources.

- Parking System – the application logic implemented as a Web Application, running on a server. System data is stored in XML files. Data consists of USSD requests, system end users and parking resources data. The system is using telecommunication APIs exposed by Orange for communication between end users.

- GSM/UMTS Operator – enables communication with the Parking System through exposed APIs interfaces.

D. Class Diagrams

Presented solution was developed in Model-View Controller (MVC) style. It consists of Controller, Model and Helper classes. All classes are presented in Fig. 6 and Fig 7. Views are simple dynamically generated HTML. The project consists of following classes:

- Controller classes – responsible for handling HTTP requests. GET and POST requests are mapped to appropriate method called action from the class. Diagram is presented in Fig. 6.
- Model classes – represents the data objects send within system. Each class consists of properties and the setter/getter methods for each.
- Helper classes – are responsible for parsing XML data, handling AJAX, sending response to mobile phone end user and are presented in Fig. 7.

Fig. 6. Controller class diagram

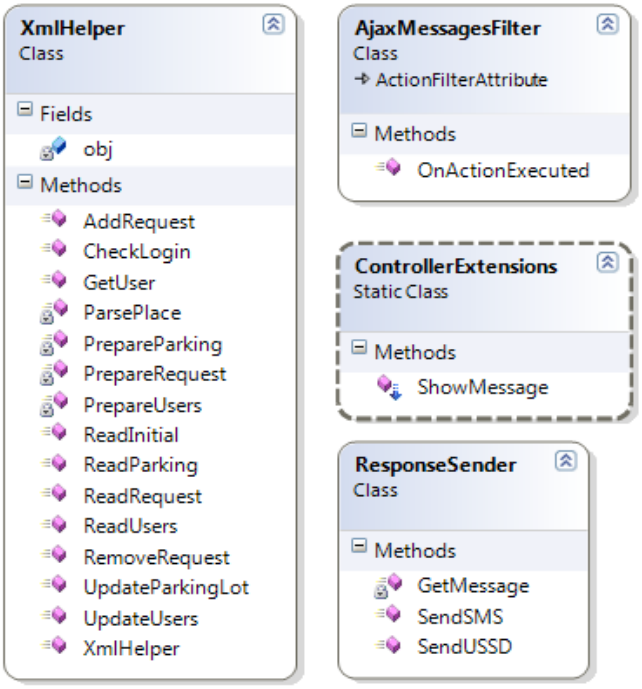


Fig. 7 Helper class diagram

IV. User Interface

The Parking system was developed using C# programming language and ASP.NET MVC 3 framework. The system has a simple, intuitive web interface. The user (parking security end user) logged in to the system, on the main page can see the map of the parking, three tabs on the right (home, make reservation, cancel reservation) and request notification messages on the bottom. On the map reserved parking places are marked with a car pictures. Home page of application is presented in Fig. 8.

Fig. 8. Parking system - home page

Application supports two types of notification messages. One for making reservation and second for canceling. Both types of notification are presented in Fig. 9.

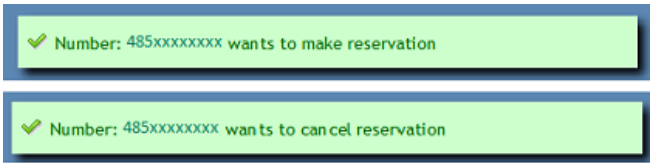


Fig. 9 Notification messages

To make or cancel reservation the parking security end user has to select the appropriate tab and fill up the form and select communication type (USSD or SMS). The forms are shown in Fig. 10 and Fig. 11.

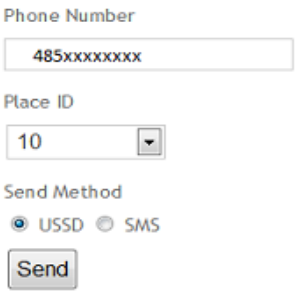


Fig 10. Reservation form

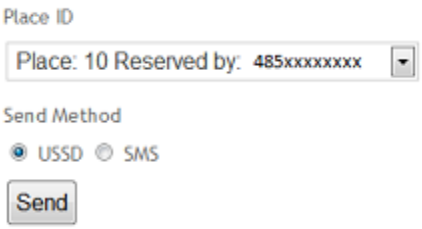


Fig. 11 Cancellation form

After performing the operations the person responsible for parking place management is redirected to the home page with parking map.

AJAX, JSON, JavaScript. technologies were used to create the user interface, Microsoft Internet Information Server was used as an application container.

V. CHALLENGES

The presented in this paper application has some possible future enhancements. It is possible to add new communication channel via SMS for parking place reservation. In the further development of the system it is also possible to create a mobile application dedicated for smartphones. One of the future challenges will be function of navigation to the reserved parking place. This functionality requires a precise localization API (e.g. based on GPS or Wi-Fi) or usage of specific and expansive hardware (e.g. sensors) [3], [4], [5] and implementation of algorithms for parking management strategies [5], [6]. Another enhancement concerns on the functionality of the Parking System. The current system version supports only two mobile end user actions: make reservation and cancel them. There are few more options available such as: check reservation status, change status, extend reservation time, make reservation for some time in future, etc. Potentially interesting idea is automation of system functionality: allowing end user to triggers the application with a USSD or SMS channel without interaction with parking guardian. Based on this the system could automatically reserve the place, recognize the car or end user mobile phone and navigate the user to the reserved place. Parking navigation system would be very helpful, when the mobile end user wants to find the car on the parking.

VI. CONCLUSION

Presented in this paper the prototype of the Parking Reservation System is low cost and effective solution. Because system is based on web application can be hosted in cloud computing environment and offered potential (companies, security agencies) as a service. Because no specific hardware requirements system can be used by everyone and need only mobile phone (smartphone are not necessary) for mobile end user and computer with Internet access for security end user. The implementation telecommunication APIs: Receive USSD, Send SMS and Send USSD in Web Services allowed creating application using standard programming tools in very short time. Using this system car driver can reserve a parking slot on the fly in very easy way by pressing call button on their mobile phone.

Prototype of Parking reservation System was developed under the program Open Middleware 2.0 Community by Orange Labs [15].

REFERENCES

[1] P. White, “No Vacancy: Park Slopes Parking Problem And How to Fix It,” http://www.transalt.org/newsroom/releases/126 [23.04.2013]
[2] Hanif, N.H.H.M. Badiozaman, M.H.; Daud, H., Smart parking reservation system using short message services (SMS) International Conference on Intelligent and Advanced Systems (ICIAS), 2010, Kuala Lumpur, Malaysia
[3] K. Ganesan, and K. Vignesh, "Automated parking slot allocation using RFID technology," Signal Processing and Its Applications, ISSPA. 9th International Symposium on, February. 2007, pp. 1-4.
[4] Gongjun Yan; Olariu, S.; Weigle, M.C.; Abuelela, M., "SmartParking: A Secure and Intelligent Parking System Using NOTICE," Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on , vol., no., pp.569,574, 12-15 Oct. 2008
[5] Hongwei Wang; Wenbo He, "A Reservation-based Smart Parking System," Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on , vol., no., pp.690,695, 10-15 April 2011
[6] Soh Chun Khang; Teoh Jie Hong; Tan Saw Chin; Shengqiong Wang, "Wireless Mobile-Based Shopping Mall Car Parking System (WMCPS)," Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific , vol., no., pp.573,577, 6-10 Dec. 2010
[7] Litwiniuk, K.; Czarnecki, T.; Grabowski, S.; Legierski, J., "BusStop — Telco 2.0 application supporting public transport in agglomerations," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.649,653, 9-12 Sept. 2012
[8] Bogusz, D.; Legierski, J.; Podziewski, A.; Litwiniuk, K., "Telco 2.0 for UC — An example of integration telecommunications service provider's SDP with enterprise UC system," Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on , vol., no., pp.603,606, 9-12 Sept. 2012
[9] A Podziewski, K Litwiniuk, J Legierski, E-health Oriented Application for Mobile Phones Informatica Special Issue: Advances in Network Systems, 12.2012
[10] Open Service Access (OSA); Parlay X web services, 3 GPP, http://www.3gpp.org/ftp/Specs/html-info/29199-01.htm, [23.04.2013]
[11] SOA Practitioners’ Guide Part 2 SOA Reference Architecture, http://www.soablueprint.com/whitepapers/SOAPGPart2.pdf, [02.07.2013]
[12] SOAP specification W3C, http://www.w3.org/TR/soap/ [23.04.2013]
[13] GSMA OneAPI http://www.gsma.com/oneapi/ [23.04.2013]
[14] L. Richardson, Sam Ruby, David Heinemeier Hansson, RESTful Web Services, O’Reilly, 2007
[15] Open Middleware 2.0 Community by Orange Labs http://www.open-middleware.pl/ [14.05.2013]

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 871–874

Student Information Delivery Platform Using Telecommunications Open Middleware APIs

Piotr Wawrzyniak, Piotr Korbel, and Anna Borowska-Terka
Institute of Electronics
Lodz University of Technology
ul. Wólczańska 211/215, 90-924 Łódź, Poland
piotr.wawrzyniak@dokt.p.lodz.pl, piotr.korbel@p.lodz.pl, anna.borowska-terka@p.lodz.pl

***Abstract*—The paper describes architecture of a prototype networked student information delivery system. Main system functionalities include interactive access to lecture room timetables and group messaging. The system exploits modern mobile technologies to allow flexible usage scenarios. The use of open APIs of telecommunications service delivery platforms in combination with e-mail messaging provides diverse ways of system information delivery. The perceived application scenario of the system is to provide ubiquitous access to up-to-date lecture room timetables and reliable ways of notifying the affected users about changes.**

***Index Terms*—student information systems, web services, mobile applications**

I. Introduction

NOWADAYS, universities worldwide offer a variety of complex campus services addressed to different groups of users: academics, administration staff, and students. Among the information systems facilitating the management of these services we can find systems supporting various fields related to the educations process, like student information database systems [1], [2], [3], [4], systems facilitating students and staff mobility and general cooperation between education institutions [5], [6], [7], learning management systems [8], and many more.

In this paper we propose a system enabling ubiquitous access to up-to-date lecture room timetables and reliable ways of notifying the affected users about unexpected changes. The proposed system exploits modern mobile technologies (mobile phones, tablets) to allow flexible usage scenarios. The use of open APIs of telecommunications service delivery platforms [9] provides reliable and fast way of delivery of system messages to the users.

The remainder of the paper is organized as follows. Section II presents overall architecture of the prototype system and provides details on the system modules. Section III describes user system interactions as well as user interfaces of the system modules. Section IV summarizes the paper.

II. System architecture

Proposed system has a distributed heterogeneous network architecture. In particular the complex solution might be divided in two independent but yet complementary branches. The first one consists of modern interactive lecture room timetable delivery platform that provides the users with the most actual timetables for the auditoriums. The other part of the solution consists of group messaging platform utilizing telecommunications open APIs [9] to deliver messages to the mobile phones.

Both parts were developed separately with the use of different open source and proprietary technologies. The joint use of interactive timetables and group messaging offers possibility to notify system users on the temporary timetable modifications. Moreover, the ability to provide user with additional information makes it possible to offer group messaging in order to improve the communications between students and academic teachers.

The remainder of the section provides detailed description of the aforementioned system modules.

A. Interactive Lecture Room Timetable Delivery Subsystem

As mentioned before in many universities among the world contemporary printed schedule boards are still in use. This way of information delivery has strong advantages among which low usage cost and effectiveness of delivery are the most important. On the other hand, the possibility of live interaction with printed timetables is impossible. Thus, any unusual situations and events require manual updates of the timetables and in urgent cases engagement of additional communication channels to pass the notification of the changes to affected groups of students and academics.

Electronic boards can present information in a similar way to the printed ones but additionally may offer an interaction channel to provide the possibility of on-line data modifications. That was the main idea that led us to design and implement the timetable delivery subsystem.

Proposed system architecture consists of four main elements as presented in Fig. 1. Central application server is responsible for controlling input data, in particular it provides timetable conflicts resolver. It is also responsible for preparing data for each lecture room display. It is also the only element that interacts with incorporated Relational Database Management System (RDBMS). Database server is used to store all the timetable data.

The management of the system is also possible with independent WebGUI service that allows authorized users to enter timetable data for supported lecture rooms. It includes state-of-the-art calendar interface which permits to define events of

978-83-60810-53-8/$25.00

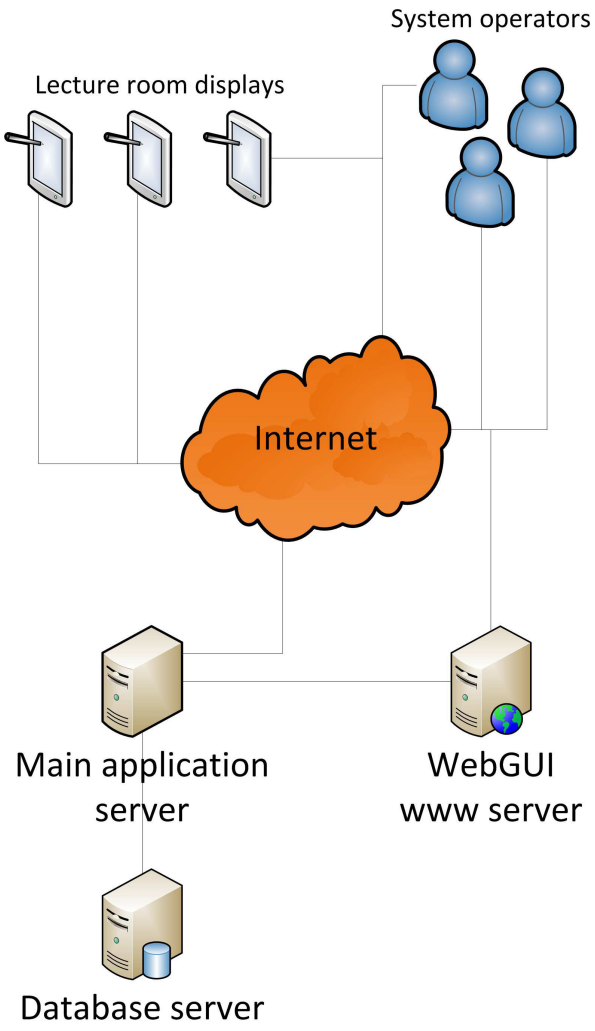


Fig. 1. Timetable delivery system architecture diagram.

different kinds, in particular single-time and periodic events are supported. The last part consists of interactive displays to be mounted at entrances to the lecture rooms.

As the result of conducted market research it was decided to use modern Android OS based tablet devices as the interactive displays. Therefore, dedicated application for Android operating system was developed. The main goal of the application is to display schedule timetable in convenient and accessible way. It is designed to be constantly active and refresh the data for up-to-date timetable delivering. Moreover, after each user interaction application returns to it default state (i.e. displays the timetable for current week).

As extensive user-timetable interaction is expected, it was also decided to develop dedicated pointing device to navigate over schedule application in order to reduce the use of touch screen of the device. The prototype device uses Bluetooth protocol to communicate with tablet device and is equipped with seven buttons. Four of them are used as the four-directional joystick, two are used to change the week view and the last one allows to display additional subject related information in a new window.

B. Student Massive Messaging Subsystem (SMMS)

The ability to send notification to a group of students is the main feature exposed by Student Massive Messaging Subsystem (SMMS). Contemporary messaging services at the university usually use university electronic mail systems as the only way to contact students. However secure and reliable, the effectiveness of information delivery is strongly affected by the necessity of user initiated mailbox checking in order to retrieve new messages. This might be insufficient in the case of unusual changes to timetable caused for example by lecture room equipment failures.

Aforementioned limitations might be omitted by incorporating the use of mobile phones and Public Land Mobile Network (PLMN) messaging services like Short Message Service (SMS) and Unstructured Supplementary Service Data (USSD). In particular these protocols are the most suitable for sending short messages that should be delivered shortly and reliably. Therefore, the proposed service involves SMS and USSD messaging services exposed by Telco 2.0 APIs of Service Delivery Platform WebGateway. The overall SMMS system architecture is presented in Fig. 2.

The main application server exposes a set of SOAP-based stateless web services that allows the following actions:

- It allows to send notifications on exceptional changes in the timetable (i.e. caused by equipment failure or important special event) to previously assigned users (both students and academics),
- It simplifies the communication between academics and groups of students by allowing academic teachers to send short notifications to selected groups of students registered in the system,
- It provides fast and reliable communication channel between university administration and students which improves the quality of administration services.
- It allows end users to sign-in for receiving the messages on selected topic. The user might sign-in in three ways: by sending a USSD code, SMS message or using Web Interface (which communicates with the main server).

This part of system was developed with the use of Windows Communication Foundation (WCF) technology. SOAP protocol was implemented in the northbound interface of the service, although RESTful web services are planned to be developed for final revision of the software.

User interacts with the system using web-based GUI offered by Web Interface server. This web portal was developed in PHP and communicates with SOAP web services from application server. The tool set also allows to manage database of users and news groups. Sensitive data, in particular user personal data and passwords are processed in accordance to widely accepted rules. Fraud detection mechanisms were also considered to prevent unauthorized use of users e-mail addresses and mobile phone numbers. In particular, verification

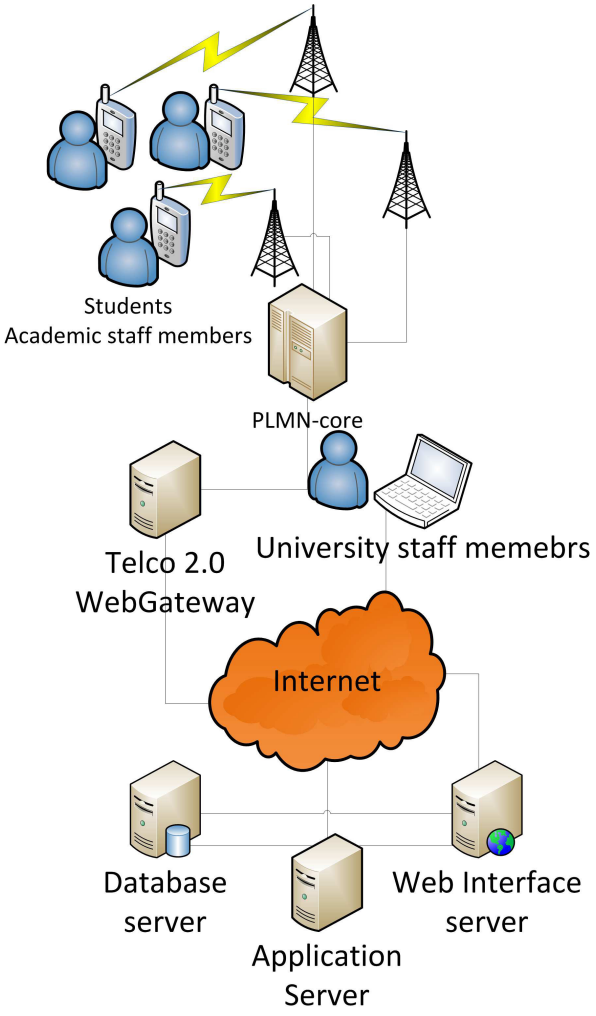


Fig. 2. Architecture diagram of the massive messaging system.

mechanisms are invoked every time the user wants to join the system or messaging group.

All the internal data, as well as user account information are stored in incorporated RDBMS which is accessible for both the application and WebGUI servers.

III. USER SYSTEM INTERACTION

Despite the capabilities of the developed solution user interaction models are the key factors of the deployment success. Thus four complementary user interfaces were proposed:

- Web-based ical User interface for timetable management. This interface allows to easily create and manage timetable for a given lecture room. The access is restricted only to previously defined users. It is also planned to offer students personalized timetables. Sample GUI view is presented in Fig. 3.
- Electronic interactive timetables (EIT) mounted at the entrance to the lecture rooms equipped with wireless 7-buttons pointing device. The primary objective of the EIT is to provide the most up-to-date contemporary timetable

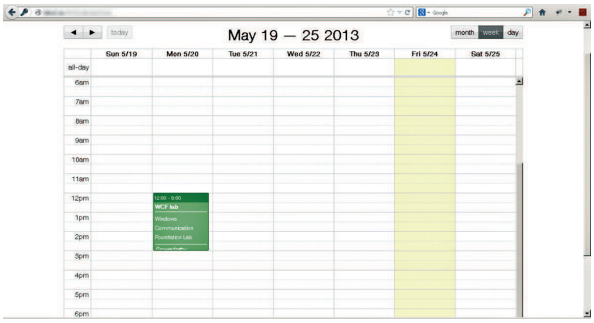


Fig. 3. Example of the timetable manager GUI.

but it is possible to obtain additional information on the selected entry thus EIT strongly extends the timetable capabilities.
- Web based GUI for management of group messaging subsystem. This interface is intended to be used by both academics and students. Staff members might create topic groups (editor role) or join existing ones (notification recipient role). Student may only join the existing groups. As mentioned in section II, to avoid frauds or improper use, e-mail addresses and mobile phone numbers submitted to the system are subjects of verification procedures.
- User might join the messaging group by sending a USSD code. This method is intended for combined use with the capabilities of the EIT in order to minimize the effort needed to join the desired messaging group. It is worth to notice that in this case there no necessity for phone number verification as the action has to be initiated on the target device.

The use of the system includes notifying affected users on changes in the timetables. Nevertheless of the change type, whether it is planned or emergency situation the system allows to improve the communications between building management team and students or teachers. It is especially useful in the case of emergency situations when all the users of the lecture room might be informed on the unexpected changes.

The latter use case scenario involves the academic teachers or administration staff (e.g. dean office). Nowadays they usually communicate with the selected student using emails. Then the leader of the student's group is obligated to pass the information received to the rest of the group. Although emails are reliable and secure mean of communications, in the case of urgency this approach might be ineffective. SMS or USSD messages provide similar level of security and reliability but due to immediate delivery are the most appropriate for such situations. The use of proposed solution creates the possibility to inform all the interested students at once just by creating notification to the target messaging group.

Aforementioned examples only show the base system functionalities and application possibilities. Due to open and distributed architecture the proposed solution might be easily adopted to new application scenarios.

IV. SUMMARY

In the paper we presented a prototype Student Information Delivery Platform. Proposed system architecture consists of two joined independent modules:

- Interactive Lecture Room Timetable Delivery Subsystem which allows to create, manage and display interactive timetables at the entrances to the lecture rooms,
- Student Massive Messaging Subsystem (SMMS) which makes it possible to send notifications to predefined groups of students or academics. The system incorporates SMS, USSD and email messaging in order to suit various delivery time and reliability needs.

Joint use of the proposed subsystems allows instant notification of the users on unexpected changes to the timetables as well as facilitates communications between academics and student groups. Future system development plans include implementation of RESTful northbound interfaces in all the key components as well as implementation of interfaces enabling integration with already deployed campus systems.

REFERENCES

[1] J. Mincer-Daszkiewicz, "Student Management Information System for Polish Universities", *in Proc. of The 8th International Conference of European University Information Systems EUNIS,* 2002.
[2] A. Materka, M. Strzelecki, and P. Dębiec, "Student's Electronic Card: A Secure Internet Database System for University Management Support", Advances in Intelligent and Soft Computing, No. 64, 2009, pp. 59–72.
[3] University Study-Oriented System, http://usos.edu.pl/, Accessed 19 May 2013.
[4] LADOK consortium, http://www.ladok.se/, Accessed 19 May 2013.
[5] J. Mincer-Daszkiewicz, F. Arcella, and S. Ravaioli, "Web-services for Exchange of Data on Cooperation and Mobility between Higher Education Institutions", *in Proc. of The 15th International Conference of European University Information Systems EUNIS,* 2009.
[6] A. Materka, P. Dębiec, P. Korbel, and M. Strzelecki, "A Networked Information Processing System for Student Mobility Support in European Higher Education Area", Electrical Review (PrzeglÄĚd Elektrotechniczny), No. 7a, 2012, pp. 289–293.
[7] RS3G (Rome Student Systems and Standards Group), http://www.rs3g.org/, Accessed 19 May 2013.
[8] Moodle community, http://moodle.org/, Accessed 19 May 2013.
[9] Open Middleware 2.0 Community, http://www.openmiddleware.pl/, Accessed 19 May 2013.

International Conference on Wireless Sensor Networks

A FEW years ago, the applications of WSN were rather interesting example than a powerful technology. Nowadays, this technology attracts still more and more scientific audience. Theoretical works, where the WSN principles were investigated, grew into the interesting practical applications integrated by this time in a real life. Various application fields, from military to healthcare, are already covered by WSN. Hand in hand with WSN application coverage expansion, still new and new tasks bringing along the new interesting problems occur. Therefore such application practices stimulate the progress of WSN theory as well as unlock new application possibilities.

Wireless sensor networks, as the spatially distributed networks consisted of relatively simple components interconnected mutually, provide quite wide application potential to be utilized in military, industry, transport, agriculture, healthcare and many other branches. However, in the near future, even higher growth of WSN application coverage could be expected. This expansion is nevertheless conditioned by solving questions related to the communication protocols standardization, to the lack of effective energy sources enabling network nodes working time prolongation, as well as to the progress in the field of ultra-low-power microelectronic components industry. An integration of WSN within the public data networks as well as within the domains where confidential and private data are proceed (e.g. E-Health) brings about problems related to the ethical and legal questions too.

The WSN is one of actual affairs getting to the fore in the European Research Area since the issue of Future network technologies research and innovation is planned to be included in the new HORIZON2020 FP7 program.

It is therefore necessary to create an experience-sharing platform for scientific researchers and experts from research institutes and WSN occupied companies that employ benefits of this modern technology and to the exchange of skills, experiences and new ideas from the WSN problematic within the all participants of iNetSApp.

Topics

Original contributions, not currently under review to another journal or conference, are solicited in relevant areas including, but not limited to, the following:

Development of sensor nodes and networks

- Sensor Circuits and Sensor devices – HW
- Applications and Programming of Sensor Network – SW
- Architectures, Protocols and Algorithms of Sensor Network
- Modeling and Simulation of WSN behavior
- Operating systems

Problems dealt in the process of WSN development

- Distributed data processing
- Communication/Standardization of communication protocols
- Time synchronization of sensor network components
- Distribution and auto-localization of sensor network components
- WSN life-time/energy requirements/energy harvesting
- Reliability, Services, QoS and Fault Tolerance in Sensor Networks
- Security and Monitoring of Sensor Networks
- Legal and ethical aspects related to the integration of sensor networks

Applications of WSN

- Military
- Health-care
- Environment monitoring
- Transportation & Infrastructure
- Precision agriculture
- Industry application
- Security systems and Surveillance
- Home automation
- Entertainment – integration of WSN into the social networks
- Other interesting applications

Event Chairs

Fouchal, Hacene, University of Reims Champagne-Ardenne, France

Hodon, Michal, University of Žilina, Slovakia

Kapitulík, Ján, University of Žilina, Slovakia

Miček, Juraj, University of Žilina, Slovakia

Segal, Michael, Ben-Gurion University of the Negev, Israel

Ševcik, Peter, University of Žilina, Slovakia

Stojmenovic, Ivan, University of Ottawa, Canada

Xiao, Yang, The University of Alabama, United States

Program Committee

Gu, Yu, National Institute of Informatics, Japan

Scholz, Bernhard, The University of Sydney, Australia

Shu, Lei, Osaka University, Japan

Staub, Thomas, Data Fusion Research Center (DFRC) AG, Switzerland

Wang, Zhonglei, Karlsruhe Intitute of Technology, Germany

Al-Anbuky, Adnan, Auckland University of Technology, New Zealand

Baranov, Alexander, Russian State University of Aviation Technology, Russia

Chaczko, Zenon, University of Technology Sydney, Australia

D'Innocenzo, Alessandro, University of L'Aquila, Italy

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 877–880

Cloud Computing System Based on Wireless Sensor Network

Wen-Yaw Chung
Institute of Electronic Engineering,
Chun-Yuan Christian University,
Chun-Li, Taiwan, R.O.C
Email: eldanny@cycu.edu.tw

Pei-Shan Yu
Institute of Electronic Engineering,
Chun-Yuan Christian University,
Chun-Li, Taiwan, R.O.C
Email: collin1027@hotmail.com

Chao-Jen Huang
Industrial Technology Research
Institute, Hsin-Chu, Taiwan,
R.O.C. Email: ephoton@itri.org.tw

***Abstract*— In this paper, the system presents an integrated wireless sensor network (WSN) to monitor the information from agriculture systems namely temperature, humidity, pondus hydrogenii (pH) value…etc. The purpose is to provide a faster and more convenient platform for the client to obtain information from an array of sensor nodes that has been set-up in an agricultural system. A WSN will collect the values of various parameters from the front-end sensors at the host end. At the client sides, one can use the internet to request for Web Services that will store this big data into distributed SQL databases which are already in our proposed cloud system. In addition, this work presents the concept of cloud computing and services. The benefits of this system include basic computing hardware and reasonable storage capacities making it suitable for any smart device which can monitor real-time farmland information anywhere. The customers can fully access our cloud service using devices that have internet capabilities.**

I. Introduction

WIRELESS sensor network (WSN) consists of a large number of sensor nodes that are interconnected to form a wide communication network. Usually, it can achieve small size, low cost, low power consumption, fewer network components and other features easily. In recent years, it has been readily implemented in agriculture, industry, environmental protection and other fields.

With the development of hardware limitations, and in pursuit of a better performance and enhancing greater computing capability, people turn to find other techniques to achieve these goals. Therefore, the concept of "Cloud" was born. In fact, as early as the Internet appeared, the "Cloud" has already existed silently providing for us some services.

In recent years, the "Cloud" concept has become more and more popular, especially on the business sector. There are also many types of cloud computing platforms such as Google, Amazon, IBM, and Microsoft…etc. However, the true essence of using "clouds" was not completely understood. "Cloud" was then not a specific technology but rather a concept.

II. Cloud Computing

"Cloud" refers to a network. In the beginning, engineers drew a network diagram in the form of a cloud to represent Wide Area Network. This had an undefined volume of interconnected computers and network routers [1]. So to the client, this "cloud" was just a means of interacting with other sides.

A. Distributed Computing

According to literal interpretation [2], the idea of distributed computing is to divide the whole work load into smaller units. Each work fragment will be given to a corresponding slave computer that will do the computing and then will send back the results to the master computer. This is shown in Fig. 2.

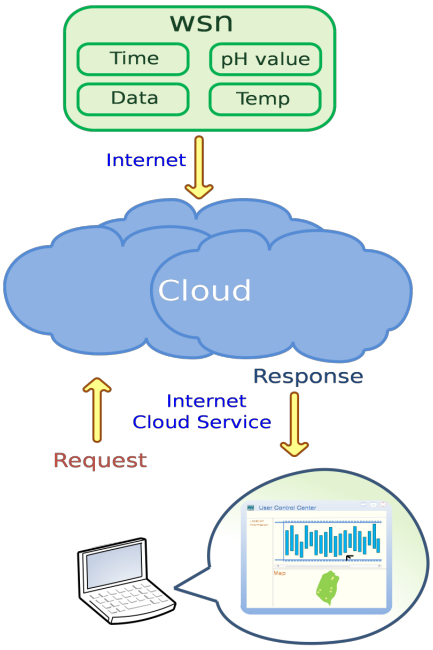


Fig. 1. Cloud concept

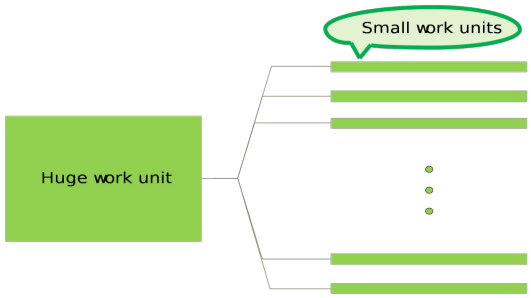


Fig. 2. Distributed computing

Through this technique, cloud computing can achieve a similar processing power to that of a "Super computer" with relatively lower cost and lesser fabrication complexity.

B. Virtualization Technology

Virtualization is a kind of software technology which can be applied to many fields. In fact, the concept of "Cloud" appeared very early [3], but due to lack of network bandwidth, insufficient storage space, and premature virtualization technology it was unstable and lacked flexibility.

However, with the development of virtualization technology, the development of clouds has been accelerated. In the past, our computing power was based on personal computers locally, but through virtualization technology we can now use cloud computing power which is centralized.

The proposed cloud system consists of a master server computer and four virtual slave server computers. The data are distributed onto the respective storage spaces of the four slave server computers which shall be used in the distributed computing process.

III. Database Design

A. Database Concept

Database refers to a particular subject or a theme arranged in a structured set of multiple and complex data in the computer. As a storage structure, the database can be broadly divided into: Hierarchical, Network, Relational, and Object-oriented. We used relational database in our project. Relational database is a database that has a collection of tables of data items which are formally described and organized based on the relational model. In this model, each table must identify a column or group of columns referred to as the "primary key" to uniquely identify each row. The rows of one table can relate to rows of another by establishing a "foreign key" which is a column or group of columns in one table that points to the primary key of another table.

B. Stored Procedure

The Stored Procedure (SP) could be divided into four instructions : Insert, Update, Delete, and Select in the database. These basic instructions can also control the data in the database and even combine more complex instructions so that the user can handle data more efficiently. In this project, we used T-SQL to write the database instructions. Every time one accesses a regular database, it has to check the syntax and this consumes a lot of time. So we implemented another technology that is "Stored procedure" which is written using T-SQL. Stored procedure will process the composition, verify the syntax by T-SQL, and then store. After that, one just needs to use it directly.

IV. Web Service Design

A. Web Service Concept

With XML (Extensible Markup Language), SOAP (Simple Object Access Protocol), WSDL (Web Service Description Language), UDDI (Universal Description, Discovery and Integration) appearing one by one, Web Service-oriented software, a new generation of distributed computing technology, and Web Services were born. Web Service through the open-type standard (e.g. XML, SOAP... etc.) of Web communication protocol alongside with the data provides services to other applications. Web Service consists of reusable components that can be published, discovered and invoked across the Web [4].

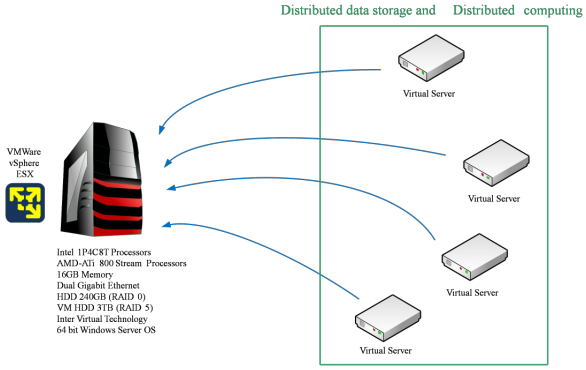


Fig. 3. Cloud system structure

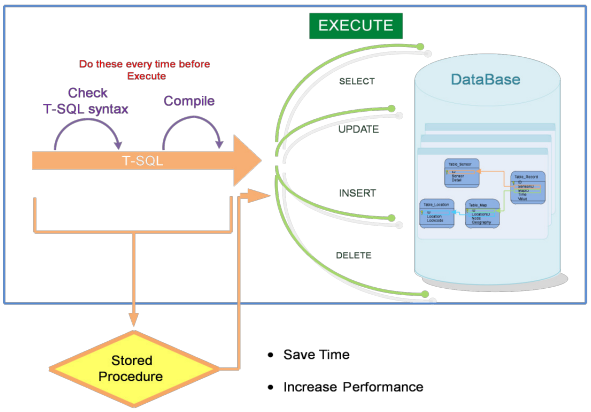


Fig. 4. Stored procedure's concept

Its basic function is to respond to the call from the client to the server. Its purpose is to let the application from different platforms intercommunicate. When using the program, one inputs a value onto it, and then the program will request the Web Service to compute. The project used C# to design the user interface; while the platform was built

using .NET that can do the integration and implement the application.

B. Related Technologies

XML is a language form and a kind of syntax which can easily be read by the user and that the computer program can easily identify. It emanated from the SGML (Standard Generalized Markup Language). XML was developed by the XML Working Group, an organization formerly known as SGML Editorial Review Board; and so, XML is reviewed by a group of SGML experts. XML does not replace HTML. HTML focuses on how the file appears in the browser; while XML focuses on how to represent them in a structured way.

SOAP, Simple Object Access Protocol, SOAP is a kind of protocol that defines the way XML data are delivered and the various transmission protocols, like HTTP, FTP, SMTP, and TCP. SOAP uses XML as the data transmission format, and combines with the aforementioned transmission protocols to send the message.

C. LINQ-to-SQL

LINQ is a kind of method to link different program languages. Without LINQ, it is hard to establish a communication link between the SQL database and the programming language. LINQ can do object-oriented programming; it can also make the T-SQL language easily readable in C#, where the basic language is different; and it can also automatically generate the corresponding data type.

V. User Control Center

The relationship between user control center and Web Service are divided into two directions: “data” and “Panorama Map”.

A. Data Curve

Due to the large amount of data coming from the front-end sensors of temperature, humidity, and pH value… etc., a virtual machine is therefore used as a storage medium whose contents shall later be extracted via the Web Service. The process is as follows: one will first get the display window size to the Web Service, then he will extract the quantity of data from the database. The cloud will then compute the result which will be fitted onto the screen. [5]

B. Panorama Map

Because the farm we can monitor can be of large scale covering entire Taiwan or even the whole world, we therefore made reference to the operation of Google map. At first, the program will send the start instruction to let the cloud side know the user side’s display window size, and then the cloud will compute the image result and relay it back to the user control center.

Seven parameters have been designed to operate the map, namely: user control center’s display window width and height, X and Y coordinates of present location map, X and Y coordinates of displacement, and the presented map hierarchy.

When the user zooms in or out of the map, the user control center will send the new map’s hierarchical parameters to the cloud to regain the map. As the map’s display is maximized, it will show the images of the farmland as well as the corresponding sensor information. When the sensor is clicked, the user will get its information and data.

C. Interface

The design of the data curve is divided into three parts, namely: initial, static and dynamic. Regardless of whether the user starts the user control center or changes the window size, the coding of the initial status will still redraw the screen. It will calculate the window size, split the X axis, and then use the amount of data it gets from the Web Service to calculate the initial screen size.

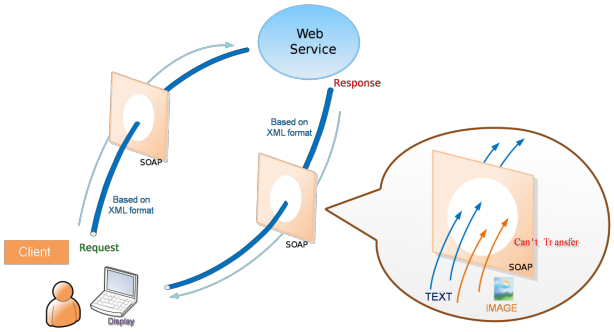


Fig. 5. Web Service concept

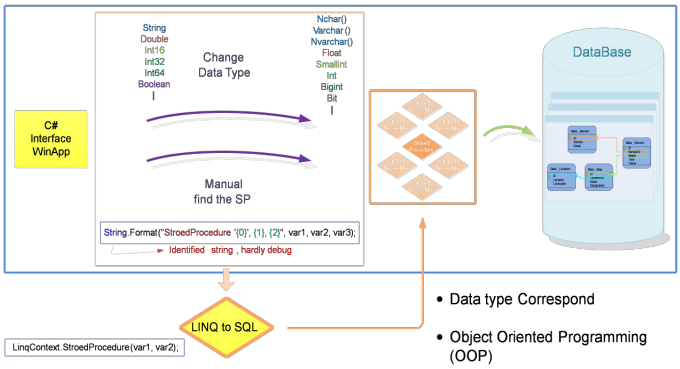


Fig. 6. LINQ-to-SQL concept

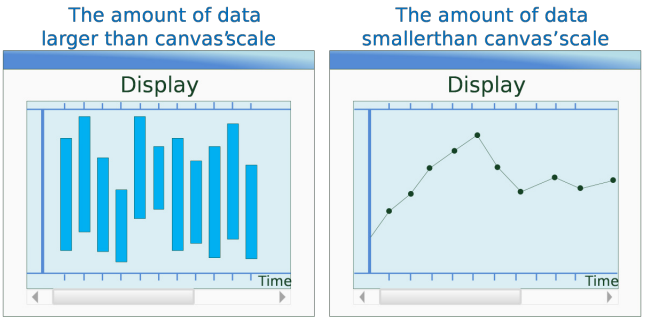


Fig. 7. Relation between data curve and data

The design of the relational database involves two concepts: one is that the data in the database should not be multiple repetitions, and the other is that two separated

tables should be established with a correct relation to ensure the consistency. A good database design depends on its characteristics. Because of the related characteristics, data duplication between tables can be minimized. This design results into saving the storage space and establishing a correct relation. Hence, one doesn't need to modify each data over and over again.

VI. Conclusion

The system proposed an internet database design by using the SQL database, LINQ-to-SQL technique, Web Service, virtualization technology and C# interface. By using this paper's method, the client's end can monitor the environmental condition of the agricultural place at any place. The major elements of this work are wireless network applications, C # interface, cloud computing system, database design, Web Service, and user control center.

Data packets were sent via a USB connection to the host-end which transmits the values of various environmental parameters coming from the front-end sensors. The number of WSN nodes can be more than 600,000 and each node can return the value for every one second. Therefore, the database will become enormous. The issue of quick access of information at the client end has been addressed by our system. C# interface will be designed to display the data curve which will aid in the decision making of the client with regard to getting better yield from his farm. Besides this, the sensor data will be uploaded to the cloud database allowing the client to use our Cloud Service as long as the user's display facility has internet connectivity.

Acknowledgment

The proponents of this work would like to acknowledge the National Science Council, Taiwan, ROC for funding this project (NSC 102-2221-E-033-066). The work would also like to extend its gratitude to Chip Implementation Center (CIC) for the technical support given.

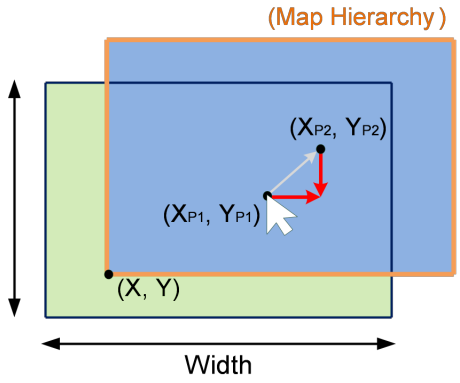


Fig. 8. Seven map parameters

Fig. 9. C# User control center interface

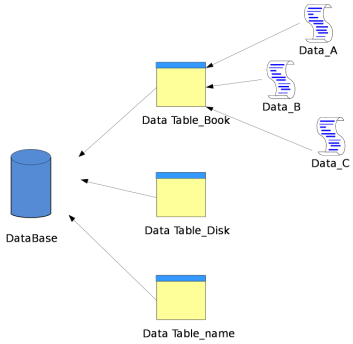


Fig. 10. The connection between the system data table and data

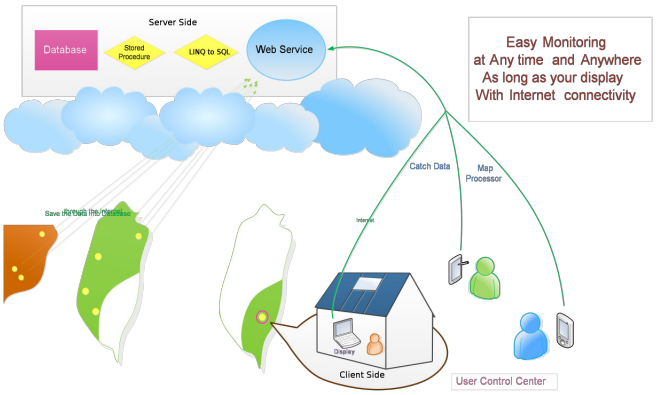


Fig. 11. The benefit of the cloud computing database system

References

[1] W.-L. Liu, (Oct. 2008). Cloud computing has been talked about rotten? Gartner recommends re-learn from the dichotomization. [Online]. Available: http://mr6.cc/?p=2281

[2] J.M. Reddy, J.M. Monika, "Integrate Military with Distributed Cloud Computing and Secure Virtualization," *IEEE Computer Society*, pp. 1200-1206, Nov. 2012.

[3] F.B Shaikh, S Haider, "Security threats in cloud computing," *IEEE Internet Technology and Secured Transactions (ICITST)*, pp. 214-219, Dec. 2011.

[4] M. Beraka, H. Mathkour, S. Gannouni, H. Hashimi, "Applications of Different Web Service Composition Standards," *IEEE International Conference on Cloud and Service Computing (CSC)*, pp.56-63, Nov. 2012

[5] W.-Y. Deng, "Design A Database Program," in *Visual 2010 C# Program Designed Strategy*, vol. 14, *Wen-Yuan Studio*, Taiwan, TW: Gotop, 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 881–888

Approaches of Wireless Sensor Network Dependability Assessment

Antonio Coronato and Alessandro Testa

Institute for High-Performance Computing and Networking - CNR
Via P. Castellino 111, 80125 Napoli, Italy
{antonio.coronato, alessandro.testa}@na.icar.cnr.it

***Abstract*—The extensive use of the Wireless Sensor Networks (WSNs) in main critical scenarios stresses the need to verify their dependability properties at design time to prevent wrong design choices and at runtime in order to make a WSN more robust against failures that may occur during its operation. In literature, several approaches have been proposed in order to evaluate the dependability of a WSN during its inception and its operating.**

In this paper we present a survey on these adopted techniques reporting aspects and characteristics of some research studies. Moreover, by means of a comparison grid, we analyze the current state-of-the-art of the approaches of WSN dependability assessment in order to identify the most performant and to discuss the ongoing challenges.

***Index Terms*—Wireless Sensor Networks, Dependability, Reliability, Fault-Tolerance**

I. Introduction

NOWADAYS Wireless Sensor Networks (WSNs) are usually involved for critical systems monitoring [1][2] and in smart environments [3], thus the recent scientific production on WSNs dependability assessment is grown .

Unexpected events, such as node crash and packet loss, may affect the dependability of the WSN and hence it is necessary evaluate its robustness from the early stages of the development process (*design phase*) onwards to minimize the chances of unexpected problems during use. It is also crucial to monitor a WSN at runtime (*operating phase*) and to detect undesired effects that cannot be analyzed before the WSN deployment.

The approaches adopted in literature to assess WSN dependability, at *design time* or *runtime*, can be categorized in four classes: *experimental*, *simulative*, *analytical* and *formal*.

The first allows to analyze dependability at runtime by means of experiments. Experimental methods are used to evaluate a real system and they require the deployment of a real WSN. They are useful in operating phase since by means of them we can perform experiments directly on the real system from which we collect data. Among the experimental approaches we consider the *Fault Injection* techniques[4] which allow to evaluate the dependability of a real WSN by injecting faults.

The simulative approaches make use of well-known simulators that can be adopted at design time; this kind of approach consists in modeling the WSN and making an estimate of the dependability. In a similar way analytical approaches are conceived; the difference is in considering mathematical models to check the WSN behavior during the design. Finally, formal approaches use specifications of correctness and they can be adopted to assess WSN dependability both at design time and at runtime. Formal approaches offer a new opportunity for studying the WSN dependability.

The aim of this paper is to revise experimental, simulative, analytical and formal approaches and tools currently used in the field of WSN dependability assessment, including related studies. We want to provide a survey on the main approaches adopted for the WSN dependability assessment evaluating the best choice to perform dependability assessment of WSNs. A comparison of related work is presented to summarize the state-of-art and reason about what is still missing and the ongoing challenges.

The rest of the paper is organized as follows. In Section II experimental approaches are presented; the Section III includes the simulative approaches. An analysis of analytical approaches is documented in Section IV and the formal approaches are discussed in Section V. Finally in Section VI we report a comparison of the discussed papers and the Section VII ends the paper with conclusions.

II. Experimental Approaches

Experimental approaches are used to measure the WSN dependability directly from a real system, during its operation. In the prototyping phase, it is possible to perform an accelerated testing, for example by forcing a fault (by means of *Fault Injection* (FI) [4], [5]); it is also possible to collect occurring failures directly from system (by means of *Field Failure Data Analysis (FFDA)* techniques [6]).

FI is defined as the dependability validation technique based on the observation of the system behavior under the presence of faults which are deliberately introduced into the system [7].

Typically FI is used to i) assess the dependability level of a target system, such as an operational systems, a system

This work has been partially supported by the "Se@me: Sustainable e-maritime @ssistance for Maritime Employees, Passengers and Yachtsmen" project, funded under the "POR Campania 2007/2013 " initiative of the Regione Campania - Italy.

978-83-60810-53-8/$25.00

TABLE I
FAULT INJECTION TOOLS

Tool	Technique	Fault Model
XCEPTION [8]	SWIFI with exception trigger	Transient faults
FERRARI [9]	SWIFI with interrupt, fork, trap	Transient and permanent faults
FIAT [10]	SWIFI with exception trigger	Bit-flip faults in the memory
NFTAPE [11]	SWIFI with exception trigger	Several types of faults (*arbitrary model*)
MESSALINE [12]	HWIFI with forcing and insertion	Faults of type *stuck-at-0*, *stuck-at-1*, *logical bridging*, *physical bridging*
AVR-INJECT [4]	SWIFI with exception trigger	Bit-flip faults in the memory area, code area and special registers

prototype, or an emulated execution environment (the last two options are used especially in the pre-deployment phase of the system), and ii) to shed some light on the design choice of a system, for instance, showing its potential dependability bottlenecks. FI tries to determine whether the response of a system matches its specifications in the presence of a defined space of faults.

The implementation of tools for injecting faults has been the focus of several studies. Table I reports a summary of well known tools for fault injection in WSN and in other types of systems. Beyond their inherent differences, they operate in a similar way: each of them performs a study of the fault-free target, obtaining a 'gold file'; then, it injects a fault (obtaining the 'fault file') and it compares the gold file with the fault file, to evaluate the system behavior in response to the fault. Among of the tools mentioned in the table I, there is the AVR-Inject Tool which has been conceived to operate with WSNs. Unfortunately the AVR-Inject tool cannot be used at design time since it needs a prototype of the system, an assembly code that runs on the sensors and thus it needs very detailed information in design phase. Cinque et al. in [13] perform a fault-injection campaign in order to analyze the dependability for three different WSN operating systems (TinyOS, MantisOS, LiteOS). They consider a fault model without taking in account some dependability metric.

Field Failure Data Analysis (FFDA) [14] of a system represents the set of fault forecasting techniques which are performed at runtime. By means of this analysis, the dependability attributes of an actual and deployed system are measured considering real conditions. A system which is in normal operation is observed and the natural occurring errors and failures are monitored and recorded in log files. The FFDA is not practical, not feasible for the WSNs since they do not provide log and they have to be lightweight [15].

Other experimental approaches are described in [16] and [17].

In [16] authors present a deployment of 27 Crossbow Mica2 motes that compose a WSN. They describe a Structure-Aware Self-Adaptive WSN system (SASA) designed in order to detect changes of the network due to unexpected collapses and to maintain the WSN integrity. Detection latency, system errors, network bandwidth and packet loss rate are measured; coverage and connection resiliency metrics are not considered. A large scale simulation is performed in order to evaluate the system scalability and reliability.

Pennington et al. [17] assert that, due to the high complexity of the WSN dynamics, it is difficult to predict problems that may occur after the deployment of the WSN. Therefore, in their paper they propose a Integrity-Checking framework which considers real inputs for the testing and validation process. No case study is shown for framework evaluation.

Experimental approaches for WSN dependability assessment allow to gain insight in the actual failure behavior of WSNs and to establish the reliability degree of the network. However, results are difficult to reproduce and for this reason research studies on WSNs have migrated towards simulative and analytical approaches.

III. SIMULATIVE APPROACHES

A simulative approach for assessing WSNs usually makes use of behavioral simulators, i.e., tools able to reproduce the expected behavior of a system by means of a code-based description. Behavioral simulators allow to reproduce the expected behavior of WSN nodes on the basis of the real application planned to be executed on nodes. However, it is not always possible to observe non-functional properties of WSNs by means of simulative approaches, since models need to be redefined and adapted to the specific network to simulate.

Typical simulative approaches to evaluate WSN fault/failure models are provided in [18] and [19].

In [18] authors address the problem of modeling and evaluating the reliability of the communication infrastructure of a WSN. Authors assume that failures can be categorized in node and network failures.

The first on-line model-based testing technique [19] has been conceived to identify the sensors that have the highest probability to be faulty. The effectiveness of the approach is evaluated in the presence of random noise using a system of light sensors; a fault classification taxonomy is introduced.

Some work like [20] and [21] provide code generation of wireless sensor network applications to perform behavioral simulation and performance analysis.

In [20], a framework for modeling, simulation and code generation of WSNs is presented. The framework is based on Simulink, Stateflow and Embedded Coder; it allows engineers to simulate and automatically generate code of sensor network applications based on MathWorks tools. By means of this tool, an application developer can configure the connectivity of the sensor nodes and can start simulation and functional verification of the application. This framework is able to generate the complete application code for several target operating systems (e.g. TinyOS and MantisOS) from the simulated model.

In [21] a model-driven development process (MDD) is presented to obtain a major effort of optimization for WSN applications. In this work a set of modeling languages is the starting point for code generation and performance analysis.

Finally, the network lifetime is analyzed in [22]; to calculate the lifetime of a WSN, the authors perform simulation by means of a Castalia-based approach [23] that models path-loss.

A. Simulators

Several simulators for WSNs have been proposed in literature, such as *NS-2*, *OMNet++*, *Prowler*, *TOSSIM*, *OPNET* and *Avrora*.

NS-2 [24] is an event-based simulation tool for WSN. It is amply adopted in academic research being open source and easy to use. The simulations are written with C++/C languages and they can be observed graphically by Network AniMator (NAM).

OMNeT++ [25] is a component-based discrete network simulator. Even this simulator is based on C++ language and it has graphical tools for simulation building and evaluating results in real time. The most recent simulation environment built on OMNeT++ is Castalia [23]. This framework was realized for Wireless Sensor Networks, Body Area Networks [26] and networks of low-power embedded devices and it allows to test distributed algorithms and protocols for WSN considering some features of a real WSN like wireless channel, power consumption and considering a real node behavior. Castalia can be used to simulate a wide set of wireless sensor platforms.

Prowler [27] is an event-driven WSN simulator conceived to operate in Matlab environment. Initially it was realized to simulate MICA motes but then it has been extended also for more general platforms. Advantages of Matlab environments are simple implementing of applications, friendly GUI interface and good visualization facilities. By means of this simulator, it is possible to perform deterministic simulation to test application code of a WSN application and to perform probabilistic simulation to observe the behavior of the sensor nodes.

TOSSIM [28], [29] is the simulator built for TinyOS applications. Actually it is an emulator rather than a simulator since it runs actual application code; it allows to simulate the hardware of a sensor but it does not provide information about WSN dependability. Moreover TOSSIM is provided of a visualization tool, TinyViz.

OPNET [30] is a discrete event, object oriented network simulator. This tool was developed initially for military purposes but its large use grew as much to be considered also for commercial use. OPNET is a powerful software that can be used for research purpose and also as a network design tool.

Finally Avrora [31] is a simulator that adopts an approach which is more oriented to the verification of behavioral properties or performance indicators, and not oriented to the observation of dependability properties. It is a low-level emulator of the AVR processor mainly used to test the behavior of WSNs application prior to their deployment. It executes the disassembled code instruction per instruction and emulates the hardware of the processor and the hardware of the node (memory, LEDs, sensors, radio channel, etc.).

IV. Analytical Approaches

The study of the performance and dependability of WSNs can be performed by means of analytical models [32]. Some of these models are based on a mathematical representation of the WSN characteristics and are solved by means of simulation.

In [33] authors introduce an approach for the automated generation of WSN dependability models, based on a variant of Petri nets.

An analytical model to predict the battery exhaustion and the lifetime of a WSN, *LEACH*, is discussed in [34].

In [35] the authors present a network state model used to forecast the energy consumption of a sensor.

AboElFotoh et al. [36] present a probabilistic technique to observe the WSN behavior and discuss about dependability of a WSN; they suppose that the main causes of the failures are related to the crashes, power failures and natural causes. The authors evaluate dependability on the basis of the number of packets received by the sink in a deterministic time (*decision interval*). The dependability is computed evaluating the delay of the expected message.

In [37] authors develop an analytical model to investigate the relation between energy saving and system performance and to observe the effects when sensor sleep/active mode vary. By means of this model, authors can obtain several performance metrics, such as the distribution of the data delivery delay. This work adopts analytical model specifically representing the sensor in sleep/active mode considering channel contention and routing issues. In this work authors model a WSN by means of Markovian techniques; they assess dependability using data delivery resiliency and power consumption metrics.

A linear programming model [38] is introduced to address the problem of "multi-hop lifetime aware routing". The authors propose a Garg-Konemann-based approach to obtain the minimum cost arborescence for reaching the sink node optimizing the lifetime of sensor nodes.

Finally in [39] the node aging problem is addressed. The authors try to solve this problem by associating a survivor function for each sensor node (using *Weibull* distribution). The aim of this work is to demonstrate that the node aging process has an important impact on the connectivity at the increasing of the hop distance. By means of a mathematical analysis and a simulation, they observe that nodes at first hop consume their energy because of the aggregation with children nodes. Hence, they assert that the consumption is related to the number of children nodes.

Each analyzed work, which applies a simulative or analytical approach, defined its own fault model making simple assumptions on network topology and on power consumption; results cannot be generalized since they are obtained by means of abstract simulations. Thus there is a lack of realistic fault

models and this is a limit of these approaches. Therefore it is necessary a further kind of approach.

V. Formal Approaches

Formal approaches offer a new opportunity for the dependability study of WSNs both before and after its deployment. They are based on formal verification that consists in checking of the correctness of a system taking in account specifications or properties, using formal methods. Until now there is no work that has proven how to use an unique formal approach to perform dependability assessment at design time and runtime.

The formal verification is performed by providing a proof on an abstract mathematical model of the system. Typically to model systems we can consider labeled transition systems, timed automata, finite state machines, Petri nets, process algebra, hybrid automata, formal semantics of programming languages such as axiomatic semantics, operational semantics and denotational semantics.

In this section we focus on main formal approaches proposed in literature such as *model checking* and *Event Calculus*. Then we discuss about papers in which formal methods have been applied for dependability assessment of WSNs.

A. Model Checking

One of the well known formal approaches is *model checking* [40]. This technique consists of a systematically exhaustive exploration of the mathematical model (this is possible for finite models, but also for some infinite models where infinite sets of states can be effectively represented finitely by using abstraction or taking advantage of symmetry). Usually this consists of exploring all states and transitions in the model, by using smart and domain-specific abstraction techniques to consider whole groups of states in a single operation and reduce computing time. Implementation techniques include state space enumeration, symbolic state space enumeration, abstract interpretation, symbolic simulation, abstraction refinement. The properties to be verified are often described in temporal logics, such as linear temporal logic (LTL) or computational tree logic (CTL) [41]. The great advantage of model checking is that it is often fully automatic; its primary disadvantage is that it does not in general scale to large systems; symbolic models are typically limited to a few hundred bits of state, while explicit state enumeration requires the state space being explored to be relatively small.

Typically model checking allows to verify if a defined property of a system is satisfied. Thus, the limit of this technique is related to the prediction of a sequence of events. In other words, by means of model checking, an user is able to control if, given an event, the correctness properties are satisfied but is not able to know what will be the behavior of the system after that given event (e.g. node crash or packet loss).

B. Event Calculus

Event Calculus was proposed for the first time in 1986 by Marek Sergot and Robert Kowalski [42] and then it was extended by Murray Shanahan and Rob Miller in the 1990s [43]. This language belongs to the family of logical languages and it is commonly used for representing and reasoning of the events and their effects [44]. *Fluent*, *event* and *predicate* are the basic concepts of Event Calculus [45]. For every timepoint, the value of fluents or the events that occur can be specified.

This language is also named *narrative-based*: in the Event Calculus, there is a single time line on which events occur and this event sequence represents the *narrative*.

The most important and used predicates of Event Calculus are: *Initiates*, *Terminates*, *HoldsAt* and *Happens*.

Since the normal and failing behavior of a WSN can be characterized in terms of an event flow (for instance, a node is turned on, a packet is sent, a packet is lost, a node stops to work due to crash or battery exhaustion, or it gets isolated from the rest of the network due to the failure of other nodes, etc.), Event Calculus, that is an event-based formal language, can be used to formally specify the occurrence of such events and the response of the WSN to them, to check if given correctness properties are verified. Moreover dependability metrics can be valuated by analyzing the *narrative* generated by a Event Calculus reasoner based on the specification of the target WSN.

Finally several techniques are considered to perform automated reasoning in Event Calculus, such as *satisfiability solving*, *first-order logic automated theorem proving*, *Answer Set Programming (ASP)* and logic programming in *Prolog*.

To check the proposed correctness properties defined in Event Calculus, the most common adopted reasoner is the *Discrete Event Calculus (DEC) Reasoner*. The DEC Reasoner [46], [47] uses satisfiability (SAT) solvers [48] and by means of this we are able to perform reasoning like deduction, abduction, post-diction, and model finding. It is documented in details in [49] in which its syntax is explained (e.g. the meaning of the symbols used in the formulas).

C. Formal approaches for WSN

Lifetime of WSN is defined and evaluated in [50] by means of a mathematical formalism. In this work a generic definition of sensor network lifetime is presented and it is conceived in such way to incorporate different application requirements, such as i) number of alive nodes, ii) time latency in the delivery process, iii) delivery ratio, iv) connectivity, v) coverage, and vi) availability.

Recently, different formal methods and tools have been applied for the modeling and analysis of WSNs, such as [51], [52] and [53].

In [51] authors apply a formal tool to wireless sensor networks, *MEDAL*. They propose a formal language to specify the WSN and a tool to simulate it. However, the formal specification has to be rewritten if the WSN under study changes.

In [52] authors propose a methodology for modeling, analysis and development of WSNs using a formal language (PAWSN) and a tool environment (TEPAWSN). They consider only power consumption as dependability metric that is

necessary but not sufficient to assess the WSN dependability (e.g. other problems of WSN such as the isolation problem of a node have been analyzed) and also they apply only simulation.

In [53] authors describe a model-driven performance engineering framework for WSNs (called Moppet). This framework uses the Event Calculus formalism to estimate the performance of WSN applications in terms of power consumption and lifetime of each sensor node; other dependability metrics like coverage, connection resiliency and data delivery resiliency are not considered. The features related to a particular WSN have to be set in the framework every time that a new experiment starts.

There are some papers ([54],[55],[56]) that have considered the formal method in real-time contexts.

In [54] authors model and study WSN algorithms using the Real-Time Maude formalism. Though authors adopt this formalism, they use NS-2 simulator to analyze the considered scenarios making the work very similar to simulative approaches.

The work presented in [55] describes a new formal model for the specification and the validation of WSN. Authors assert the use of rigorous formal method in specification and validation can help designers to limit the introduction of potentially faulty components during the construction of the system. They consider a WSN as a Reactive Multi-Agent System consisting of concurrent reactive agents. In this paper dependability metrics are not treated and calculated and authors just describe the structure of a Reactive Decisional Agent by means of a formal language. Also, no case studies are reported to validate their proposal.

Patrignani et al. in [56] consider policies to monitor wireless sensor network applications in a WSN middleware characterized by a Component and Policy Infrastructure (CaPI); by means of a formalization they are able to catch dangerous or undesired effects which may compromise the correct behavior of a WSN application. In this work it has been developed a prototype that operates on the basis of a application topology in terms of communicating nodes and a set of properties to satisfy. Even if authors confirm that one of the most important benefits of formal approach is that problems occurring at runtime can be detected, they model a static and not dynamic network configuration, focusing only on security (encryption and decryption messages) and resource usage problems and in their scenario they do not consider other dependability metrics (coverage, data delivery resiliency, ...).

In [57] a methodology to investigate the correctness of the design of a WSN from the point of view of its dependability is proposed. The methodology is based on the *event calculus* formalism and it is backed up by a support tool aimed to simplify its adoption by system designers. The tool allows to specify the target WSN in a user-friendly way and it is able to generate automatically the event calculus specifications used to check correctness properties and evaluate dependability metrics such as coverage and connection resiliency but not data delivery resiliency and power consumption.

TABLE II
APPROACH CLASSIFICATION

Approach	Assessment	
	Design time	Runtime
Experimental	×	✓
Simulative	✓	×
Analytical	✓	×
Formal	✓	✓

VI. DISCUSSION

All the analyzed work provides interesting methods and/or techniques which give a contribution for the dependability assessment in WSN. These methods have been grouped in four categories: experimental, simulative, analytical and formal.

In table II a classification of the presented approaches is shown. Experimental methods are used to evaluate a real system and therefore they need for a existent prototype; they are useful at runtime since through these methods do experiments directly on the real system from which they collect data. Simulative and analytical may be adopted in the design phase: they model a system and make an estimate of reliability before of the system release. Finally formal methods make use of correctness specifications and they can be used at design time and at runtime too by means of runtime verification techniques.

Moreover, in this section, it is shown and discussed a comparison of the related work presented in the previous sections in which it emerges a lack of a work that allows to perform WSN dependability assessment both at design and at runtime.

In the grid, shown in figure 1, on the rows there is the analyzed work (approaches, tools and models); on the columns there are the properties chosen to highlight the differences.

In particular we have considered the following features:

- *Experimental Approach* to determine if the related work is based on experiments;
- *Simulative Approach* to determine if the related work is based on simulations;
- *Analytical Approach* to determine if the related work is based on analytical models;
- *Formal Approach* to determine if the related work is based on some formal method (e.g. model checking, Event Calculus) and in particular if the work adopts an approach that provides *Separated specifications*: we want to verify if the related work applies a modular solution considering two logical sets of specifications: a general correctness specification, valid independently of the particular WSN under study, and a structural specification related to the properties of the target WSN (e.g., number of nodes, topology, channel quality, initial battery charge);
- *Design time* to determine if the related work performs dependability assessment at design time;
- *Runtime* to determine if the related work performs dependability assessment at runtime;
- *WSN Dependability metrics* to determine if the related work considers the following dependability metrics: coverage, connection resiliency, data delivery resiliency,

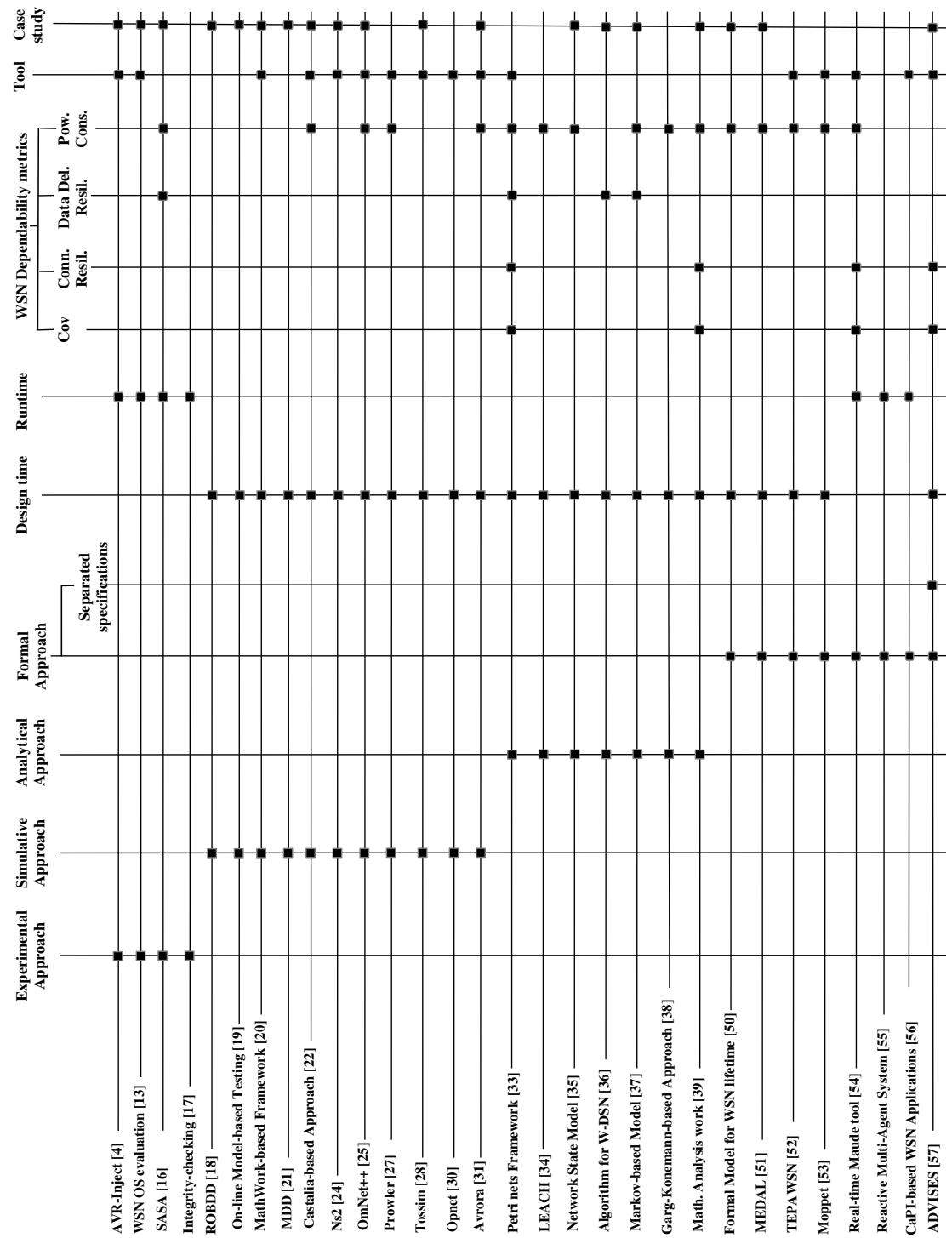


Fig. 1. Comparison of WSN dependability assessment studies

power consumption;
- *Tool* to determine if the related work proposes a novel tool to support designers;
- *Case study* to determine if the related work considers case studies in order to validate the proposed work.

From the survey of the literature it is possible to assert that among the most important dependability metrics, the power consumption is the only one that has been considered extensively, instead data delivery resiliency and connection resiliency are the least analyzed.

The majority of papers propose a tool and present results by means of a case study observing the behavior of the WSN under determined circumstances.

Therefore, looking the figure 1, there is no work that describes a framework conceived in order to perform WSN dependability assessment both at design and runtime measuring all the main dependability metrics. Many studies address the WSN dependability assessment at design time, few studies at runtime.

Moreover, we think that formal methods may be considered as a new and attractive solution for the assessment of dependability both at design time that at runtime by defining one specification for the system suitable for both purposes since the lack of a formal approach that can be applied for doing static and dynamic assessment of WSN dependability remains an open issue.

Thus, in the field of WSN research, a study of a framework that applies an approach to assess WSN dependability by means of a formal approach, before and after the deployment of a WSN, can be advantageous and innovative.

VII. Conclusions

In this paper, we have reported a survey on the approaches of WSN dependability assessment grouped in experimental, simulative, analytical and formal. What appears clear is that the path towards the production of an optimal approach to check the dependability level of WSN both at design and runtime is still long, and more research effort is needed to reach this compelling goal.

To achieve this goal, we think that applying formal techniques is a good approach since they could join the benefits of the experimental approaches (for dependability evaluation at runtime) and the simulative and analytical approaches (for dependability evaluation at design time). The idea of performing a complete check of the dependability degree on the WSN behavior, to enforce the fulfillment of correctness properties, seems a promising one to achieve more stable and dependable WSN-based systems in the future.

References

[1] C. Di Martino, G. D'Avino, and A. Testa, "icaas: An interoperable and configurable architecture for accessing sensor networks," *International Journal of Adaptive, Resilient and Autonomic Systems (IJARAS)*, vol. 1, no. 2, pp. 30–45, 2010.

[2] A. Coronato, G. Pietro, J.-H. Park, and H.-C. Chao, "A framework for engineering pervasive applications applied to intra-vehicular sensor network applications," *Mobile Networks and Applications*, vol. 15, no. 1, pp. 137–147, 2010. [Online]. Available: http://dx.doi.org/10.1007/s11036-009-0163-8

[3] A. Coronato, G. De Pietro, and M. Esposito, "A semantic context service for smart offices," in *Hybrid Information Technology, 2006. ICHIT '06. International Conference on*, vol. 2, 2006, pp. 391–399.

[4] M. Cinque, D. Cotroneo, C. Di Martino, S. Russo, and A. Testa, "Avr-inject: A tool for injecting faults in wireless sensor nodes," in *Parallel Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, may 2009, pp. 1 –8.

[5] M. Cinque, D. Cotroneo, C. Di Martino, and A. Testa, "An effective approach for injecting faults in wireless sensor network operating systems," in *Computers and Communications (ISCC), 2010 IEEE Symposium on*. IEEE, 2010, pp. 567–569.

[6] M. Cinque, "Dependability evaluation of mobile distributed systems via field failure data analysis," Ph.D. dissertation, PhD Thesis, Dipartimento di Informatica e Sistemistica, Universitá di Napoli Federico II, www. mobilab. unina. it/tesiDottorato. html, 2006.

[7] M.-C. Hsueh, T. K. Tsai, and R. K. Iyer, "Fault injection techniques and tools," *Computer*, vol. 30, no. 4, pp. 75–82, 1997.

[8] H. Madeira and J. G. Silva, "Xception: Software fault injection and monitoring in processor functional units," in *in Processor Functional Units, DCCA-5, Conference on Dependable Computing for Critical Applications*, 1995, pp. 135–149.

[9] G. Kanawati, N. Kanawati, and J. Abraham, "Ferrari: a flexible software-based fault and error injection system," *Computers, IEEE Transactions on*, vol. 44, no. 2, pp. 248–260, Feb 1995.

[10] J. Barton, E. Czeck, Z. Segall, and D. Siewiorek, "Fault injection experiments using fiat," *Computers, IEEE Transactions on*, vol. 39, no. 4, pp. 575–582, Apr 1990.

[11] D. T. Stott, B. Floering, D. Burke, Z. Kalbarczyk, and R. K. Iyer, "Nftape: A framework for assessing dependability in distributed systems with lightweight fault injectors," in *In Proceedings of the IEEE International Computer Performance and Dependability Symposium*, 2000, pp. 91–100.

[12] J. Arlat, M. Aguera, L. Amat, Y. Crouzet, J.-C. Fabre, J.-C. Laprie, E. Martins, and D. Powell, "Fault injection for dependability validation: a methodology and some applications," *Software Engineering, IEEE Transactions on*, vol. 16, no. 2, pp. 166–182, Feb 1990.

[13] M. Cinque, C. D. Martino, and A. Testa, "Analyzing and modeling the failure behavior of wireless sensor networks software under errors," in *IWCMC*, 2012, pp. 1136–1141.

[14] G. Carrozza and M. Cinque, "Modeling and Analyzing the Dependability of Short Range Wireless Technologies via Field Failure Data Analysis," *Journal of Software*, vol. 4, pp. 707–716, 2009.

[15] G. Carrozza, "Software faults diagnosis in complex ots-based critical systems," Ph.D. dissertation, PhD Thesis, Dipartimento di Informatica e Sistemistica, Universitá di Napoli Federico II, www. mobilab. unina. it/tesiDottorato. html, 2008.

[16] M. Li and Y. Liu, "Underground coal mine monitoring with wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 5, no. 2, pp. 10:1–10:29, Apr. 2009. [Online]. Available: http://doi.acm.org/10.1145/1498915.1498916

[17] S. Pennington, T. Bauge, and B. Murray, "Integrity-checking framework: An in-situ testing and validation framework for wireless sensor and actuator networks," in *Sensor Technologies and Applications, 2009. SENSORCOMM '09. Third International Conference on*, 2009, pp. 575–579.

[18] A. Shrestha, L. Xing, and H. Liu, "Infrastructure communication reliability of wireless sensor networks," in *Dependable, Autonomic and Secure Computing, 2nd IEEE International Symposium on*, 29 2006-oct. 1 2006, pp. 250 –257.

[19] F. Koushanfar, M. Potkonjak, and A. Sangiovanni-Vincentelli, "On-line fault detection of sensor measurements," in *Sensors, 2003. Proceedings of IEEE*, vol. 2, oct. 2003, pp. 974 – 979 Vol.2.

[20] M. Mozumdar, F. Gregoretti, L. Lavagno, L. Vanzago, and S. Olivieri, "A framework for modeling, simulation and automatic code generation of sensor network application," in *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on*, june 2008, pp. 515 –522.

[21] R. Shimizu, K. Tei, Y. Fukazawa, and S. Honiden, "Model driven development for rapid prototyping and optimization of wireless sensor network applications," in *Proceedings of the 2nd Workshop on Software Engineering for Sensor Network Applications*, ser. SESENA '11. New York, NY, USA: ACM, 2011, pp. 31–36. [Online]. Available: http://doi.acm.org/10.1145/1988051.1988058

[22] K. Doddapaneni, E. Ever, O. Gemikonakli, I. Malavolta, L. Mostarda, and H. Muccini, "Path loss effect on energy consumption in a wsn," in *Computer Modelling and Simulation (UKSim), 2012 UKSim 14th International Conference on*, march 2012, pp. 569 –574.

[23] A. Boulis, "Castalia: revealing pitfalls in designing distributed algorithms in wsn," in *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*. New York, NY, USA: ACM, 2007, pp. 407–408.

[24] "The Network Simulator NS-2," http://www.isi.edu/nsnam/ns/.

[25] C. Mallanda, A. Suri, V. Kunchakarra, S. S. Iyengar, R. Kannan, A. Durresi, and S. Sastry, "Simulating Wireless Sensor Networks with OMNeT++." [Online]. Available: http://csc.lsu.edu/sensor_web/final%20papers/OMNet++-IEEE-Computers.pdf

[26] K. Y. Yazdandoost, K. Sayrafian-Pour *et al.*, "Channel model for body area network (ban)," *IEEE P802*, vol. 15, 2009.

[27] A. K. Sharma and D. Gupta, "Performance evaluation of routing protocols for wsns based on energy-aware routing with different radio models," *International Journal of Computer Applications*, vol. 3, no. 12, pp. 6–14, July 2010, published By Foundation of Computer Science.

[28] P. Levis and N. Lee, *Tossim: A simulator for tinyos networks*, 2003, p. 24.

[29] P. Levis, N. Lee, M. Welsh, and D. Culler, "Tossim: accurate and scalable simulation of entire tinyos applications," in *Proceedings of the 1st international conference on Embedded networked sensor systems*, ser. SenSys '03. New York, NY, USA: ACM, 2003, pp. 126–137. [Online]. Available: http://doi.acm.org/10.1145/958491.958506

[30] P. Jurčík and A. Koubâa, "The ieee 802.15.4 opnet simulation model: Reference guide v2.0," [online] http://www.open-zb.net/publications/HURRAY_TR_070509.pdf, IPP-HURRAY!, Tech. Rep., May 2007.

[31] B. L. Titzer, D. K. Lee, and J. Palsberg, "Avrora: scalable sensor network simulation with precise timing," in *Proceedings of the 4th international symposium on Information processing in sensor networks*, ser. IPSN '05. Piscataway, NJ, USA: IEEE Press, 2005. [Online]. Available: http://dl.acm.org/citation.cfm?id=1147685.1147768

[32] C. Di Martino, "Resiliency assessment of wireless sensor networks: a holistic approach," Ph.D. dissertation, PhD Thesis, Dipartimento di Informatica e Sistemistica, Universitá di Napoli Federico II, www. mobilab. unina. it/tesiDottorato. html, 2009.

[33] C. Di Martino, M. Cinque, and D. Cotroneo, "Automated generation of performance and dependability models for the assessment of wireless sensor networks," *IEEE Trans. Comput.*, vol. 61, no. 6, pp. 870–884, Jun. 2012. [Online]. Available: http://dx.doi.org/10.1109/TC.2011.96

[34] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, jan. 2000, p. 10 pp. vol.2.

[35] A. F. Mini, B. Nath, and A. A. F. Loureiro, "A probabilistic approach to predict the energy consumption in wireless sensor networks," in *In IV Workshop de Comunicao sem Fio e Computao Mvel. So Paulo*, 2002, pp. 23–25.

[36] H. AboElFotoh, S. Iyengar, and K. Chakrabarty, "Computing reliability and message delay for cooperative wireless distributed sensor networks subject to random failures," *Reliability, IEEE Transactions on*, vol. 54, no. 1, pp. 145 – 155, march 2005.

[37] C.-F. Chiasserini and M. Garetto, "Modeling the performance of wireless sensor networks," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 1, march 2004, pp. 4 vol. (xxxv+2866).

[38] J. Stanford and S. Tongngam, "Approximation algorithm for maximum lifetime in wireless sensor networks with data aggregation," in *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2006. SNPD 2006. Seventh ACIS International Conference on*. IEEE, 2006, pp. 273–277.

[39] J.-J. Lee, B. Krishnamachari, and C.-C. J. Kuo, "Impact of energy depletion and reliability on wireless sensor network connectivity," in *Proceedings of SPIE*, vol. 5440, 2004, pp. 169–180.

[40] A. Biere, A. Cimatti, E. M. Clarke, O. Strichman, and Y. Zhu, *Bounded model checking*. Elsevier, 2003, vol. 58.

[41] T. Hafer and W. Thomas, "Computation tree logic ctl and path quantifiers in the monadic theory of the binary tree," *Automata, Languages and Programming*, pp. 269–279, 1987.

[42] R. Kowalski and M. Sergot, "A logic-based calculus of events," *New Gen. Comput.*, vol. 4, no. 1, pp. 67–95, Jan. 1986. [Online]. Available: http://dx.doi.org/10.1007/BF03037383

[43] R. Miller and M. Shanahan, "Reasoning about discontinuities in the event calculus," in *in Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*. Morgan Kaufmann, 1996, pp. 63–74.

[44] F. Van Harmelen, V. Lifschitz, and B. Porter, *Handbook Of Knowledge Representation*, ser. Foundations of Artificial Intelligence. Elsevier, 2008. [Online]. Available: http://www.worldcat.org/isbn/0444522115

[45] M. Shanahan, "The Event Calculus Explained," *Lecture Notes in Computer Science*, vol. 1600, pp. 409–430, 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.3267

[46] E. T. Mueller, "Event calculus reasoning through satisfiability," *Journal of Logic and Computation*, vol. 14, p. 2004, 2004.

[47] E. Mueller, "Decreasoner," http://decreasoner.sourceforge.net.

[48] E. T. Mueller, "A tool for satisfiability-based commonsense reasoning in the event calculus," in *FLAIRS Conference'04*, 2004, pp. –1–1.

[49] E. T. Muller, "Discrete event calculus reasoner documentation," p. http://decreasoner.sourceforge.net/csr/decreasoner.pdf, 2008.

[50] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *Communications Letters, IEEE*, vol. 9, no. 11, pp. 976 – 978, nov. 2005.

[51] K. Kapitanova and S. Son, "Medal: A compact event description and analysis language for wireless sensor networks," in *Networked Sensing Systems (INSS), 2009 Sixth International Conference on*, 2009, pp. 1–4.

[52] K. L. Man, T. Vallee, H. Leung, M. Mercaldi, J. van der Wulp, M. Donno, and M. Pastrnak, "Tepawsn - a tool environment for wireless sensor networks," *Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on*, pp. 730–733, May.

[53] P. Boonma and J. Suzuki, "Moppet: A model-driven performance engineering framework for wireless sensor networks," *Comput. J.*, vol. 53, no. 10, pp. 1674–1690, 2010.

[54] P. Ölveczky and J. Meseguer, "Specification and analysis of real-time systems using real-time maude," *Fundamental Approaches to Software Engineering*, pp. 354–358, 2004.

[55] R. Romadi and H. Berbia, "Wireless sensor network a specification method based on reactive decisional agents," in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, april 2008, pp. 1 –5.

[56] M. Patrignani, N. Matthys, J. Proenca, D. Hughes, and D. Clarke, "Formal analysis of policies in wireless sensor network applications," in *Software Engineering for Sensor Network Applications (SESENA), 2012 Third International Workshop on*, june 2012, pp. 15 –21.

[57] A. Testa, A. Coronato, M. Cinque, and J. C. Augusto, "Static verification of wireless sensor networks with formal methods," in *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on*. IEEE, 2012, pp. 587–594.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 889–894

Analysis of the influence of radio beacon placement on the accuracy of indoor positioning system

Krzysztof Piwowarczyk, Piotr Korbel, Tomasz Kacprzak
Institute of Electronics,
Lodz University of Technology,
ul. Wólczańska 211/215, 90-924 Łódź, Poland
Email: krzysiekpiwo@gmail.com, piotr.korbel@p.lodz.pl, tomasz.kacprzak@p.lodz.pl

***Abstract*—This paper discusses factors influencing accuracy of estimating localization of radio networks terminals in indoor environment. It introduces parameters that can be useful to describe the quality of localization of radio landmarks. The paper presents a software for computer aided reference radio stations placement inside the buildings and shows the results of exemplary simulations carried out with the use of proposed algorithms.**

***Index Terms*—indoor positioning systems, Location Based Services, wireless local area networks, radiolocation**

I. Introduction

NOWADAYS many telecommunication systems use information about location of mobile terminal from GPS and GSM/UMTS systems to estimate user terminal location. But these systems have common defect, both do not work indoors. These limitation results from strong signal attenuation introduced by the outer walls of buildings and from the strong multipath propagation effects present in indoor areas. To overcome these problems there are created dedicated systems like sensor networks and used dedicated networks like WLAN, Bluetooth, ZigBee, RFID, UWB to work in a small area by having a lower signal strength which causes smaller signal interference in comparison to outdoor solutions. Many of these network systems are used in daily life to exchange information between terminals. But due to the growing interest in localization in indoor environment, a problem arises how to fast and correct locate reference landmarks, taking into consideration size of the area in which localization is possible, number of reference nodes and desired localization accuracy. If we place reference nodes correctly we can use these systems for example to help a blind or visually impaired person to move inside the building, or to help a disabled person to reach the medical room. Another important field of application of positioning systems includes navigation in industrial and manufacturing facilities [1].

This article discusses the factors influencing on location accuracy in indoor environment based on the RSS (Received Signal Strength). Then, it explains how to evaluate the quality of the reference station distribution based on the aforementioned factors. Next, it presents an algorithm which improves quality of placement of the reference nodes taking into consideration the aforementioned assessment methods. Then it deals with the results of the simulations in exemplary rooms with obstacles.

II. Factors Influencing Localization Accuracy

A. Lack of signals from minimum three reference radio stations

This is the most important factor that influences the possibility to estimate unknown position of the mobile terminal. When we have information about the distance from one reference station, we can only say that the terminal is somewhere around the circle with radius being equal to this distance. If we have information on the distances from two reference stations, we can limit our searching to two points resulting from intersection of two circles. To obtain unambiguous information about position of the terminal, information from at least three reference stations is required.

B. Adverse reference nodes geometry

Distance measurements between reference stations and localized terminal is not noiseless. The size of noise depend on the strength and types of signal used and environment surrounding reference station. All it caused a distance measurements error, that is shown as ring around the station, Fig. 1. As a result of imposition of three rings, we can indicate common areas where the terminal is likely to be localized in. This area is called the area of uncertainty, Fig. 1.

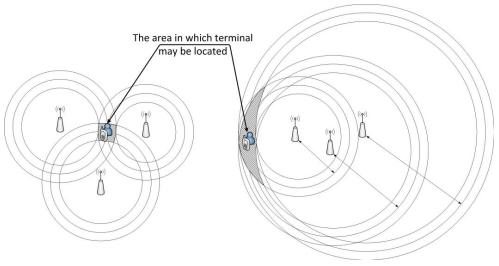


Fig. 1. Localization error for different geometry of reference nodes a) nodes located in an equilateral triangle b) nodes located on straight line.

The size of the area of uncertainty represents, in a sense, the error of localization. The uncertainty area is the smallest when the localized terminal is placed between the reference stations and the angles between adjacent stations are equal. On the contrary, the biggest area of uncertainty is created when the reference stations are placed along the line. This results in the highest error of unknown terminal localization.

978-83-60810-53-8/$25.00

C. Multipath propagation of radio signals

Multipath propagation results in imposition of several copies of the same signal reaching the receiver. The signal components travel along different paths and exhibit different power levels and phase shifts. This leads to the strong fluctuations of signal power, depending on the distance between the transmitter and the receiver [2], [3]. This factor has a particular importance in the localization methods based on the received signal strength and phase measurements [4], [5]. A graph shown in Fig. 2 presents the level of signal attenuation as a function of the distance between the antennas. Fluctuations of the signal power result from interfering of two waves: direct and reflected ones.

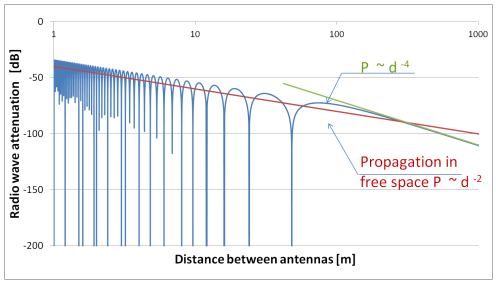


Fig. 2. Signal attenuation as a function of transmitter and receiver antennas separation

However, in a real indoor environment, the transmitting and receiving antennas are surrounded by walls and furnishing. In such conditions signals reflected from the obstacles reach the receiver, propagating along many different paths, thus having different phases. The number of interfering waves can be very high and difficult to estimate. In fact, there is no visible relationship between received signal strength and the distance, which makes difficult to determine the distance using the RSS method correctly. As a result, the calculated position shows significant errors. Another problem is the correct estimation of reflection coefficient of the radio wave for the obstacle. A huge coefficient variation dependent on the material which the obstacle is made of prevents from the accurate calculation of the radio wave reflected from the obstacle attenuation.

D. Radio signal attenuation in non line of sight case

If a line of sight (LOS) between localized object and reference station exists, we can assume that the measurement of signal strength is not affected by additional errors resulting from unspecified signal attenuation and delays due to reflections. In indoor environment it is often not possible to provide the LOS conditions in the entire room. In non line of sight case (NLOS), the signal reaches the receiver with additional attenuation resulting in the increase of the localization error.

E. Incorrect or inaccurate signal propagation model

Incorrectly chosen model of radio wave propagation, based on RSS method, can be a source of significant errors in terminal position calculation. The choice of appropriate propagation model should strictly depend on indoor environmental conditions. In practice, there is no ability to ensure that we choose the proper model, corresponding to temporary environmental conditions. For this purpose there are commonly used empirical models, based on a large number of measurements and statistical surveys.

III. Assessing the Distribution of Reference Stations

A. Evaluation of reference nodes geometry

To evaluate reference stations geometry in relation to the localized point we used two factors have been used: the first factor we propose is a simple formula determined on the trigonometric formulas. The second one is a modified factor HDOP (Horizontal Dilution Of Precision) used in GPS system to assess the layout geometry of satellites [6], [7], [8]. Fig. 3 shows the case where the terminal position is determined based on signals coming from four reference stations (according to [1]).

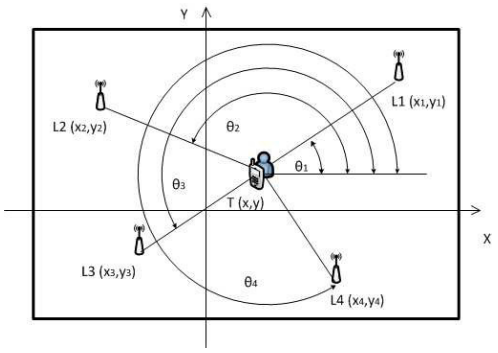


Fig. 3. Two dimensional localization schema.

In Fig. 3 the angle θ formed between the straight line passing through the localized terminal and a line parallel to the x axis are marked.
We assume that α is the angle between neighboring reference stations:

$$\alpha_1 = \theta_2 - \theta_1\,, \quad \alpha_2 = \theta_3 - \theta_2\,, \quad \alpha_3 = \theta_4 - \theta_3\,; \qquad (1)$$

The factor g(l) is obtained on the basis of trigonometric formulas:

$$g(l) = \frac{\sin^2\varphi_1 + \sin^2\varphi_2 + \sin^2\varphi_3 + \ldots + \sin^2\varphi_n}{n} \qquad (2)$$

where:

$$\varphi_n = \alpha_n - \left(\frac{2\pi}{n} - \frac{\pi}{2}\right) \qquad (3)$$

and:
α_n - the angle between the straight lines connecting the localized terminal with neighboring reference stations,
n - number of stations within the range of the terminal being localized.

The classification of the assumed geometry indexes is presented in Table I. This index measures the effect of the

TABLE I
EVALUATION OF GEOMETRY INDEX VALUES

Value of g(l) index	Evaluation of the geometry of the reference stations placement
0,5<g(l)≤1	Very good
0,35<g(l)≤0,5	Good
0,2<g(l)≤0,35	Sufficient
g(l)≤0,2	Bad

geometric configuration of the reference points on the position estimation [8].

The second factor that can be used to evaluate the distribution of radio stations is HDOP index, which is typically used to assess the potential localization error of GPS satellite navigation systems. HDOP factor for the three-dimensional case can be derived from matrix A:

$$\mathbf{A} = \begin{vmatrix} \frac{(x_1-x)}{R_1} & \frac{(y_1-y)}{R_1} & \frac{(z_1-z)}{R_1} & -1 \\ \frac{(x_2-x)}{R_2} & \frac{(y_2-y)}{R_2} & \frac{(z_2-z)}{R_2} & -1 \\ \frac{(x_3-x)}{R_3} & \frac{(y_3-y)}{R_3} & \frac{(z_3-z)}{R_3} & -1 \end{vmatrix} \tag{4}$$

where,

$$R_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \tag{5}$$

and:
R_i - the distance between the coordinates (x_i, y_i, z_i) of the reference station and coordinates (x, y, z) of object to be localized. Introducing matrix Q:

$$\mathbf{Q} = (\mathbf{A}^T \mathbf{A})^{-1} \tag{6}$$

$$\mathbf{Q} = \begin{vmatrix} d_x^2 & d_{xy}^2 & d_{xz}^2 & d_{xt}^2 \\ d_{xy}^2 & d_y^2 & d_{yz}^2 & d_{yt}^2 \\ d_{xz}^2 & d_{yz}^2 & d_z^2 & d_{zt}^2 \\ d_{xt}^2 & d_{yt}^2 & d_{zt}^2 & d_t^2 \end{vmatrix} \tag{7}$$

the HDOP index for two dimensional case is determined.

$$HDOP = \sqrt{d_x^2 + d_y^2} \tag{8}$$

Based on the HDOP index, we can specify geometry index factor according to Table II [8].

TABLE II
EVALUATION OF HDOP INDEX VALUES

Value of g(l) index	Evaluation of the geometry of the reference stations placement
<1	Measurements error or redundancy
1	Ideal
1-2	Very Good
2-5	Good
5-10	Medium
10-20	Sufficient
>20	Bad

B. Quality of radio signal in non line of sight case

To evaluate the quality of radio signals reaching the receiver, we additionally assumed the criterion of maximum acceptable additional signal attenuation due to the presence of obstacles between the localized terminal and a reference station. We assumed the following classification of radio propagation conditions:

- conditions are considered to be ideal when there is no additional attenuation in radio path caused by obstacles,
- conditions are very good, if the additional attenuation in radio path caused by obstacles does not exceed 2 dB, for each station,
- conditions are good, if the additional attenuation caused by an obstacle exceed 2dB is from 0% to 25% number of stations
- conditions are bad, if the additional attenuation caused by an obstacle is greater than 2 dB is from 25% to 50% number of stations,
- conditions are very bad, if the additional signal attenuation caused by obstacles is greater than 2 dB from more than 50% of the stations.

IV. ANALYSIS OF THE REFERENCE NODES PLACEMENT

A. Assumptions

The following assumptions regarding the assessment of reference radio stations placement have been taken into account in the article.

- In every part of room, radio signals from at least three reference radio stations should reach the receiver. It is also assumed that the number of stations within range should not exceed four. The increase of the number of stations does not necessarily improves the location accuracy but significantly increases the cost of network construction.
- The power levels of the received signals should exceed some minimum threshold level.
- Reference stations should be as far separated as possible, which allows to reduce the spatial density of the stations.
- Another important aspect is to provide direct visibility of reference stations (LOS) at each point in the room. Alternatively, signal attenuation due to the obstacles should not exceed the assumed maximum level.
- In the ideal case the angles between the reference stations in relation to the terminal position should be equal.

B. Simulation software

To examine the impact of the reference stations placement on the localization accuracy, a dedicated simulation software has been developed. The software allows evaluation of reference nodes placement for an assumed two-dimensional room layout using the criteria presented in Section III. Apart from the assessment of the reference nodes placement, the software allows to optimize the initial locations of the nodes in order to maximize the values of quality of placement coefficients. We implemented an iterative algorithm based on the idea

described in [9], [10]. Every placement of the reference nodes can be described with an overall system of a so-called "energy function". The energy of the system is computed as the sum of the energies assigned to test locations in the room. The test locations are randomly distributed within the area under investigation, and the energy assigned to a single location is a function of coefficients defined in Section III.

$$E = \sum_{n=1}^{N} \quad (a_1 Q_1 + a_2 Q_2 + a_3 Q_3 + \cdots + a_1 Q_1) \qquad (9)$$

where:
E - an overall system energy,
Q_k - value of quality factor,
a_k - weight quality factor,
N - number of all measurement points,
n - measuring point number,

The model of the system is based on resilience phenomenon, i.e. when the distance between the two neighboring reference nodes increases, the force repelling these nodes decreases. The goal of the algorithm is to minimize the overall system energy and to maximize coefficients describing the quality of the reference nodes placement. In every iteration, the software estimates reference nodes placement quality coefficients as well as system energy and forces repelling the neighboring nodes. Additionally, a random Brownian motion of the nodes is assumed to minimize the risk of stopping the optimization algorithm in some local minima. As a result, new positions of the reference stations are estimated.

$$V = \sum_{n=1}^{N} \quad (ad_n + b) + V_B) \qquad (10)$$

where:
N - number station in range
a, b - coefficient
d - distance between stations
V_B - Brown motion vector

The flow diagram of the reference nodes placement optimization algorithm implemented in the software is presented in Fig. 4.

In the simulations, to calculate signal attenuation and to estimate the distance between the antennas a Multi-Wall indoor radio propagation model [2], [3] was used. For distances d < 1m we assume free space signal propagation loss:

$$L(d < 1)_{dB} = 40 + 20 \log(d) \qquad (11)$$

For distances d > 1m we compute the propagation loss with the use of the Multi-Wall model:

$$L_{MWM}[dB] = 40 + 10 \cdot 4 \log(d) + \sum_{s=1}^{s=Sn} \quad (n_{ws} \cdot L_{ws}) \qquad (12)$$

where:
L - signal attenuation,

Fig. 4. Reference stations placement optimization algorithm.

d - distance between the antennas,
n - number of walls on the signal propagation path,
S - type of wall material is made,
Sn - number of wall types,
Lws - attenuation of a single wall,

C. Simulation results

Firstly, we try to see the relationship between placement of the reference stations and the size of the area where the terminal can receive signals from at least three of the beacon stations. For the simulations we assumed an example of 25 m x 25 m room layout and the following algorithm parameters: transmit power level is equal to 5 dBm and minimum received signal strength required at the receiver antenna equal to -85 dBm. Fig. 5 and Fig. 6 present results of the reference stations placement. Black rectangles denote obstacles and blue circles indicate reference stations. Areas where the quality requirements are not met are marked with crosses.
To assess localization conditions, we evaluate quality criteria for test points distributed in the area of the room layout. The test points are located randomly in the room and the number of points is determined by the program user. The higher the number of points, the greater the accuracy but also the longer the computation time. Fig. 7 shows sample results obtained for the scenario presented in Fig. 6, and the bars indicate percentage of the room area as a function of a number of visible reference stations.

To evaluate the performance of reference stations placement optimization algorithm, we performed a series of simulations for sample room layout and for various initial placements of

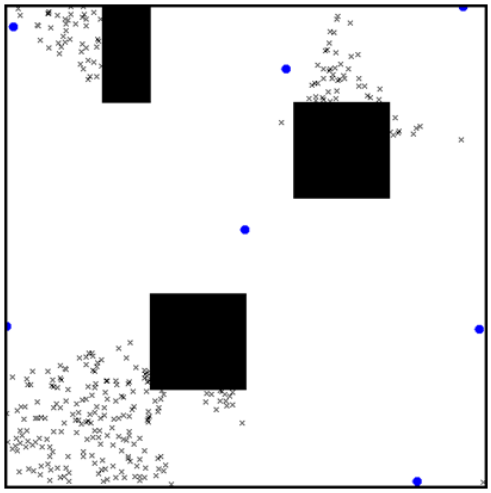


Fig. 5. Example placement of 7 reference stations in room No.1 25m x 25m.

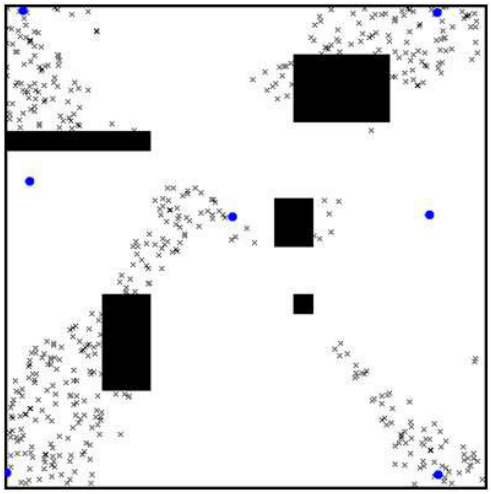


Fig. 6. Example placement of 7 reference stations in room No.2 25m x 25m.

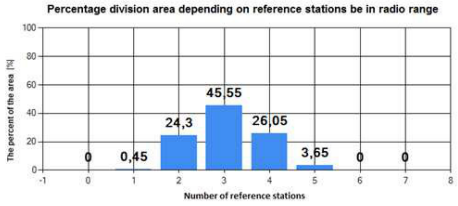


Fig. 7. Percentage division area depending on reference stations being in the radio range.

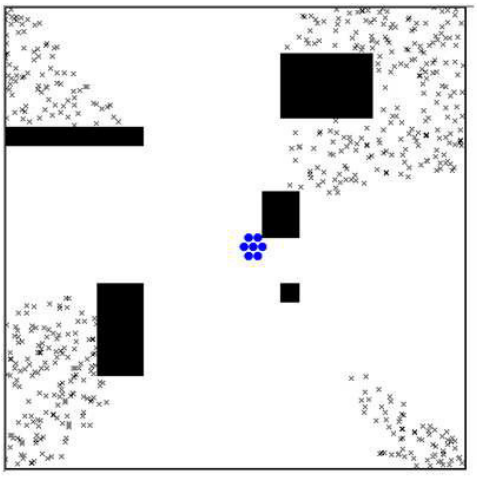


Fig. 8. Initial simulation conditions ij a room with a number of reference stations placed in its center.

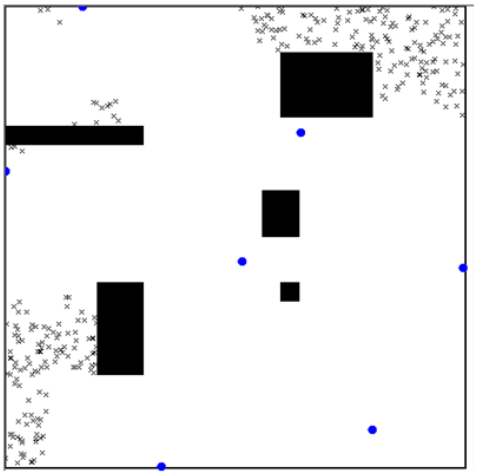


Fig. 9. The result of the reference stations placement optimization.

the stations. Fig. 8 and Fig. 9 show the initial conditions and optimization results for the same room layout as in Fig. 6.

V. Summary

In the paper, we described parameters that can be used to evaluate the quality of reference nodes placement and its influence on the network terminal positioning accuracy. We also presented a software for computer aided optimization of the reference stations of indoor positioning systems. The software implements an optimization algorithm based on resilience phenomenon and Brownian motion model. The results of simulations carried out for sample room layouts proved the suitability of the proposed approach to finding such locations of the reference stations that maximize positioning quality criteria. However, it must be noted that the resulting reference stations placement depends on the initial positions of the nodes. The goal of the further research is to evaluate results the optimization of reference stations placement for sample real

life scenarios. Measurement campaigns verifying simulation results are planned in a university campus buildings. The measurement data will be also used to further adjust the proposed algorithms.

REFERENCES

[1] J. Zhou, J. Shi, X. Qu, "Landmark Placement for Wireless Localization in Rectangular-Shaped Industrial Facilities", IEEE Transactions on Vehicular Technology, Vol. 59, No. 6, 2010, pp. 3081–3090.

[2] S. R. Saunders, A. Aragon-Zavala, *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons Ltd, 2007.

[3] R. J. Katulski, *Radio Wave Propagation for Wireless Communications*, (*Propagacja fal radiowych w telekomunikacji bezprzewodowej*), Wydawnictwo Komunikacji i Łączności, Warsaw, Poland, 2009. (in Polish)

[4] Yihong Qi, Hisashi Kobayashi, "On geolocation accuracy with prior information in non-line-of-sight environment", *in Vehicular Technology Conference, Proceedings, VTC 2002-Fall,* 2002 IEEE 56th, pp. 285–288, 2002.

[5] D. Jourdan, D. Dardari, and M.Z. Win, "Position Error Bound and Localization Accuracy Outage in Dense Cluttered Environments", *in The 2006 IEEE 2006 International Conference on Ultra-Wideband,* pp. 519–524, 2006.

[6] R. Yarlagadda, I. Ali, N. Al-Dhahir, and J. Hershey, "GPS GDOP metric", Radar, Sonar and Navigation, IEE Proceedings, vol.147, no.5, 2000, pp. 259–264.

[7] N. Levanon, "Lowest GDOP in 2-D scenarios", Radar, Sonar and Navigation, IEE Proceedings, vol.147, no.3, 2000, pp. 149–155.

[8] K. Bronk, J. Stefański, "Analysis of the influence of measurement stations configuration on TDOA based positioning", ("Analiza wpływu konfiguracji stacji pomiarowych na dokładność pozycjonowania w metodzie TDOA", *in Proceedings KKRRiT 2007, Zeszyty Naukowe Politechniki Gdańskiej,* 2007. (in Polish)

[9] M. Rutecki, T. Kacprzak, "Optimization of radio base station placement for localization of moving terminals indoors", *in 17th International Conference on Microwaves, Radar and Wireless Communications MIKON,* Wroclaw, Poland, 2008.

[10] K. Piwowarczyk, *Analysis of influence of radio beacon placement on accuracy of localization of radio network terminals*, MSc Thesis, Lodz University of Technology, Łódź, Poland, 2012. (in Polish)

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 895–900

Development of Special Smartphone-Based Body Area Network: Energy Requirements

Jana Púchyová
University of Žilina
Faculty of Management Science and Informatics
Department of Technical Cybernetics
Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: jana.puchyova@fri.uniza.sk

Michal Kochláň, Michal Hodoň
University of Žilina
Faculty of Management Science and Informatics
Department of Technical Cybernetics
Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: {michal.kochlan, michal.hodon}@fri.uniza.sk

***Abstract*—In recent years, smart-devices became very popular among people of all ages around the world. Very important is especially their usage in health applications. Special Body Area Network (BAN) for the stress monitoring is currently being developed within the authors' department. Android-based smartphone is employed as the main control unit of the sensor network built on the star architecture. Since the power consumption of the smart-phone as well as of the single sensor node is one of the key limitations of the network, special attention has to be given on it. In this article, energy requirements necessary for the data transmission among the network is analysed in detail. For this purpose, communication solution based on 2.4 GHz proprietary RF transceiver is implemented.**

I. Introduction

THESE fast moving, hurried times we live in bring along different health risks to the people's lives. In particular, the growing number of people suffering from stress, which induces high numbers of severe diseases such as cardiovascular disease, impaired immune system, asthma, peptic ulcer disease, indigestion, headaches, migraines and depression, is one of the typical problems of nowadays. Many people do not realize soon enough that their current stress level is harmful for their health. Therefore, it is necessary to have the stress issue under control and manage it somehow. For this reason, this paper presents the proposed system for monitoring the vital signs of a human body that are related to the stress issues. The system is based on the networking principles of Body Area Network (BAN). The sensor nodes around a human body that communicate in a coordinated fashion create BAN [1]. BAN requirements include low energy specifications and this fact is preferred for various applications including the field of e-Health [1], [2], [3]. Wearable sensor components enable monitoring anywhere, anytime and during wide range of activities (at home, at work, indoors, outdoors, during sports, etc.). Mostly, a comprehensive health image is obtained in comparison with the traditional diagnosing methods [3]. In addition, prompt disease identification typically leads to successful treatment even in case of serious illness [4]. Mostly, a central and most powerful network node is the coordinator of the network [1], [3]. The network coordinator for this application has been chosen to be a smartphone. The reason for this comes from smartphone features - embedded microprocessor, touch screen, capabilities to perform calls (emergency calls), send short text messages and connect to the internet [2].

This work was supported by Foundation Volkswagen Slovakia taking advantage of the Texas Instruments Sample Program.

Various such systems, using the smartphone assets, have been already introduced. In [6], a special body monitoring system for detection of the human body temperature by thermopile sensor, electrical activity of brainwaves by electrocardiogram and electrical activity of the heart by electroencephalogram have been introduced. Smartphone had been in this case used as the communication gateway interfacing the sensors' data with the remote monitoring server, in spite of its management and local-storage functions.

In [7], more complex telemedical system measuring ECG, heart rate, heart rate variability, pulse oximetry, plethysmography and fall detection was presented for the purpose of the patients' physiological parameters outside the clinical environment monitoring and recording. The smartphone was except the common features used for the data visualization and patient's localization matters in order to the emergency communication with a clinical server will be guaranteed on a certain level of quality.

An agent-based approach was presented in [8]. A multi-agent architecture for mobile health monitoring interacting doctor and patient beyond the episodes of visits is presented, involving a team of Java-based intelligent agents that collate patient's data through Bluetooth compliant monitoring device and recommend actions to patients and medical staff in a mobile environment. Agents at the smartphones are also able to monitor the patient's environment through an integrated VGA camera to track patient actions and relay images back to medical staff.

System in [9] is devoted to the monitoring of the athletes during their training process providing them information about the body response to fatigue. Different sensors are utilized within the system, including the photoplethysmographic (PPG) sensor for the heart beat-variability monitoring that is used for the arrhythmias or arterial stenosis and occlusions detection, and the earlobe sensor for the tissue impedance measurement that defining an amount of physical effort due to the ions'

978-83-60810-53-8/$25.00

concentration analysis. Whilst the smartphone was employed as the main control unit for the data collection and analysis, the system provide non-invasive method for the athletes monitoring without interfering their training.

Energy optimization through scheduled communication, Bluetooth parameter tuning and protocol optimization was stressed within the system in [10]. There was developed a Bluetooth-based body sensor network consisted of one smartphone as the network coordinator, multiple sensor nodes for the human body monitoring, and a wristwatch as the user interface. Since the Bluetooth is characterized by large power overhead, its duty cycle was minimised. The developed platform was supplied with a set of APIs for applications on the phone to manage the network, collect data from the sensors, and interact with users via the watch. Data from sensors, which had been most time in a sleep mode, were on account of the measurement request reported to the Internet server or to the phone which could perform its own analysis displayed afterwards through the wristwatch.

In [11], the design and implementation considerations of a smartphone-centred platform for low-cost continuous health monitoring based on commercial-off-the-shelf wireless wearable biosensors were introduced. The platform approach was implemented utilizing PPG biosensors and different smartphones to measure heart rate, breathing rate, oxygen saturation, and estimate obstructive sleep apnea.

The first part of the article describes radio frequency solutions and their comparison in order to better evaluate their power consumption requirements. Smartphone connectivity is investigated in the following part. The next section discusses the mere composition of the network, as well as the principles of communication and timing in the network. Efforts to reduce energy requirements are described in the last part. The proposed sensors are able to capture the temperature, humidity and heart-rate characteristics. Such, values are able to detect if a human is under the stress. Thus, by the analysis of these parameters, the level of human's stress can be found out.

II. Radio Frequency Solutions

An important task of the system design is to select the proper communication mean for the BAN. If one considers a selection of wireless connectivity for a smartphone, two major technologies come to mind. The first is IEEE 802.11.4 (WiFi), which is very powerful, but it is able to drain out the battery in a quite short time. The second option for smartphone wireless connection is IEEE 802.15.1 (Bluetooth). On the subject of the energy consumption, both standards require significant amount of energy and considerably reduce the smartphone's lifetime [12].

Except mentioned technologies, there are also other possibilities which are much more suitable for the BAN purposes, especially in the frame of power consumption. In [7], IEEE 802.15.4 ZigBee platform was implemented within the developed BAN due to the energy-saving reasons. However, though is ZigBee primarily appointed for the low-power, low-cost, multihop networks, it does not exactly meet demands of IEEE 802.15.6 standard for wireless communications supporting ultra-low power devices operating in or around the human body. Therefore, proprietary radio functioning on 2.4GHz ISM was chosen for the BAN purposes. This choice allows definition of the unique low-energy-consuming communication protocol which can be further modified according to the latest release of the BAN standard. Several embedded solutions are available on the market nowadays. The Table I briefly compares the current consumption of the selected transceivers for 2.4 GHz ISM band.

TABLE I
BAN Compliant Transceivers' Current Consumption Comparison in Different Power Modes for Proprietary 2.4GHz ISM Band Radio Modules

Power mode	Quasar RFM 70 [13]	Microchip MRF24J40MA [14]	Nordic NRF24E2 [15]	Microchip MRF89XAM9A [16]
RX mode	17.50 mA	19.00 mA	22.00 mA	3.00 mA
TX mode	14.57 mA	23.00 mA	27.00 mA	25.00 mA
Power down	3.00 μA	2.00 μA	-	1.05 μA
Standby	50.00 μA	-	30.00 μA	-

TABLE II
TI System-on-chip Current Consumption Comparison in Different Power Modes

Power mode	CC2511 [17]	CC2530 [18]	CC2543 [19]	CC2545 [20]
Active mode	3.97 mA	6.27 mA	4.50 mA	4.50 mA
RX mode	19.12 mA	24,80 mA	21.20 mA	20.80 mA
TX mode	21.50 mA	33,93 mA	27.70 mA	28.25 mA
Power mode 0	4.28 mA	-	3.80 mA	3.75 mA
Power mode 1	0.22 mA	0.25 mA	0.24 mA	0.24 mA
Power mode 2	0.75μA	1.50 μA	0.90 μA	0.90 μA
Power mode 3	0.65 μA	0.70 μA	0.40 μA	0.40 μA

Transceiver solution requires a microcontroller (MCU) to process the measured data and to initialize wireless communication. Furthermore, additional space on the PCB (printed circuit board) is required when taking advantage of the standalone transceiver interconnected with MCU via some interface. As the sensors ought to be as small as possible we decided to investigate system-on-chip (SoC) 2.4 GHz RF solutions as well. Whilst Texas Instruments (TI) is probably the best producer of such solutions in the sense of current consumption, selected SoC representatives from TI family were chosen for comparison in Table II.

Common features of these SoCs and transceivers include low-cost 2.4 GHz radio solution, ultra-low power requirements, suitability for portable applications and several advanced low-power operating modes in order to save the power. Since ultra-low power requirements are necessary for proposed application the power requirements parameter was key part of the communication subsystem selection. All values in the Table II are average values of the current consumption based on the datasheet information.

Along with the table, it is possible to conclude that the best power consumption requirements had the first SoC solution TI

CC2511. Therefore, the main MCU that is connected directly to the smartphone via USB, functioning as the communication gateway for the BAN, has been chosen the mentioned CC2511. This chip was, however, unsuitable for the sensor nodes since it lacks the most important communication interfaces - SPI and I^2C [17]. Thus, it was necessary to select another SoC solution that comprises these peripherals. The second best solution provided in the Table II was TI CC2545 that had all desired peripherals and features satisfying the sensors' claims [20].

III. Smartphone USB Connectivity Constraints

Universal Serial Bus (USB) was employed as the main communication interface as well as the power source for the communication gateway board - control node (Fig. 1 and Fig. 2) - since it provides bus-power. That is one of the key advantages, because the device obtains power from the bus and no extra cables are required. In the following subsections, the necessary basics of the USB interconnectivity management will be summarized.

A. USB Device Types

The device specifies its power consumption in 100 mA load units in the configuration descriptor [21]. The device cannot increase its power consumption above the declared amount of load units. Three classes of USB devices exist [22]:

- *Low-power bus powered devices (LPBPDs)* - draw all necessary power from the bus and cannot draw more than one load unit. This class of devices has to be designed to work in 4.40 V up to a maximum 5.25 V voltage range. Therefore, many devices require LDO regulators;
- *High-power bus powered devices (HPBPDs)* - all necessary power is drawn from the bus and cannot draw more than one unit load until it has been configured. After the configuration, the device may drain up to five load units (max. 500 mA) provided in the configuration descriptor. Such devices have to operate in 4.40 V - 5.25 V voltage range. When operating at a full unit load, the minimal voltage level is 4.75 V. Once again, a LDO regulator is needed for many devices;
- *Self-powered devices (SPDs)* - may draw up to one load unit from the bus and the rest of the necessary power may derive from an external source. One load unit allows reliable detection and enumeration of the devices without main/secondary power applied;

No USB device, whether bus powered or self-powered can drive the bus (in sense of the power). If the power is lost, the device has 10 seconds to remove power from the pull-up resistors on the USB data pins that are used for speed identification. Another very important consideration for implementation is the inrush current that has to be limited. The inrush current contributes to the capacitance of the device between the USB power and the USB ground [22]. The maximum decoupling capacitance stated in USB specifications is 10 μF. When the device disconnects, a large voltage peak may occur. Therefore, at least 1 μF decoupling capacitance has to be implemented for safe USB operation [21].

Fig. 1. Central unit - smartphone connection

B. USB Suspend Mode

Support for suspend mode is mandatory for all devices. During this mode, additional constrains apply. The maximum suspend current is proportional to the load unit. For one load unit the maximum suspend current is 500 μA [21]. This includes current from the pull-up resistors on the bus. Another consideration for many devices is the 3.3 V regulator. Regular voltage regulator has average quiescent currents around 600 μA [22]. For one unit load it is necessary to implement more efficient and sophisticated voltage regulators. In most cases, microcontroller clock has to be stopped or slowed down to fall within the 500 μA limit [22].

A USB device will enter suspend when there is no activity on the bus for greater than 3.0 ms. Then, the device has another 7 ms to shut down and draw no more than the designated suspend current [21]. In order to maintain connected, the device has to provide power to its pull up speed selection resistors also during suspend mode.

USB specification determines frame packet start as well as periodical sending of keepalive packets. This prevents an idle bus from entering suspend mode during the data absence. High speed bus has micro-frames sent every 125.0 μs, full speed bus sends frames each 1.0 ms and low speed bus sends the keepalive frames every 1 ms only in the absence of any low speed data [21].

IV. Characteristic of the Proposed BAN

Proposed network for the vital function monitoring senses three characteristics of human body, assuming that the stress can be determined with change of sweating, heart-rate and body temperature. Due to the reasons of energy-saving, temperature and humidity was being monitored each time period on temperature/humidity module (TH module), with the time period set to $1s$. If a significant change from the normal value was spotted, the control node/central unit (CU) sent the information to oximeter module (O module) for the need of heart rate measurement. In case that the stress was indicated also from heart rate, signal is sent to the smartphone and according to the stress rate, the smartphone application will choose form available methods and will offer suitable activity for stress dismantle.

Communication structure of the BAN modules was based on the star network architecture, where the central unit interconnecting whole BAN was set to be CU. More about the architecture of BAN was written in [2], [3]. The interfaces

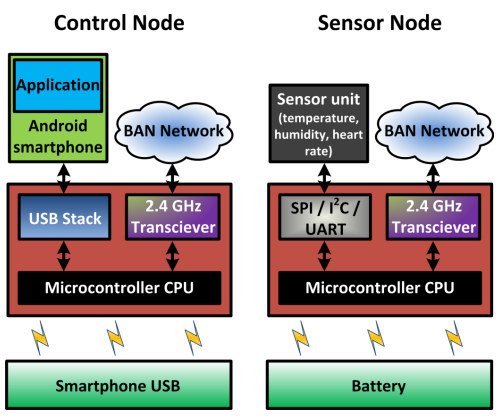


Fig. 2. BAN concept

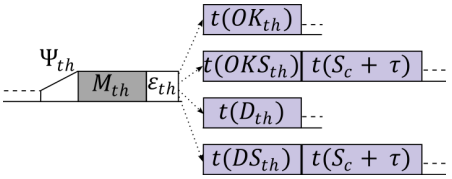


Fig. 3. Time diagram for temperature-humidity module

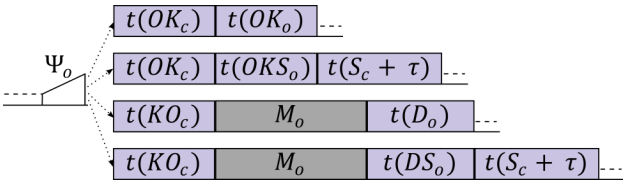


Fig. 4. Time diagram for oximeter module

utilized within the particular nodes are depicted in Fig. 2. The central unit was connected to the smartphone through USB.

For temperature and humidity monitoring, HIH-6131-021-001 sensor is used. The sensor goes into sleep mode when not taking a measurement and consuming only 1 μA of power. In full operation consumes 1 mA and the time of power-on to data ready is 60 ms [23]. Standard way of the pulse oximetry with emitting diode and photodetector is used for the heart-rate measures in a fingertip.

V. Timing and Communication Slots

Aforementioned modules communicate with each other in window slots (TDMA). CU which collects information from sensor modules, is responsible not only for measured data evaluation but also for synchronization of measuring modules.

Because the aim is to decrease the energy requirements, CU is active in communication slots for each module. Then it evaluates data, sends them to smartphone and goes into sleep mode for the next communication period.

In Fig. 3 time diagram for TH module is depicted. The module sleeps and wakes in the time Ψ_{th} [s] before measurement. Ψ_{th} depends on the wake-up-time of the selected core TI CC2545 necessary for getting from power mode 3 to active mode as well as on the assurance constant value. Ψ_{th} was set to $150\,\mu s$. Because the measurement is required before each communication slot, the RX/TX initialization is made at the same time as the measurement is made. Measurement of used temperature/humidity sensor HIH-6131-021-001 is ready after $M_{th} = 60\,ms$. The measurement is sent to core through SPI interface and the data are concerned for sending. Data processing and communication of the core with the sensor last for ε_{th} [s]. If data do not get over the change limit, the data are not sent, only information about good condition of module is sent in packet OK_{th}. Type of communication after data processing is visible in the Fig. 3 with respect to the Table IV. After last communication in corresponding communication period the module returns back to the sleep mode.

Measured data are processed in CU. If the values are in the scope where the stress was not identified, CU sends only OK_c packet to O module in corresponding time slot and the O module does not measure heart rate. Because measuring is energy-demanding, in this way it is assured that energy consumption is suppressed. Time flow of oximeter communication period is depicted in Fig. 4. O module does not measure before each communication slot, but it needs to wake up in the time Ψ_o before the packet arrives from the CU. $\Psi_o = 650\,\mu s$ depends on the ability of module to change from power mode 3 to active mode, on the time necessary for the activation of receiving as well as on the subsumption of the assurance constant.

Because the communication with each module is different, for communication purposes not the same packets are used. Form of the packet is visible in Table III, acceptable types of packet and length of data are mentioned in Table IV.

Because the communication under the communication rate $v_c = 1\,Mb/s$ is aimed to be accomplished within the system, the time for communication of worst case is when the packet size $PS = 7B$ (according to Table III and IV) is sent. The length of communication slot t_s is set as:

$$t_s = \frac{v_c}{PS} + \lambda \approx 60.5\,\mu s, \quad (1)$$

where λ is assurance constant for resynchronization and was set to $0.5\,\mu s$. λ assures, that also when the crystal inaccuracy

TABLE III
BAN Packet Structure

Byte \ Bit	0	1	2	3	4	5	6	7
0	Synchro bits				ID			
1	Packet length				Type of packet			
2-5	Data[a]							
	CRC[b]							

[a] number of data bytes depends on type of packet
[b] one byte after data block, if is

TABLE IV
BAN PACKET TYPES

Packet type	Data length	Description
D_{th}	4 B	Data from temperature/humidity module
D_o	1 B	Data from oximeter module
DS_{th}	4 B	Data with synchronization request from temperature/humidity module
DS_o	1 B	Data with synchronization request from oximeter module
OK_c	0 B[a]	Information packet from central unit for oximeter module that there is no need for measurement
KO_c	0 B[a]	Information packet from central unit for oximeter module that there is need for measurement
OK_o	0 B[a]	Acknowledgement from oximeter module about the packet OK_c from central unit
OKS_o	0 B[a]	Acknowledgement about the packet OK_c from central unit and synchronization request from oximeter module
OK_{th}	0 B[a]	Information packet from temperature and humidity module for central unit that everything is all right and the measurement does not change after treshold
OKS_{th}	0 B[a]	Information packet from temperature and humidity module for central unit that everything is all right, measurement does not change after treshold and there is need for synchronization
S_c	4 B	Synchronizing time from central unit

[a]Information packet differs only in packet header. There is no need to transfer data.

will occur, the communication in corresponding time slot will be possible.

Because the communication is performed between the nodes with two different cores CC2511 and CC2545, they oscillators have different crystal frequency (f_{CC2511}, f_{CC2545}) and accuracy (acc_{CC2511}, acc_{CC2545}) . Time of communication period is set to 1s, error of crystal for core CC2511 Δ_{2511} is set to:

$$\Delta_{2511} = \frac{1}{f_{CC2511}} - \frac{1}{f_{CC2511} + acc_{CC2511}} \approx 1.538x10^{-12}\,s \quad (2)$$

Error of core CC2545 Δ_{2545} is set to:

$$\Delta_{2545} = \frac{1}{f_{CC2545}} - \frac{1}{f_{CC2545} + acc_{CC2545}} \approx 1.875x10^{-12}\,s \quad (3)$$

set for $60.5\,\mu s$ and the number of synchronized periods p_s is set as:

$$p_s = \frac{\frac{\lambda}{2}}{\Delta_{2511} + \Delta_{2545}} = \frac{0.25x10^{-6}\,s}{3,413x10^{-12}\,s} \approx 73\,249 \quad (4)$$

It means that the synchronization is necessary after 73 249 cycles. As was mentioned, cycle for TH measure is long 1 second, so the synchronization is used 2 times a day.

When each of sensing modules need to synchronize, it sends the synchronization packet to CU. The type of synchronization packets are mentioned in Table IV.

VI. Energy saving in TH module

Low power consumption is one of major condition in many applications. Amount of required energy is very important in monitoring applications where an effort to achieve long operation time is emphasized. It is not only the problem of applications from BAN area, but also from many other monitoring applications in the field of wireless sensor network [24].

One of the possibility is to use sleep mode as was mentioned in previous part. For TH module, communication period lasts $1s$ and is visible in Fig. 3. TH module is active for time t_a:

$$t_a = \Psi_{th} + M_{th} + \varepsilon_{th} + t_s \approx 0.066\,s \quad (5)$$

and sleeps for $t_{sleep} = 0.934\,s$. Concerning the consumptions listed in Table II, average current I_{avg} for TH node was estimated as $2\,mA$. According to used battery CR2450 which capacity is $600\,mAh$, the TH module should operate more than 12 days.

VII. Conclusion

This paper presents the smartphone-based Body Area Network for monitoring of human vital signs regarding the stress situation. In particular, energy requirements are discussed. The results showed that the designed BAN is able to last up to 2 weeks without any intervention. On the other hand the smartphone in not capable of such long lifetime. Therefore, until now, the weakest link of the network is the smartphone and the coordinator module attached to it. This application is eligible for eHealth sector as well as general wellness to serve as front-line in cardiovascular disease detection. The future work is focused on implementation of the developed system into different application fields of modern life, primarily into the field of intelligent transportation systems aiming the problematic of the professional drivers' behaviour on the road with the comparison to their actual stress level.

Acknowledgment

This publication is the result of the project implementation *"Vývoj špeciálnej BAN pre monitoring pacienta v domácich*

podmienkach za využitia smartfónu ako riadiacej jednotky i koordinátora siete" with grant no. 050/12 in grant program *"Vysokoškoláci a technika"* supported by Foundation Volkswagen Slovakia.

This publication is co-funded by the project implementation *Centre of excellence for systems and services of intelligent transport II.* ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

"Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ"

REFERENCES

[1] J. Miček, O. Karpiš and P. Ševčík, "Body area network: analysis and application areas," *in International journal of engineering research and development,* vol. 6, ISSN 2278-800X, 2013, pp. 22–26.

[2] M. Kochláň, M. Hodoň and J. Púchyová , "Vital functions monitoring via Sensor Body Area Network with smartphone network coordinator," *in MEMSTECH 2013 : perspective technologies and methods in MEMS design : proceedings of the IXth international conference,* 16-20 April 2013, Polyana, Ukraine, Lviv: Lviv Polytechnic Publishing House, 2013, ISBN 978-617-607-424-3, pp. 143–147.

[3] M. Kochláň, M. Hodoň and J. Púchyová , "Body Area Network for Monitoring Human Vital Signs Using Smartphone," *in TRANSCOM 2013 : 10-th European conference of young researchers and scientists,* Žilina, June 24-26, 2013, Slovakia, Žilina: University of Žilina, 2013, pp. 41–44.

[4] E. Montón, J.F. Hernandez, J.M. Blasco, T. Herve, J. Micallef, I. Grech et al., "Body area network for wireless patient monitoring," *Telemedicine and E-health Communication Systems*, Vol. 2, February 2008, pp. 215-222.

[5] EUROSTAT, *Eurostat regional yearbook 2012*, Chapter Health, ISBN 978-92-79-24940-2, 2012.

[6] W. Song, H. Yu, C. Liang, Q. Wang and Yunfeng Shi, "Body Monitoring System Design Based on Android Smartphone", *IEEE 2012 World Congress on Information and Communication Technologies*, October 30 - November 2 2012, Trivandrum, India, Print ISBN: 978-1-4673-4806-5, pp. 1147–1151.

[7] M. Wagner, B. Kuch, C. Cabrera, P. Enoksson, A. Sieber, "Android Based Body Area Network for the Evaluation of Medical Parameters", *IEEE 10th International Workshop on Intelligent Solutions in Embedded Systems*, 2012, 5-6 July 2012, Klagenfurt, Austria, Print ISBN: 978-1-4673-2464-9, pp. 33–38.

[8] V. Chan, P. Ray and N. Parameswaran, "Mobile e-Health monitoring: an agent-based approach", *IET Communications*, Volume 2, issue 2, February 2008, pp. 223–230.

[9] A. Depari, A. Flammini, S. Rinaldi, A. Vezzoli, "Multi-sensor system with Bluetooth connectivity for non-invasive measurements of human body physical parameters", *Sensors and Actuators A: Physical (2013)*, in press, http://dx.doi.org/10.1016/j.sna.2013.05.001.

[10] L. Zhong, M. Sinclair, R. Bittner, "A Phone-Centered Body Sensor Network Platform: Cost, Energy Efficiency & User Interface", *IEEE International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006)*, 3-5 April 2006, Cambridge, Massachusetts, USA, ISBN 0-7695-2547-4.

[11] S.C.K. Lam; K. L. Wong, K. O. Wong, W. Wong, "A smartphone-centric platform for personal health monitoring using wireless wearable biosensors", *7th IEEE International Conference on Information,Communications and Signal Processing*, 7-10 December 2009, Macau Fisherman's Wharf, Macau, ISBN: 978-1-4244-4656-8, pp. 1–7.

[12] P. Smith, *Comparisons between Low Power Wireless Technologies*, CSR plc whitepaper CS-213199-AN.

[13] Quasar Datasheets, "Low Power High Performance 2.4 GHz GFSK Transceiver Module RFM70 v1.0," http://www.quasaruk.co.uk/acatalog/RFM70.pdf.

[14] Microchip Datasheets, "2.4 GHz IEEE Std. 802.15.4 RF Transceiver Module - MRF24J40MA," http://ww1.microchip.com/downloads/en/DeviceDoc/70329b.pdf.

[15] Nordic Semiconductor Product Specifications, "2.4GHz RF transmitter with embedded 8051 compatible microcontroller and 9 input, 10 bit ADC - nRF24E2," http://www.nordicsemi.com/eng/content/download/2742/34227/file/Product_Specification_nRF24E2_1_3.pdf.

[16] Microchip Datasheets, "915 MHz Ultra Low-Power Sub-GHz Transceiver Module - MRF89XAM9A," http://ww1.microchip.com/downloads/en/DeviceDoc/75017B.pdf.

[17] Texas Instruments Datasheets, "Low-Power SoC (System-on-Chip) with MCU, Memory,2.4 GHz RF Transceiver, and USB Controller - CC2510Fx/CC2511Fx," http://www.ti.com/lit/ds/symlink/cc2511f32.pdf.

[18] Texas Instruments Datasheets, "A True System-on-Chip Solution for 2.4-GHz IEEE 802.15.4 and ZigBee Applications - CC2530F32, CC2530F64, CC2530F128, CC2530F256," http://www.ti.com/lit/ds/symlink/cc2530.pdf.

[19] Texas Instruments Datasheets, "System-on-Chip for 2.4-GHz RF Applications - CC2543," http://www.ti.com/lit/ds/symlink/cc2543.pdf.

[20] Texas Instruments Datasheets, "System-on-Chip for 2.4-GHz RF Applications - CC2545," http://www.ti.com/lit/ds/symlink/cc2545.pdf.

[21] USB.org, "Universal Serial Bus Revision 2.0 specification," http://www.usb.org/developers/docs/usb_20_040413.zip.

[22] C. Peacock, "USB in a NutShell," http://www.beyondlogic.org/usbnutshell/usb2.shtml

[23] Honeywell Datasheets, "Honeywell HumidIcon™ Digital Humidity/Temperature Sensors: HIH-6130/6131 Series," http://www.farnell.com/datasheets/1443945.pdf.

[24] J. Miček and J. Kapitulík, "WSN sensor node for protected area monitoring," *in Federated Conference on Computer Science and Information Systems (FedCSIS),* Wroclaw, ISBN 978-83-60810-51-4, 2012, pp. 803–807.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 901–906

SENTIOF: An FPGA Based High-Performance and Low-Power Wireless Embedded Platform

Khurram Shahzad, Peng Cheng, Bengt Oelmann
Department of Electronics Design
Mid Sweden University
Sundsvall, Sweden
{khurram.shahzad, cheng.peng, bengt.oelmann}@miun.se

***Abstract*—Traditional wireless sensor nodes are designed with low-power modules that offer limited computational performance and communication bandwidth and therefore, are generally applicable to low-sample rate intermittent monitoring applications. Nevertheless, high-sample rate monitoring applications can be realized by designing sensor nodes that can perform high-throughput in-sensor processing, while maintaining low-power characteristics. In this paper, a high-performance and low-power wireless hardware platform is presented. With its compact size and modular structure enabling there to be an integrated customized sensor layer, it can be used for a wide variety of applications. In addition, the flexibility provided through dynamically configurable interfaces and power management, helps optimizing performance and power consumption for different applications.**

***Keywords*—FPGA, wireless platform, high-throughput, low-power**

I. Introduction

With the advent of low-cost, low-power, and miniature size electronics, wireless sensor nodes have emerged as a low-cost alternative to fixed wired-based sensing solutions. These nodes are typically designed using a low-power micro-controller, a radio transceiver, and one or more sensors to measure a particular physical phenomenon. Based on the limited computational performance, communication bandwidth, and lifetime, these nodes have traditionally been applied to particular low-sample rate intermittent monitoring applications in different field [1]-[4]. However, their potential advantages have provided the motivation to explore them in relation to high-sample rate applications.

Given the limited bandwidth and high-power consumption of radio transceivers that are typically used in these nodes, the major challenge lies in transmitting a large amount of data generated by high-sample rate applications. In order to overcome this challenge, researchers have proposed in-sensor processing [5]-[8] in which, raw data is processed locally in a sensor node, and only results are transmitted wirelessly. Therefore, it reduces communication activity and the associated power consumption. However, it is observed that typical low-power micro-controllers integrated in sensor nodes, often lack high-throughput performance when complex mathematical and signal processing algorithms are involved. Therefore, to meet the demands of high computational performance, the prototype design reported in [5] employs a high-throughput micro-controller for in-sensor data processing, in addition to a low-power 8-bit micro-controller. In [6], a Field Programmable Gate Array (FPGA) based in-sensor processing solution is demonstrated in relation to an image processing application. In [7] and [8], FPGA-based commercial evaluation kits and micro-controller based wireless nodes were used to investigate their performance and energy consumption for high-sample rate applications. Based on the presented results in these articles, it can be concluded that high-throughput processing requirements can only be fulfilled through an FPGA, however, by accurately distributing the processing, communication and control specific tasks on a micro-controller and an FPGA, it is possible to optimize both the performance and power consumption for an application.

By analyzing the published literature relating to high-performance wireless nodes, it was found that there are only two such nodes, [9] and [10] that are able be used for high-throughput in-sensor processing. Both nodes are very similar in their design, as they are built on a similar layered structure and they integrate similar modules. For example, each has a sensor layer, a processing layer consisting of micro-controller and/or an FPGA, a communication layer to enable wireless communication and, in addition, to a power supply layer. On both nodes, the processing layer consists of an FPGA that can be used for in-sensor processing. However, both these FPGAs (Spartan-IIE, and Spartan-IIIE) were built on a technology that is now a decade old and therefore, does not offer the high-performance and low-power consumption that is the case with the modern FPGAs. Apart from that, inter module communication interfaces such as between FPGA and micro-controller, and between wireless transceiver and micro-controller are fixed in these nodes and therefore, cannot be re-configured for different applications without modifying the hardware design. Additionally, the compactness of these nodes is compromised because of the extra height, resulting from the interface connectors on both sides of a layer and also stacking of four layers on top of each other to form a complete working node.

To overcome the above mentioned problems, we have developed a high-performance, low-power and compact wireless embedded platform, the SENTIOF that is shown in Fig. 1. It integrates a micro-controller, an FPGA, an SRAM, a FLASH, and a radio transceiver on a single printed circuit board (PCB), while providing easy integration of a customized sensing layer,

978-83-60810-53-8/$25.00

in order to use it in different applications. The SENTIOF is designed with the following key characteristics.

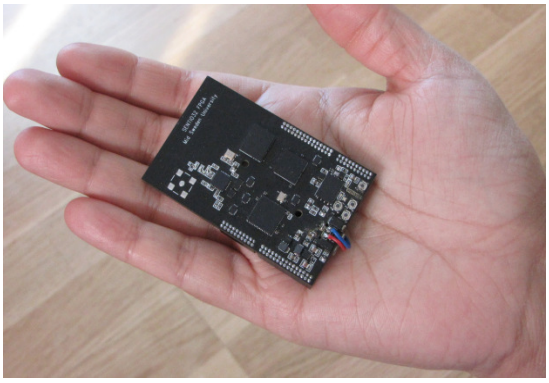


Figure 1. The SENTIOF platform

High-Performance In addition to sequential processing capabilities, there must be support for hardware acceleration in order to perform computationally intensive processing.

Low-power The integrated components should be operable at different clock frequencies so as to optimize the power consumption and performance. In addition, the platform must support a dynamic power management to switch-off unused components.

Sensor integration The platform must support easy integration of application dependent sensors. Therefore, in addition to different power supply voltages, a large number of input-output interfaces from a micro-controller and FPGA should be provided for this purpose.

Flexibility In addition to the easy integration of sensors either with a micro-controller, FPGA or both, the platform must support the easy realization of application dependent communication between the micro-controller and the FPGA.

Compact size The platform should have small dimensions so that it can be deployed in applications with space constraints.

The details regarding the design, performance and power consumption of the SENTIOF are the main focus of this paper. The remaining sections are organized as follows. Section II describes the hardware design aspects of the SETNIOF. Section III is devoted to software related details. In section IV, the performance and the power consumption analysis are presented, while concluding remarks are given section V.

II. Hardware Design

A simplified design of the SENTIOF, representing all major components and their interconnections is shown in Fig. 2. With the exception of power source, all other components are integrated in the SENTIOF platform, which is discussed in detail in the following.

A. Plateform Components

Based on the functional description of the integrated components, the SENTIOF can be divided into different logical units, such as processing, communication, memory, power management and sensor interface, and can be then described accordingly.

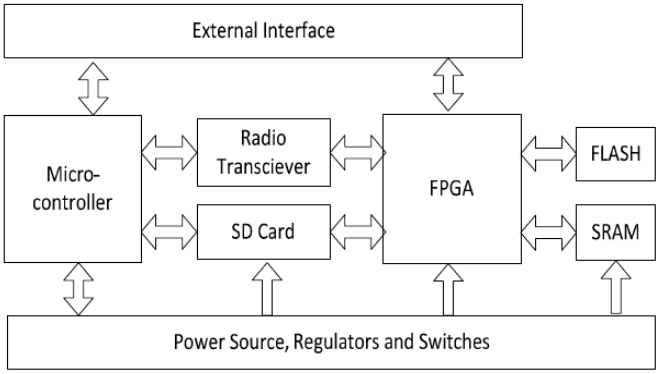


Figure 2. Architectural depiction of the SENTIOF

Processing Unit

In order to achieve a high-computational performance, two processing resources, a 32-bit micro-controller ATUC3B0512 [11], and a Spartan-6 FPGA XC6SLX16-2CPG196 [12], are integrated in the SENTIOF.

The micro-controller provides high computational capabilities by means of a compact single-cycle instruction set including DSP instructions and an operational clock frequency of up to 60 MHz. Additionally, it includes an extensive set of input-output (IO) interfaces and a number of low-power modes in order to optimize the power consumption for an application. The micro-controller can be used to process the application data, read the sensor's data, and/or communicate with the radio transceiver in order to transmit or receive data/results. In addition to its application dependent data processing, the micro-controller performs all the control specific operations such as power and FPGA configuration management and therefore, acts as a central control unit in the SENTIOF.

The micro-controller can be used as the main processing unit for a wide variety of monitoring applications. However, in order to achieve a high-throughput computational performance as required in many high-sample rate applications, the FPGA is integrated into the SENTIOF. Due to hardware parallelism, it is possible for multiple independent tasks to be realized using dedicated hardware logic in the FPGA, which thus be performed simultaneously at high speed. In addition, the re-configurability of the FPGA allows synthesizing different algorithms with optimized control and data paths, without modifying any real hardware. The selected FPGA is enriched with sufficient amount of basic logic cells, in addition to other resources such as block RAM, multiply-and-accumulator (MAC) units, and PLLs. A summary of these resources is given in Table I.

Radio transceiver

In order to enable wireless communication in the SENTIOF, an IEEE 802.15.4 compliant low-power radio transceiver, CC2520 [13] is integrated in the design. The

transceiver operates at the 2.4 GHz license free band and provides a maximum throughput of 250 kbps. In addition to 6 reconfigurable GPIOs for optional command and interrupt signals, the radio transceiver includes a SPI interface to communicate and exchange data either with the micro-controller or with the FPGA. The interface to both the micro-controller and the FPGA is achieved by means of multiplexer/de-multiplexer switches and is therefore, able to be configured dynamically. This additional flexibility can be exploited to optimize the performance and the power-consumption of the communication activity according to the requirements of an application.

TABLE I.
A SUMMARY OF LOGIC RESOURCES OF THE SPARTAN-6 FPGA XC6SLX16-2CPG196

Logic Cells	LUT size	Dis-tribu-ted RAM	Block RAM	MAC units	PLL	Number of IOs
14,579	6-input	136 kb	576 kb	32	2	106

Memory

High-sampling rate applications generally process a large amount of raw data, and generate both intermediate and finals results that must be stored in fast memories in order to achieve a high performance. Therefore, in addition to a 96 kB of SRAM in the micro-controller and 89 kB in the FPGA, a 4 MB of additional SRAM, CY62177DV30 [14] is integrated in the SENTIOF. The low-power SRAM, interfaced with the FPGA, provides a 16-bit wide data path for read and write operations that can be performed in 55 ns.

The FPGA requires re-configuration each time it is powered-on. Therefore, a high-speed and low-power FLASH based configuration memory, W25Q64BV [15][14] is used for this purpose. The selected FLASH memory provides a rapid access rate of 85 MHz and a quad serial peripheral interface (SPI), which enables the configuration of the FPGA in the minimum time. In addition to the configuration data of 4 Mb, the 64 Mb FLASH can be used to store application data.

For long term data storage, the SENTIOF is designed to support a micro-SD card. In a similar manner to that for the radio transceiver, the interface to the SD card can also be dynamically configured either to the micro-controller or to the FPGA, as it is interfaced to both these processing units through multiplexer/de-multiplexer switches.

Power

The SENTIOF is powered by means of a single DC power source, with output voltage between 3.6V and 6V. This DC voltage is fed to the SENTIOF and is then regulated and converted to four different levels as shown in Fig. 3. In relation to these four voltages, the 1.2V regulated voltage is used to power the FPGA's core while the 1.8V is used to power the radio transceiver and the core of the micro-controller. Apart from the radio transceiver, the cores of the FPGA and the micro-controller, all other components on the SENTIOF platform are powered with 3.3V. The boost convertor that can be adjusted to provide up to 6.5V, is included for external devices/sensors that may require a higher voltage than is provided by attached DC power source, i.e. typically a 3.6V. In addition to a DC power source, the SENTIOF can also be powered through a USB interface. This provides an added advantage of continuous long term power during an application development process.

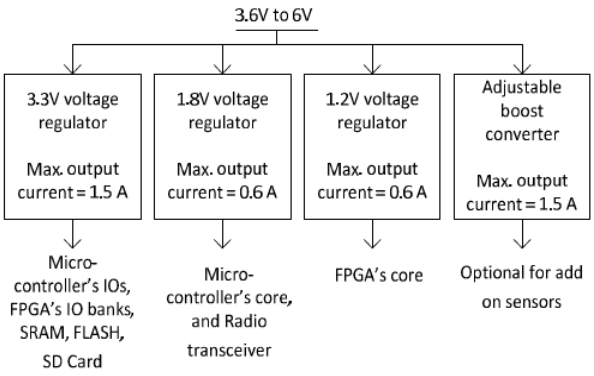


Figure 3. Power supply distribution in the SENTIOF

External Interface

The external interface provided through the 1.27 mm-pitch connectors serves three purposes. Firstly, it provides a large set of IOs, both from the micro-controller and the FPGA, in addition to all regulated power sources in order to integrate a customized sensor board with the SENTIOF. Secondly, the JTAG interfaces required to program/configure each of the micro-controller, the FPGA, and the FLASH memory from the computer are also made accessible through this external interface. Lastly, the application specific communication requirements between the micro-controller and the FPGA can also be fulfilled through this interface.

B. PCB Design

In order to ensure a high-performance and low-production cost, special consideration was given to the PCB design. This includes the separation of different ground planes, the minimization of trace length between the high-speed devices, limiting the design to a reduced number of layers, and avoiding micro and buried vias.

To minimize noise in relation to all the different types of components, four ground planes are used. A generic ground plane, GND, is used to connect ground plane of the DC power source to other planes through 0 Ohm resistors, as shown in Fig. 4. In these planes, the digital ground, DGND serves as a return path for all digital components including the FPGA, micro-controller, SRAM etc. The power ground, PGND is used to minimize the noise from other planes onto the power supply components. As there are no analog components mounted on the SENTIOF, the analog ground AGND is not used. However, it is provided to ensure that analog sensors can be reliably interfaced with the SENTIOF.

In relation to all the components mounted on the SENTIOF, it is the FPGA that has the highest pin count of 196 pins, which

are packed into 8x8 mm BGA package organized in 14 x 14 rows and columns. The 0.5 mm horizontal and vertical pin pitch resulting from this small footprint was a challenging factor, as it determined the routing and clearance rules in addition to the number of the routing layers. For example, if the manufacturer's guidelines regarding the PCB design of the FPGA [16] are strictly followed, then it requires a minimum of 7 PCB layers using micro-vias, while restricting the trace width and clearance to 0.075 mm. The production cost of the resulting PCB is then significantly higher in comparison to an equivalent PCB with a via diameter of more than 0.15 mm, and a trace width and clearance of 0.1 mm or more.

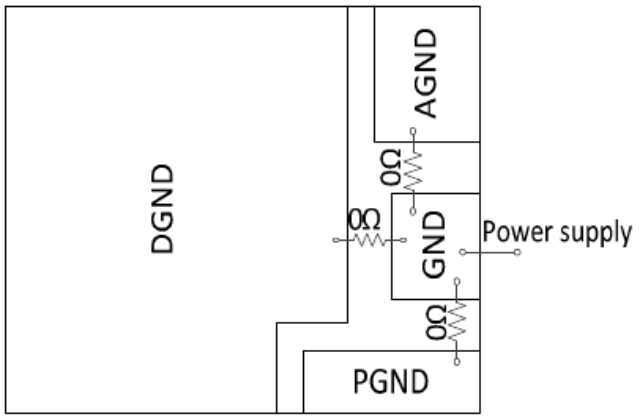


Figure 4. Ground planes used in the SENTIOF.

Therefore, to relax the production requirements, the via-in-pad option was used for the FPGA. This allowed to use larger pad size and, eventually, larger via hole diameter of 0.2 mm, a minimum trace width and clearance of 0.1 mm. In addition, it also helped to reduce the PCB design to 6 layers and without the necessity of buried vias.

Among other components, it was the radio transceiver that required the strictest observation of the manufacturer's guidelines, in order to ensure proper functionality. Therefore, the layout of the radio transceiver was completely matched with the reference layout. In addition to providing an interface for an external antenna, a PCB antenna is also incorporated for the radio transceiver.

C. Physical Structure

The size of the PCB is 65 x 40 mm, with 5 mm of interface height on one side that is used to attach a sensor module. A sensor module comprising of one or more sensors, is electrically connected with the SENTIOF platform through a rigid and strong header interface. In addition, two mounting holes are created in order to provide structural reinforcement of the platform and the sensor module, which may be required for certain applications.

III. Software Design

In order to use the SENTIOF for an application, supporting software must be developed for both the micro-controller and the FPGA. Currently, this is achieved using two different software development environments, the AVR32 Studio and the Xilinx ISE for the micro-controller and FPGA, respectively.

The integrated development environment for the micro-controller that can be obtained from Atmel's website free of charge and it integrates a software framework and a GNU tool chain allowing easy and rapid application development in C/C++. To further enhance the software development process of the micro-controller in relation to the SENTIOF, application programming interfaces (APIs) are developed for all control and data transfer specific operations including power management, radio communication and reading/writing to SD card.

The Xilinx ISE Design Suite integrates all the tools to support complete design development, starting from an RTL design specification to the generation of a programming file for the FPGA. It also includes a wide variety of IP cores that can be integrated into a design, thus resulting in rapid development. In relation to the SENTIOF, we have developed interface APIs for both the SRAM and the FLASH, which can be re-used in other designs. Similar APIs for the radio transceiver and the SD card interface will also be developed in the near future.

IV. Performance and Power Consumption Analysis

The SENTIOF was designed to achieve a high computational performance and low power consumption goals, and these were ensured at each development stage including architectural decisions, component selection, and PCB layout. Performance of both the micro-controller and the FPGA is dependent on an underlying application. Therefore, a measure of clock frequency, on which these two can be operated, is discussed for analysis purposes. However, the precisely measured power consumption for various activities is used for analysis in the this section.

D. Clock frequency

The micro-controller can either be clocked from an internal oscillator producing a clock frequency of about 115 kHz or from an external oscillator producing 16 MHz of clock frequency. The clock from the internal oscillator is often too slow for a computationally intensive application and therefore, may be used in either very low-speed applications or during sleep modes. However, to achieve different performance levels, the micro-controller can be operated at a wide frequency range, with an upper limit of 60 MHz. In such cases, the desired frequency can be synthesized from a 16 MHz external oscillator by using a built-in Phase Locked Loop (PLL) in the micro-controller.

During the SENTIOF design it was observed that the available oscillators with frequencies higher than 19 MHz, as is required in the FPGA, consume a significant amount of power that is undesirable in a low-power platform. Therefore, a global clock generation feature of the micro-controller was instead used to provide the clock for the FPGA. The frequency of this input clock can be further increased to 375 MHz using a PLL in the FPGA.

E. FPGA Configuration

In order to conserve power, it may be desirable to switch off the FPGA for idle time periods in an application. However, upon power-up, it must be reconfigured to resume its job. The (re)configuration for the FPGA integrated in the SENTIOF is accomplished by loading nearly 3.6 Mb of configuration data into the FPGA, from the associated FLASH memory. Depending upon the clock frequency and the bus width options that are selected to configure the FPGA, the actual configuration time can vary from one application to another.

The minimum configuration time of 15.16 ms was recorded by applying a maximum supported speed of 66 MHz and a maximum bus width option of 4 bits, also known as quad SPI.

F. Power Consumption

The power consumption in the SENTIOF can be optimized through dynamic power management, where all major components including the micro-controller, FPGA, SRAM, FLASH, and the radio transceiver can be switched to low-power modes at run time. This allows a reduction in the power consumption of each component to a minimum level, typically from tens of micro-watts to a few milli-watts depending upon the actual component. Therefore, to further minimize the power consumption during the idle state, the SENTIOF is designed to dynamically switch-off/on the FPGA, SRAM, FLASH, and SD card as they consume significant power in their low-power modes. To realize this power on/off mechanism, a very low-power metal-oxide semiconductor (MOS) transistor is used as a switch between the power supply and the power connection of the components. The transistor is then switched on/off through the micro-controller, which performs all control specific operations including power management. It should be noted that radio transceiver that typically consumes 1μW in low-power mode and, it is not enabled to be switched on/off using MOS transceiver, as the MOS transistor consumes almost the same amount of power as that of the radio transceiver in low-power mode and therefore, no significant power can be conserved by the power on/off method.

Depending on the application, the SENTIOF may be in active mode, where it performs data acquisition, processing, and/or result transmission, or it may be in sleep mode for certain time duration before it is operational again. The power consumption in these modes differs significantly and therefore, it is discussed with respect to these two modes in the following.

Sleep mode

In sleep state, with the exception of the micro-controller, other components including the FPGA, SRAM, FLASH, and the SD card are switched-off. The micro-controller however, is switched to a low-power mode known as DeepStop, in which it is not only able to keep track of the sleep duration by using real time counter (RTC), but is also capable of switching all components including itself to the active state.

In sleep mode, the average current drawn by the SENTIOF from a 3.6V supply source was measured to be 95 μA. During this mode, both the 1.8V and 3.3V voltage regulators which remained fully functional to ensure the required voltage levels to the micro-controller were responsible for nearly 70% of the reported current consumption.

Active mode

Unlike the sleep mode, the power consumption of the SENTIOF during its active mode is dependent on an frequencies and the time duration for which different components are active. Nevertheless, to provide a rough idea, the current consumption was measured for a number of application scenarios involving almost all the major components on the SENTIOF, and is summarized in Table II. Each reported value represents the total amount of current that was drawn by the SETNIOF for a corresponding scenario, from a 3.6V power source.

For application scenarios 1, 2, and 3 given in Table II, the micro-controller was in the active state performing a simple addition operation repeatedly. Apart from the 3.3V and 1.8V voltage regulators that are required to provide power to the mi-

TABLE II. THE AVERAGE CURRENT CONSUMPTION OF THE SENTIOF FOR DIFFERENT APPLICATION SCENARIOS

S. No.	Actions	Average Current (mA)
1.	The micro-controller is clocked at 16 MHz and is in active mode, where it performs an addition operation repeatedly.	5.54
2.	Same as of 1, except the clock frequency is 20 MHz.	7.49
3.	Same as of 1, except the clock frequency is 60 MHz.	21.34
4.	The micro-controller is clocked at 16 MHz and is in Frozen mode, in which it generates 20 MHz of clock frequency for the FPGA.	4.28
5.	Same as of 4 + both the FPGA and the Flash are ON, and the FPGA loads bit-stream from the FLASH at 66 MHz using quad SPI interface.	30.3
6.	Same as of 4 + FPGA is active and is running a design in which, a 100 MHz of clock is synthesized through internal PLL, and then is used to update a 27-bit counter.	22.31
7.	Same as of 6, except that the FPGA is in standby mode.	8.30
8.	Same as of 6 + the FPGA reads 16-bit data word from the SRAM repeatedly at a rate of 16.6 MHz.	34.58
9.	Same as of 6 + the FPGA writes 16-bit data word to the SRAM repeatedly at a rate of 16.6 MHz.	44.62
10.	Same as of 6 + the FPGA erases the FLASH memory.	48.17
11.	Same as of 6 + the FPGA reads data from the FLASH memory at a rate of 50 MHz.	29.90
12.	Same as of 6 + the FPGA writes data to the FLASH memory.	38.89
13.	Same as of 2 + the micro-controller reads data from the SD card at a rate of 20 MHz.	20.64
14.	Same as of 2 + the micro-controller writes data to the SD card at a rate of 20 MHz.	24.37
15.	Same as of 1 + the radio transceiver operates in receive mode and received packets are transferred to the micro-controller.	31.34
16.	Same as of 1 + the data packets from the micro-controller are transmitted by the radio transceiver at rate of 250 kpbs using 5 dBm power.	39.15

cro-controller, all other modules were in the off state. The micro-controller was clocked through a 16 MHz external oscillator, and then synthesized to 20 MHz and 60 MHz through an internal PLL.

The average current drawn by the SENTIOF while configuring the FPGA from the associated FLASH is given in Scenario 5. Scenarios 8 and 9 list the current consumption when the FPGA performs read and write operations on the SRAM. Other important scenarios include 15 and 16, in which the radio transceiver is operated in receive and transmit modes, respectively.

V. Conclusion

In this paper, we presented a high-performance and low-power wireless hardware platform, SENTIOF that is capable of performing high-throughput in-sensor processing, as typically required for high-sample rate monitoring applications.

The SENTIOF integrates a micro-controller, an FPGA, an SRAM, a FLASH, and a radio transceiver on a single PCB in order to realize a high-performance compact wireless platform. This application independent platform allows for the integration of any kind of sensors through an application specific and customized sensor layer. In addition, the dynamic power management and reconfigurable architecture can be exploited to optimize performance and power consumption according to different applications.

The flexibility provided through dynamically configurable interfaces, customizable communication between the micro-controller and the FPGA, and dynamic power management can also be used to explore efficient architectures for different monitoring applications.

References

[1] R. G. Luis, L. Lunadei, P. Barreiro, and I. Robla, "A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends". IEEE Journal on Sensors, vol. 9, pp. 4728-4750,2009.

[2] A. Sharma, R. Chaki, and U. Bhattacharya, "Applications of wireless sensor network in Intelligent Traffic System: A review", 3rd International Conference on Electronics Computer Technology (ICECT), vol.5, pp.53-57, 8-10 April 2011.

[3] S. H. Lee, S. Lee, H. Song, and H.S. Lee, "Wireless sensor network design for tactical military applications: Remote large-scale environments", IEEE Conferenece on Military Communications MILCOM 2009, pp.1-7, 18-21 Oct. 2009.

[4] A. Pantelopoulos, and N. G. Bourbakis, "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.40, no.1, pp.1-12, Jan. 2010.

[5] J. P. Lynch, "An overview of wireless structural health monitoring for civil structures", Philosophical Transactions of the royal Sociaety A-Mathematical Physical and Engineering Sciences, vol. 365, pp. 345-372, Feb. 15, 2007.

[6] C. H. Zhiyong, L.Y. Pan, Z. Zhenxing, and M.Q.H Meng, "A novel FPGA-based wireless vision sensor node", IEEE International Conference on Automation and Logistics, 2009, pp.841-846, 5-7 Aug. 2009.

[7] K. Khursheed, M. Imran, A.W. Malik, M. O'Nils, N. Lawal, and B. Thörnberg, "Exploration of Tasks Partitioning between Hardware Software and Locality for a Wireless Camera Based Vision Sensor Node", 26th International Symposium on Parallel Computing in Electrical Engineering (PARELEC), pp.127-132, 3-7 April 2011.

[8] K. Shahzad, P. Cheng,B. and Oelmann, "Architecture exploration for a high-performance and low-power wireless vibration analyzer", IEEE Journal on Sensors, vol.13, no.2, pp.670-682, Feb. 2013.

[9] J. Portilla, T. Riesgo, and A. de Castro, "A Reconfigurable FPGA-Based Architecture for Modular Nodes in Wireless Sensor Networks", 3rd Southern Conference on Programmable Logic, pp.203-206, 28-26 Feb. 2007.

[10] J. S. Bellis, K. Delaney, B. O'Flynn, J. Barton, K. M. Razeeb, and C. O'Mathuna, "Development of field programmable modular wireless sensor network nodes for ambient systems", Elseiver Journal on Computer Communications, vol. 28, issue 13, pp. 1531-1544, Aug. 2005.

[11] AT32UC3B0512, San Jose, CA: Atmel, 2013. [Online]. Available: www.atmel.com.

[12] Spartan-6 XCLS6X16-2CPG196 and XC6SLX45T FGG484-3C, San Jose, CA: Xilinx Inc. 2013. [Online]. Available:www.xilinx.com.

[13] CC2520, Dallas, TX: Texas Instruments, 2013. [Online]. Available:www.ti.com.

[14] CY62177DV30, San Jose, CA, USA: Cypress Semicondcutor, 2013. [Online]. Available:www.cypress.com.

[15] W25Q64BV, Taichung City 428, Taiwan: Winbond Electronics, 2013. [Online]. Available: www.winbond.com.

[16] "Spartan-6 FPGA PCB Design and Pin Planning", San Jose, CA: Xilinx Inc. 2013. [Online]. Available:www.xilinx.com.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 907–910

Wireless Indoor Positioning System for the Visually Impaired

Piotr Wawrzyniak, Piotr Korbel
Institute of Electronics
Lodz University of Technology
ul. Wólczańska 211/215, 90-924 Łódź, Poland
Email: piotr.wawrzyniak@dokt.p.lodz.pl

***Abstract*—The paper presents a prototype radio network aiding the visually impaired to navigate in indoor areas. The main purpose of the system is to provide accurate and reliable location information as well as to enable access to location related context information. The nodes of the network operate in two modes providing basis for both rough and precise user position estimation. The data transmitted by the nodes are used to get access to additional services, e.g. to retrieve position related context information.**

***Index Terms*—Context-aware services, indoor radio communication, location services, personal communication networks, pervasive computing, radio navigation, wireless sensor networks**

I. Introduction

THERE ARE about 285 million visually impaired people living around the world [1]. The inability to sense the surrounding environment strongly affects the possibility of utilizing public spaces, including urban areas, transportation systems and public buildings [2]. Recently, a number of electronic systems aiding the visually impaired in travel and mobility have been developed [3-8]. Most of these systems require accurate information on current user location. Obtaining precise information on user location can facilitate access to public services offered in large buildings (e.g. city halls, hospitals) by aiding to locate rooms or by giving a remote guidance on how to get to the target destination. Contemporary satellite navigation systems like GPS provide positioning services sufficient for successful navigation of pedestrians in typical outdoor scenarios and thus are often incorporated in electronic travel aids (ETA) for the visually impaired. However, the use of satellite positioning systems is limited to outdoor areas only. The GPS positioning accuracy also decreases in dense urban environments where multipath propagation and strong signal attenuation result in insufficient quality of satellite beacons. Therefore, there is a need to develop dedicated systems aiding the blind and the visually impaired in urban navigation. Most of electronic mobility aids offered on the market make use of local networks of reference stations that transmit infrared [3] or radio signals [4], [5], [6]. The transmitters are used to identify various points of interest (POI) like bus stops, entrances to public buildings, etc. One of the first indoor positioning systems that used radio beacons and Received Signal Strength Indicator (RSSI) measurements was the RADAR system developed by Microsoft Research in the beginning of the 21st century [9]. From that time the problem of indoor positioning and navigation has been widely addressed around the world [10-22]. Hence maintaining low deployment and maintenance costs is among the most important objectives of the research, majority of the solutions reported in the literature rely on radio signal strength measurements. Signal strength readouts can be incorporated to location services in at least two ways. First of all, radio wave indoor propagation models can be used to determine the possible location of the terminal. This approach requires detailed description of the propagation environment and thus is difficult to implement. Another approach involves the use of database search methods to calculate user position. Therefore, it is necessary to provide reference RSSI measurements (i.e. measurements taken in predefined locations) that are stored in the reference database, and which are then used by location estimation algorithms. In the paper, we present a wireless indoor positioning system developed as a part of a complex solution aiding the visually impaired in independent travel and mobility.

This work was partially supported by the National Centre for Research and Development of Poland under grant no. NR-02 0083-10 in years 2010-2013.

II. System architecture

The architecture of the proposed indoor positioning system consists of a local localization server, a local database server and an optional global localization server. A wide range of portable user devices (PDAs, smartphones, notebooks, etc.) operating in different wireless networks may be used as system terminals. The terminals should have the capability to measure strength of the signals transmitted by system reference stations mounted inside a building. Thus, a dedicated software or hardware is necessary to make use of the measurement data, especially to pass the results to the local positioning server. The tasks of the local positioning server are: to keep information about the layout of the area it serves (e.g. an office building), to make the use of local database engine to store reference measurement data, and to compute the probable user location based on the RSSI measurement values reported by the terminal. Moreover, the local server is also responsible for communication with the global localization server, if available. The global localization server can be also

978-83-60810-53-8/$25.00

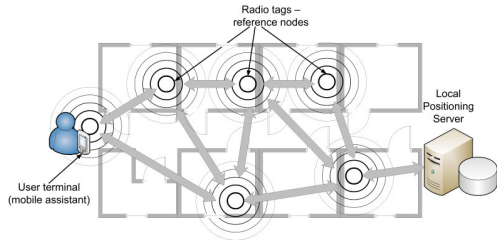


Fig. 1. Architecture of the electronic system for guidance of visually impaired in indoor areas.

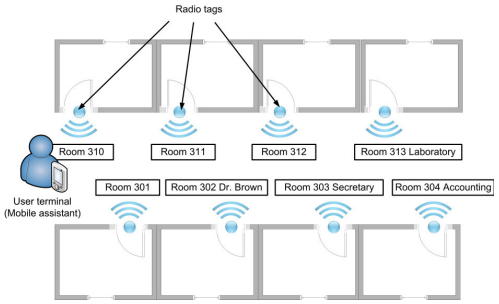


Fig. 2. Proximity detection – presentation of indoor location of the terminal and position related context information (room number).

used to deliver new positioning algorithms. The architecture of the proposed system is shown in Fig. 1. All the components of the system communicate using XML/JSON and SOAP-based Web Services.

III. Wireless Positioning Techniques

A variety of techniques can be employed to estimate the position of a wireless network terminal. In majority of systems measurements of signal parameters transmitted by system reference stations are used. Then, the position of the terminal is estimated based on calculation of distances of the terminal to at least some of the reference nodes. The most commonly used signal properties include propagation time, angle of arrival, and received signal strength [10].

A. Proximity Detection

The most straightforward method to estimate the position of a radio terminal is to determine whether it is within the coverage of some reference station. The accuracy of positioning with this approach strongly depends on the range of reference transmitters. However, when reference stations transmit signals with relatively low power, the position of the user terminal may be well approximated by the known location of the reference transmitter. This approach is called proximity detection. Practical implementation of this positioning technique involves installation of many reference nodes, often called radio tags. However, due to simple tag's construction the overall system installation cost might remain low. This technique offers good accuracy, however it strongly depends on the number of installed reference tags. The idea of positioning system using proximity detection is shown in Fig. 2.

B. Database Search-based Indoor Positioning

Distance estimation techniques involving radio wave propagation modeling are widely used in positioning systems. However, due very high complexity of indoor radio wave propagation environment, applicability of these methods is limited to outdoor areas. In typical indoor scenarios, strong multipath propagation effects make it impossible to unambiguously relate measured signal parameter value to a distance from the transmitter. Even along short propagation paths, signal parameters may exhibit very strong variability. Another factor limiting performance of positioning methods is time variability of indoor radio channel characteristics. For example, depending of the time of the day, the offices may either be crowded or almost empty what may result in significant changes of the reported values. Therefore, there is a need to search for new positioning methods for indoor applications. One of the approaches that is adequate for indoor systems assumes the use of correlation analysis of reported signal parameter values with some reference data recorded at predefined locations. As database search methods rely on evaluation of similarity of measured signal characteristics at actual location to the reference datasets, these methods are not so prone to multipath and shadowing effects as the methods based on radio wave propagation modeling. Despite of the fact that database correlation methods may be based on analysis of any available signal parameters, most of practical implementations involve received signal strength measurements. The advantage of the use of RSSI is that most of contemporary radio receivers provide possibility to monitor RSSI level and a wide range of devices can be used with a positioning system without a need to implement any hardware modifications. The use of database search methods makes it also possible to reduce the influence of RSSI time variability by the use of normalization to the value read from a given reference signal source.

C. Proposed Positioning Technique

The proposed implementation of a wireless indoor positioning system involves received signal strength (RSSI) measurements to estimate terminal position. It makes use of the advantages of both aforementioned approaches, i.e. proximity detection and database search methods. Hence proximity detection is most effective and accurate when area served by a single reference tag is relatively small, the tags should be equipped with radio transmitters supporting low transmit power modes. On the other hand, database search method accuracy increases with the number of sources of reference signals. In that case, the nodes should be capable to transmit reference signals over relatively large areas. Moreover, the coverage areas of neighboring transmitters should overlap. It is worth mentioning, that position information returned in the form of absolute geographical coordinates of the user is not the most expected output from indoor navigation and positioning systems. Geographical coordinates are more suitable for outdoor positioning, mainly due to easy integration with GIS systems. Moreover, in indoor applications accurate and

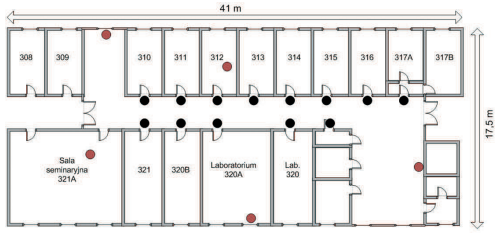


Fig. 3. Test site with reference points (black dots) and reference tags (red dots) used in the second experiment.

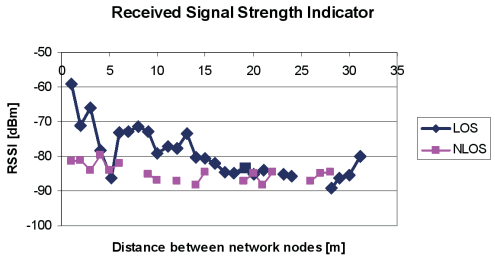


Fig. 4. RSSI as a function of distance from the transmitter.

reliable altitude estimation is required. Although in outdoor scenarios the use of absolute altitude above ground or sea level as altitude descriptor is the most convenient, in indoor scenarios floor index should be considered as the natural way of expressing in-building altitude of travelling people. Therefore, the proposed indoor positioning system makes use of area-based context-related positioning. Area-based positioning systems provide end users with context information related to the current zone of the building. The system output data set includes but is not limited to:

- floor index or name (if applicable),
- zone within a building (e.g. “north wing”),
- room or office number or its name (e.g. “kitchen” or “auditory no. 416”),
- additional site-related information (like name of current lecture in an auditory room).

Moreover, the proposed system returns absolute coordinates of the user terminal to ensure backward compatibility.

IV. Experiment Results

The prototype system was built with the use of Texas Instrument’s CC1110 radio transceivers operating in 868 MHz unlicensed band and transmitting with output power ranging from -30 dBm to +10 dBm. The reference nodes and the user terminal were equipped with omnidirectional antennas having +2.2 dBi gain. The receiver sensitivity was -110 dBm. The reference nodes were mounted in an office building as shown in Fig. 3. All network nodes were utilizing SimpliciTI protocol to communicate with each other. The test premises consisted of about 41 meters long and 3.5 meters wide corridor with doors leading to several offices and laboratories. In order to evaluate the positioning accuracy of the implemented methods a number of experiments have been conducted. The goal of the first experiment was to estimate the size of the zones covered by the reference nodes when packets are being sent with the lowest available transmit power, i.e. -30 dBm. During the experiment one of the tags has been placed in the building entrance area while the second one served as data receiver. Measurements of Received Signal Strength Indicator (RSSI) have been recorded along 36 meter long path. The experiment was repeated twice in order to examine system behavior in Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) conditions. The results of the first phase of experiments have been summarized in Fig. 4. As the result of the experiment the maximum radius of the zone covered by a single reference tag was estimated to be about 25 meters for LOS, and about 10 meters for NLOS conditions. In practical system application, node proximity will be detected only when the RSSI values exceed predefined threshold values. It must be also noted that the use of miniaturized radio tag modules equipped with ceramic antennas will result in significant decrease of the coverage area of a single tag. The goal of the next experiment was to evaluate door identification accuracy with the use of database search methods. Reference Points of Interests (POI) have been distributed at the entrance doors to the rooms as shown in Fig. 3. The database search algorithms has been used to determine the user position on the basis of RSSI measurements from five reference tags distributed in the test premises. Results of the experiment have been summarized in Table I and Fig. 5.

TABLE I
Summary of Experiment Results

Indicator	Result
Average positioning error (25 %)	0 m
Average positioning error (50 %)	2.38 m
Average positioning error (75 %)	3.23 m
Average positioning error (100 %)	4.47 m

As presented in Table I database search method resulted in mean positioning error of 4.47 meters and 2.38 meters for 50 % of cases. It is worth to mention that for 29.76 % of analyzed test cases user position was determined correctly (therefore 25 % error rate counts 0) and for another 15.48 %

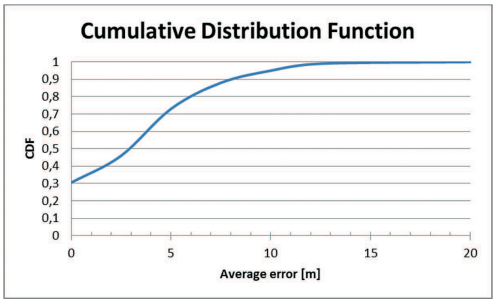


Fig. 5. Cumulative Distribution Function of positioning error from conducted experiment.

of cases user position was estimated at neighboring door. Cumulative Distribution Function is presented in Fig. 5. It can be noticed that maximum positioning error in conducted experiment reached 20 meters (for a single case) and for 90 % of cases positioning error was lower than 8 meters.

V. System user interface

Indoor positioning system presented in the article is a part of a complex solution designed for aiding the blind and the visually impaired in independent mobility and travel. Therefore, described positioning system shares components like reference tags or user terminals with the remaining part of the system. The use of Text-To-Speech enabled smartphone as a mobile electronic aid makes it possible to provide the visually impaired users with voice messages presenting position related information e.g. description of rooms the user is passing by. Moreover, an interactive plan showing part of the building where user was localized is simultaneously displayed on the screen of the mobile phone. This function makes the system suitable for a wider group of target users. Displaying additional context-related information may be helpful in effective navigation in large and unknown buildings.

VI. Conclusion

In this paper indoor positioning system for short range radio communications network has been proposed. The system is a part of complex solution designed for aiding navigation of visually impaired in independent travel and mobility. The positioning system combines proximity sensing and database search methods. The experiments conducted in a large office building resulted in average positioning error not exceeding 4.47 meters. Positioning results are returned as contextual information regarding the area where the user was localized. The use of smartphone as a user terminal makes it possible to present the results to the users in the form of voice messages. Future development works assume incorporation of multi-system positioning. As the result, the positioning methods implemented in the system will benefit from the use of data from generally available radio networks, like public Wi-Fi or mobile cellular telephony networks.

Acknowledgment

This work was partially supported by the National Centre for Research and Development of Poland under grant no. NR-02 0083-10 in years 2010-2013.

References

[1] World Health Organization, http://www.who.int/mediacentre/factsheets/fs282/en/, Accessed 22 February 2013.

[2] P. Strumiłło, "Electronic Interfaces Aiding the Visually Impaired in Environmental Access, Mobility and Navigation," *in Proc. 3rd International Conference on Human System Interaction,* Rzeszów, Poland, 2010, pp. 17–24.

[3] Talking Signs, http://www.talkingsigns.com/, Accessed 21 May 2013.

[4] K. Radecki, K. Łukaszewicz, "Lokalne radiowe systemy orientacji dla osób niewidomych w środowisku miejskim," ("Local radio systems supporting orientation of the blind in urban environment") *in Proc. Krajowa Konferencja Radiokomunikacji, Radiodyfuzji i Telewizji KKRRiT 2004,* Warsaw, Poland, 2004. (in Polish)

[5] S. Bohonos, A. Lee, A. Malik, C. Thai, and R. Manduchi, "Universal Real-Time Navigational Assistance (URNA): An Urban Bluetooth Beacon for the Blind," *in Proc. 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environment,* New York, 2007, pp. 83–88.

[6] J. Marski, P. Bajurko, K. Radecki, and T. Buczkowski, "Miniaturowe radiolatarnie i terminale z sygnalizacją RSSI do wspomagania orientacji osób niewidomych," ("Miniature radio beacons and terminals with RSSI signaling to support the orientation of the blind") Telecommunication Review – Telecommunication News (Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne), Vol. 6, 2010, pp. 320–323. (in Polish)

[7] P. Barański, M. Polańczyk, and P. Strumiłło, "A Remote Guidance System for the Blind," *in Proc. 12th IEEE International Conference on e-Health Networking, Applications and Services HealthCom 2010,* Lyon, France, 2010.

[8] M. Polańczyk, P. Skulimowski, B. Sujecki, and D. Sulmowski, "Personal Navigation System for the Blind based on Points of Interest," *in Proc. II Forum Innowacji Młodych Badaczy 2011,* Łódź, Poland, 2011.

[9] P. Bahl, V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," *in INFOCOM 2000, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies,*, vol.2, pp. 775–784, vol.2, 2000.

[10] Y. Gu, A. Lo, I. Niemegeers, "A Survey of Indoor Positioning Systems for Wireless Personal Networks," IEEE Communications Surveys & Tutorials, No. 1, 2009, pp. 13–32.

[11] M. Mendalka, K. Bizewski, Ł. Kulas, and K. Nyka, "Pattern matching localization In ZigBee wireless sensor networks," *in Proc. 18th International Conference on Microwave, Radar and Wireless Communications MIKON-2010,* Vilnius, Lithuania, 2010.

[12] N. Patwari, J.N. Ash, S. Kyperountas, A.O. Hero III, R.L. Moses, and N.S. Correal, "Locating the Nodes," IEEE Signal Processing Magazine, Vol. 22 No. 4, 2005, pp. 54–69.

[13] Ekahau Real Time Location System (RTLS), http://www.ekahau.com/products/real-time-location-system/overview.html, Accessed 20 May 2013.

[14] A. Kushki, K.N. Plataniotis, A.N. Venetsanopoulos, WLAN Positioning Systems, Cambridge University Press, Cambridge, 2012.

[15] G. Gonçalo G, H. Sarmento, "Indoor Location System using ZigBee Technology," *in Proc. Third International Conference on Sensor Technologies and Applications SENSORCOMM 2009,* Athens, Greece, 2009.

[16] K. Radecki, P. Lewicki, and J. Marski, "Lokalizacja terminala radiowego z wyjściem RSSI wewnątrz korytarza budynku," ("Localization of radio terminal with RSSI output inside the corridor") Telecommunication Review – Telecommunication News (Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne), Vol. 6, 2011, pp. 479–482. (in Polish)

[17] I. Maly, Z. Mikovec, and J. Vystrcil, "Interactive Analytical Tool for Usability Analysis of Mobile Indoor Navigation Application," *in Proc. 3rd International Conference on Human System Interaction,* Rzeszów, Poland, 2010, pp. 259–266.

[18] Jun-geun Park et al., "Growing an organic indoor location system,"*in Proceedings of the 8th international conference on Mobile systems, applications, and services,* ACM, 2010, pp. 271–284.

[19] J. Hightower, G. Borriello, "Particle filters for location estimation in ubiquitous computing: A case study," *in Proc. UbiComp 2004: Ubiquitous Computing,* 2004, pp. 88–106.

[20] L. Zekeng, I. Barakos, and S. Poslad, "Indoor location and orientation determination for wireless personal area networks," Mobile Entity Localization and Tracking in GPS-less Environments, 2009, pp. 91–105.

[21] C. Laoudias, G. Constantinou, M. Constantinides, S. Nicolaou, D. Zeinalipour-Yazti, and C. Panayiotou, "The airplace indoor positioning platform for android smartphones," *in Proc. 13th International Conference on Mobile Data Management (MDMï¿½12),* 2012.

[22] L.A. Guerrero, F. Vasquez, and S.F. Ochoa, "An Indoor Navigation System for the Visually Impaired," Sensors No. 12, 2012, pp. 8236–8258.

Information Technology for Management, Business & Society

IT4MBS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the disciplines of information technology and information systems. The IT4BMS area emphasizes the issues relevant to information technology and necessary for practical, everyday needs of business, other organizations and society at large. This area takes a sociotechnical view on information systems and relates also to ethical, social and political issues raised by information systems.

Events that constitute IT4BMS are:

- **ABICT'13 –** 4th International Workshop on Advances in Business ICT
- **Agent Day'13**
- **AITM'13 –** 11th Conference on Advanced Information Technologies for Management
- **IT4L'13** – 2nd Workshop on Information Technologies for Logistics
- **KAM'13 –** 19th Conference on Knowledge Acquisition and Management
- **MMT'13 –** 3rd Workshop on Modeling Multi-commodity Trade: Towards Smart Systems
- **TAMoCo'13** – Techniques and Applications for Mobile Commerce

4th International Workshop on Advances in Business ICT

ABICT focuses on Advances in Business ICT approached from a multidisciplinary perspective. It will provide an international forum for scientists/experts from academia and industry to discuss and exchange current results, applications, new ideas of ongoing research and experience on all aspects of Business Intelligence. ABICT will be also an opportunity to demonstrate different ideas and tools for developing and supporting organizational creativity, as well as advances in decision support systems.

We kindly invite contributions originating from any area of computer science, information technology and computational solutions for different applications areas, data integration and organizational implementation of ABICT, as well as practical ABICT solutions.

Topics

Topics include (but are not limited to):

- Advanced Technologies of Data Processing, Content Processing and Information Indexing
- Business Applications of Social Networks
- Business Data Mining and Knowledge Discovery
- Business Intelligence, Business Analytics
- Business Rules
- Business-oriented Time Series Data Mining, Analysis, and Processing
- Data Warehousing
- Information Forensics and Security, Information Management, Risk Assessment and Analysis
- Information Systems in Enterprise Management
- Information Technologies in Enterprise Logistics
- Information Technologies in Enterprise Management, Information Systems,
- Service Oriented Architectures (SOA)
- Knowledge Management
- Recommender Systems
- Semantic Web and Ontologies in Business ICT
- Virtual Enterprise
- Web-Based Data Management Systems
- Web 2.0 and Web 3.0 in fusing Business Intelligence systems and Decision Support Systems
- Knowledge Management for better Decision Support, Collaboration and Competitiveness
- Creativity Support Tools
- Customer Relationship Management, social Customer Relationship Management

Event Chairs

Mach-Król, Maria, Katowice University of Economics, Poland

Olszak, Celina M., University of Economics in Katowice, Poland

Pełech-Pilichowski, Tomasz, AGH University of Science and Technology, Poland

Program Committee

Abramowicz, Witold, Poznań University of Economics, Poland

Badica, Amelia, University of Craiova, Romania

Berio, Giuseppe, Universite de Bretagne Sud, France

Butleris, Rimantas, Kaunas University of Technology, Lithuania

Chiu, Dickson K. W., Dickson Computer Systems, Hong Kong S.A.R., China

Christozov, Dimitar, Amrican University in Bulgaria, Bulgaria

Dostal, Petr, Brno University of Technology, Czech Republic

Druzdzel, Marek, University of Pittsburgh, Bialystok University of Technology, United States

Dustdar, Schahram, Vienna University of Technology, Austria

Gackowski, Zbigniew, California State University Stanislaus, United States

Grzech, Adam, Wroclaw University of Technology, Poland

Józefowska, Joanna, Poznań University of Technology, Poland

Kacprzyk, Janusz, Institute of Computer Science, Polish Academy of Sciences, Poland

Khachidze, Manana, Tbilisi State University, Georgia

Konikowska, Beata, Institute of Computer Science, Poland

Koohang, Alex, Macon State Collage, United States

Korwin-Pawlowski, Michael L., Universite du Quebec en Outaouais, Canada

Kulczycki, Piotr, Systems Research Institute, Polish Academy of Sciences, Poland

Levy, Yair, Nova Southeastern University, United States

Ligęza, Antoni, AGH University of Science and Technology, Poland

Loucopoulos, Peri, Harokopio University of Athens, Greece

Maamar, Zakaria, Zayed University, Afghanistan

Madeyski, Lech, Wroclaw University of Technology, Poland

Miliszewska, Iwona, Victoria University, Melbourne, Australia

Ogihara, Mitsunori, University of Miami, United States

Owoc, Mieczyslaw, Wroclaw University of Economics, Poland

Petryshyn, Lubomyr, AGH University of Science and Technology, Poland

Pinheiro, Carlos Andre Reis, Katholieke Universiteit Leuven, Belgium

Prasad, T. V., Visvodaya Technical Academy, India

Pulvermueller, Elke, University Osnabrueck, Germany

Reimer, Ulrich, University of Applied Sciences St. Gallen, Switzerland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 915–921

A Hierarchical Approach for Configuring Business Processes

Mateusz Baran, Krzysztof Kluza,
Grzegorz J. Nalepa, Antoni Ligęza
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {matb,kluza,gjn,ligeza}@agh.edu.pl

***Abstract*—Business Process models in the case of real life systems are often very complex. Hierarchization allows for managing model complexity by "hiding" process details into sub-levels. This helps to avoid inconsistencies and fosters reuse of similar parts of models. Configuration, in turn, gives the opportunity to keep different models in one configurable model. In the paper, we propose an approach for configuring Business Processes that relies on hierarchization for more expressive power and simplicity. Our goal is achieved by allowing arbitrary n-to-m relationships between tasks in the merged processes. The approach preserves similar abstraction level of subprocesses in a hierarchy and allows a user to grasp the high-level flow of the merged processes.**

***Index Terms*—BPMN, Business Processes, Business Process Hierarchization, Business Process Configuration**

I. Introduction

ENTERPRISES take advantage of using Business Processes (BP) in their everyday practice. These processes define the way the company works by describing control flow between tasks. Design and development of such processes, especially more and more complex ones, require advanced methods and tools, e.g. [1], [2], [3].

Business Process Model and Notation (BPMN) [4] is a visual language used to model Business Processes. It is a set of graphical elements denoting such constructs as activities, splits and joins, events etc. (see Figure 1). These elements can be connected using control flow and provide a visual description of process logic [5]. Thus, a visual model is easier to understand than textual description and helps to manage software complexity [6].

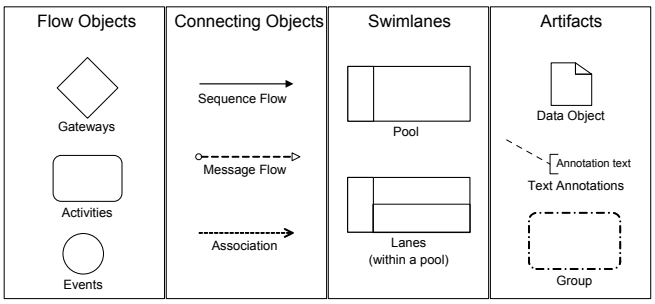


Figure 1. BPMN core objects

Although the BPMN notation is very rich when considering the number of elements and possible constructs, apart from the notation rules, some style directions for modelers are often used [7]. To deal with the actual BP complexity, analysts use various modularization techniques. Mostly, they benefit from their experience and modularize processes manually during the design because these techniques are not standardized. Modularization issue is important in the case of understandability of models [8]. Thus, guidelines for analysts, such as [9], emphasize the role of using a limited number of elements and decomposing a process model.

In the case of large collection of processes [10], especially modeled by different analysts, processes can be modularized in different ways on distinct granularity level [8]. Moreover, the processes in the collection can be similar [11], but this similarity is lost when the models are kept separately.

Automatic hierarchization and configuration, which is the subject of this paper, can help in preventing these problems. Hierarchization provides different abstraction levels and, properly developed, can ensure the same way of modularization for all the processes. Configuration gives the opportunity of unification of processes and enables to keep different models in one configurable model.

The rest of this paper is organized as follows: Section II presents motivation for our research. Section III and IV describe related works in the hierarchizaton and configuration areas. In Section V, we present our configuration BP approach which takes advantage of hierarchization. We evaluated the proposed approach based on the issue tracker case study. The paper is summarized in Section VI.

II. Motivation

Modern business applications require advanced modeling solutions to deal with the model complexity. Thus, several challenges in Business Process modeling can be distinguished:

- the *granularity modularization* challenge – how to model different processes similarly, especially in respect of abstraction layers [12].
- the *similarity capturing* challenge – it is not easy to grasp similarities in the collection of only partially similar models [13], especially if they are modularized differently.
- the *collection storing* challenge – how to store a large collection of models in some optimized way [14].

978-83-60810-53-8/$25.00

In our research, we deal with these challenges using automatic hierarchization that allows us to preserve similar abstraction level of subprocesses in a hierarchy. To address two other challenges, we propose a new BP configuration technique, that allows us to express similarities between different BP models in a simple, but comprehensive way. Our configuration is based on the hierarchical model prepared previously.

The aim of this paper is to present our approach for configuring Business Processes based on hierarchization. We present the automatic hierarchization algorithm that takes advantage of task taxonomy and the algorithm for configuration.

The proposed approach has several advantages. Thanks to the use of hierarchization, the obtained model structure incorporates similar activities, and the configuration step can be simplified. Thus, the configurable process diagram is easy to comprehend. Contrary to the existing configuration methods, in our approach we can bind not only one task with another, but also groups of tasks which were considered in hierarchization step. Such configuration technique address the abovementioned problems by managing both process model complexity and diversity.

III. Hierarchization Issues in Business Process Models

La Rosa et al. [15] distinguished 3 abstract syntax modifications that are related to modularization for managing process model complexity. These are:

1) vertical modularization – a pattern for decomposing a model into vertical modules, i.e. subprocesses, according to a hierarchical structure
2) horizontal modularization – a pattern for partitioning a model into peer modules, i.e. breaking down a model into smaller and more easily manageable parts, especially assigned to different users in order to facilitate collaboration.
3) orthogonal modularization – a pattern for decomposing a model along the crosscutting concerns of the modeling domain, such as security or privacy, which are scattered across several model elements or modules.

Our approach is consistent with the vertical and horizontal patterns. Although we decompose a model into subprocesses, in fact we use some additional information to decompose it, such as task assignment or task categories, which is an example of the second pattern instance. The last one, orthogonal pattern, requires to extend the notation, as in [16]; thus, it is not our case.

It is important to notice that such decomposition has several advantages:

- increases their understandability by "hiding" process details into sub-levels [8],
- decreases redundancy, helps avoid inconsistencies and fosters reuse by referring to a subprocess from several places [17], [18], [13], [19],
- decreases the error possibility [20],
- increases maintainability of such processes [9].

IV. Business Process Configuration

Business Process configuration is a tool for expressing similarities between different Business Process models. There are mechanisms for managing and comparing processes in large repositories [21], [22], refactoring of such repositories [14] as well as automatic extraction methods for cloned fragments in such process model repositories [14], [17]. However, our case differs from the existing approaches because we do not base on any directly visible similarity, but on previously defined taxonomy of states or roles in the processes etc. Moreover, our hierarchization algorithm forces the generation of similar models as a result.

There are a few methods of extraction of configurable processes. They focus on different goals. Analyzing digest configurable Business Process reveals high-level workflow that might not be apparent in particular models. The structure is partially lost in the process so this does not concern our approach.

The method of interest in this article are models merged into configurable model [23]. They allow the analyst to see several processes as special cases of one configurable model. The model emphasizes similarities preserving all the details.

Configurable Business Processes are also a good alternative to current reference model bases such as SAP. Instead of presenting the analyst a few example models, a more general, configurable solution can be delivered. It makes producing final models faster and less error-prone [24].

There is an active research field in the area of configurable Business Processes. In [25], Rosemann et al. describe an approach focused on hand-made diagrams for the purpose of reference modeling. La Rosa et al. [24] extend it with roles and objects.

Variant-rich process models were explored in PESOA project [26], [27]. They enable process designer to specify a few variants of a task.

Our approach is a specialized version of solution proposed by La Rosa in [28]. Hierarchization algorithm produces models of very specific structure and this fact is exploited in our approach.

V. Hierarchization-based Business Process Configuration Approach

We propose an approach for configuring Business Processes that relies on hierarchization for more expressive power and simplicity. The first goal is achieved by allowing arbitrary n-to-m relationships between tasks in merged processes. Taxonomy of tasks provides level of flexibility which many of current state-of-the-art solutions are lacking [25], [28].

Simplicity is the effect of constructing a very specific type of models during hierarchization. These models do not require general configuration algorithms that often produce complicated, hard to analyze diagrams. Sufficient configuration algorithm is described in Section V-C.

General flow of data in the whole approach is depicted in Figure 2.

Figure 2. General flow of data in proposed approach

A. Case Study

To present our configuration approach, we chose 3 different BPMN models of bug tracking systems: Django[1], JIRA[2] and the model of the issue tracking approach in VersionOne[3].

A bug tracking system is a software application which helps in tracking and documenting the reported software bugs (or other software issues in a more general case). Such systems are often integrated with other software project management applications, such as in VersionOne, because they are valuable for the company.

Thus, apart from popularity, we selected such a case study because these kinds of processes have similar users assigned to similar kinds of tasks, the processes of different bug trackers present the existing variability, and such an example can be easily used to present our algorithm in a comprehensive way.

B. Automatic Hierarchization Algorithm

The hierarchization algorithm is given a BPMN model, a set of high-level BPMN tasks and an assignment of BPMN model's tasks to the high-level tasks. Using this information it constructs two-level hierarchical diagram. The lower level contains one diagram for each high-level task. Higher level diagram contains high level tasks (with lower-level diagram as subprocesses). This is done in such way to maximize simplicity and preserve semantics of original model.

Hierarchization is performed in several steps.

1) **Introduction of high-level expanded subprocesses**
In the first step, expanded subprocesses are introduced. Tasks are assigned to them according to given specification. Gateways are placed outside of all subprocesses unless all their incoming and outgoing flows lead to tasks of the same process. Intra-subprocess flows are kept. Inter-subprocess flows are replaced by:
   a) flow from source element to end event (in subprocess),
   b) OR-gateway right after subprocess (one per subprocess),
   c) flow from introduced gateway to target of initial flow with condition 'subprocess ended in event introduced in 1a'.

   This step is depicted in Figure 3. After this step is performed, assumption 1 of configuration algorithm (Section V-C) is fulfilled.

[1]See: https://code.djangoproject.com/
[2]See: http://www.atlassian.com/software/jira/
[3]See: http://www.versionone.com/

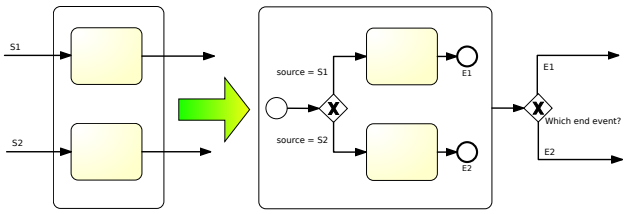


Figure 3. First step of hierarchization algorithm

2) **Gateway simplification**
The previous step introduced new gateways. The original diagram may contain unnecessary gateways too. Creating configurable diagram in proposed approach requires a very specific structure of high-level model. It can be achieved through gateway simplification.
The process of gateway simplification is depicted in Figure 4. In a simplified model one gateway G is placed after every subprocess $S(G)$ (unless it has only one outgoing flow which does not end in a gateway). Gateway G has outgoing flows to all subprocesses and end events reachable in original model from $S(G)$. Conditions labeling these flows are determined as follows.
Let $flow_N(G,T)$ and $flow_O(G,T)$ be the flows in the new and old graphs respectively from gateway G to target item T. Let $C_N(G,T)$ and $C_O(G,T)$ be the conditions on flow $flow_N(G,T)$ and $flow_O(G,T)$ respectively. Let $P(G,T)$ be the set of all paths in old graph (all gateways appear at most once) from G to P. Let $L(G)$ be the set of loops in gateway graph reachable from gateway G. Then the following hold:

$all((G_1, G_2, \ldots, G_k), T)$ holds iff $C_O(G_k, T)$ and
for all $i \in \{1, 2, \ldots, k-1\}$ holds $C_O(G_i, G_{i+1})$
$C_N(G,T)$ holds iff exists $p \in P(G,T)$ such that $all(p,T)$

Presented procedure works as long as graph of gateways is acyclic. Cycles need additional compensation for the fact that infinite looping is possible. We propose a solution where a new task "loop infinitely" is added and connected by flow from all gateways that allow looping. Condition on the new flow may be defined by analogy to the previous case:

$all((G_1, G_2, \ldots, G_k))$ holds iff $C_O(G_k, G_1)$ and
for all $i \in \{1, 2, \ldots, k-1\}$ holds $C_O(G_i, G_{i+1})$
$C_N(G,T)$ holds iff exists $p \in L(G)$ such that $all(p)$

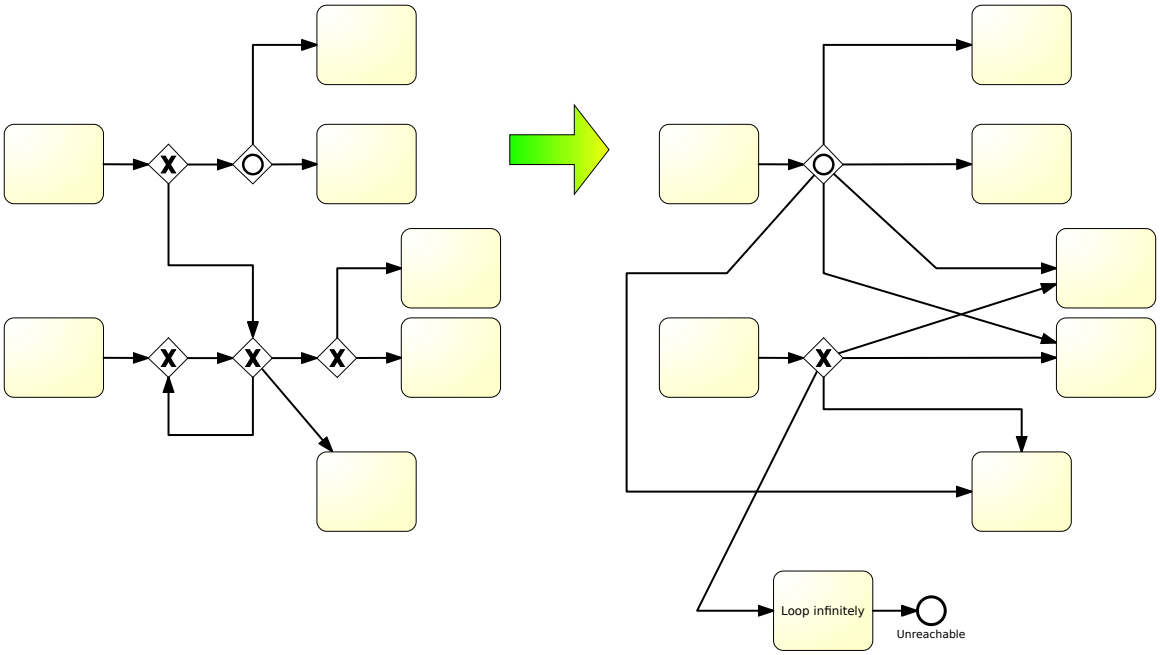


Figure 4. Second step of hierarchization algorithm (gateway simplification)

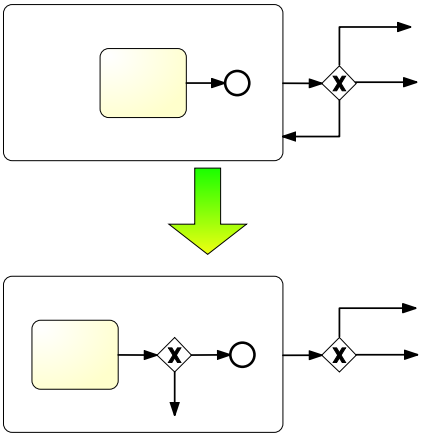


Figure 5. The idea of third step of hierarchization

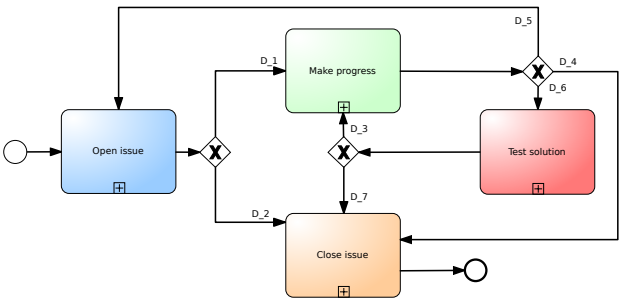


Figure 6. High-level diagram of Django issue tracking

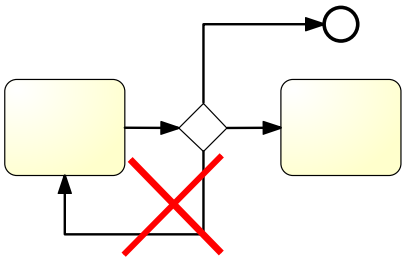


Figure 7. Possible flows to and from a gateway

Figure 4 shows a graph of gateways before and after simplification. Simplification assures that requirements 2 and 3 from Section V-C are fulfilled.

3) **Removal of recurring flows**
Last step of hierarchization is elimination of recurring flows. By this a flow from a gateway to activity preceding this gateway is meant (see Figure 5).
Before each end event in the subprocess that can result in recurring flow a new XOR gateway is placed. It has two output flows: one to end event and one to task or gateway a recurring flow would lead to (see Figure 5). The condition on the latter flow is created according to condition on original recurring flow.
This step of hierarchization algorithm makes requirement 4 of configuration process fulfilled. The result of hierarchization of Django issue tracking process can be seen in Figure 6.

C. Process Configuration

Let us be given N BPMN models such that:

1) all of them share the same set of tasks,
2) flows outgoing from tasks or start event end in a different task, end event or an (XOR or OR) gateway (Figure 7),
3) flows outgoing from gateways always end in tasks or an end event,
4) no flow outgoing from a gateway leads to a task that has a flow to this gateway.

Then the configurable model that entails all the given models can be defined as follows:

1) configurable diagram has one start event, all the specified tasks and all the end events from N given models,
2) for all tasks and start event (let i be the current item):
   a) If all diagrams have flow outgoing from i that ends in (the same) task or the only end event then the same flow exists in merged diagram.
   b) If in at least one model the flow f ends in a gateway, the merged model has a configurable gateway after i. It is a configurable type gateway if there are diagrams with two different types of gateways (or one without gateway).
   c) If and only if any input diagram gateway after i has a flow to an item, the configurable gateway has a flow to this item too. The flows are labeled with model number and condition from that model.

D. Approach Evaluation

As we tested our approach on the three issue tracking systems, the results we got are optimistic. The obtained model is simple and comprehensible. Figure 9 compares initial and hierarchical versions of Django system. The three hierarchized models, simplified by the algorithm, can be simultaneously compared on high level and on the subprocess level. The final high level configurable model is presented in Figure 8.

One of the drawbacks of our approach is that conditions on control flows outgoing from gateways may become complex after hierarchization. However, it is not an obstacle in understanding of high level flow in the process, which is the goal of the approach.

VI. Conclusion and Future Work

The research presented in this paper addresses three challenges in Business Process modeling, which we distinguished in Section II. These are granularity modularization, similarity capturing and collection storing challenges.

In the paper, we proposed automatic hierarchization algorithm that takes advantage of task taxonomy and allows us to preserve similar abstraction level of subprocesses in a hierarchy. A Business Process configuration technique, based on the hierarchization result is presented as well. It allows for expressing similarities between different BP models in a simple but comprehensive way. Thanks to this, a user can grasp the high-level flow of the merged processes.

In comparison to other approaches, our hierarchization algorithm supports arbitrary n-to-m relationships between tasks in the merged processes.

To get a proof of concept of our approach, we narrowed our attention to the subset of BPMN, similarly expressive to EPC. Thanks to the use of the taxonomy shared by the three models and the hierarchization algorithm, the configuration approach is straightforward.

In future work, we consider to extend the approach in several ways, e.g. to allow more BPMN elements or multi-level diagrams [1], and to integrate it with Business Rules [29], especially in the XTT2 representation [30], [31], in order to use control flow as inference flow [32] and to allow for automatic verification of models [33], [34], [35]. Moreover, automatic generation of taxonomy using some process metrics [11], [36] is also considered, as well as automatic assignment of tasks to subprocesses based on Natural Language Processing.

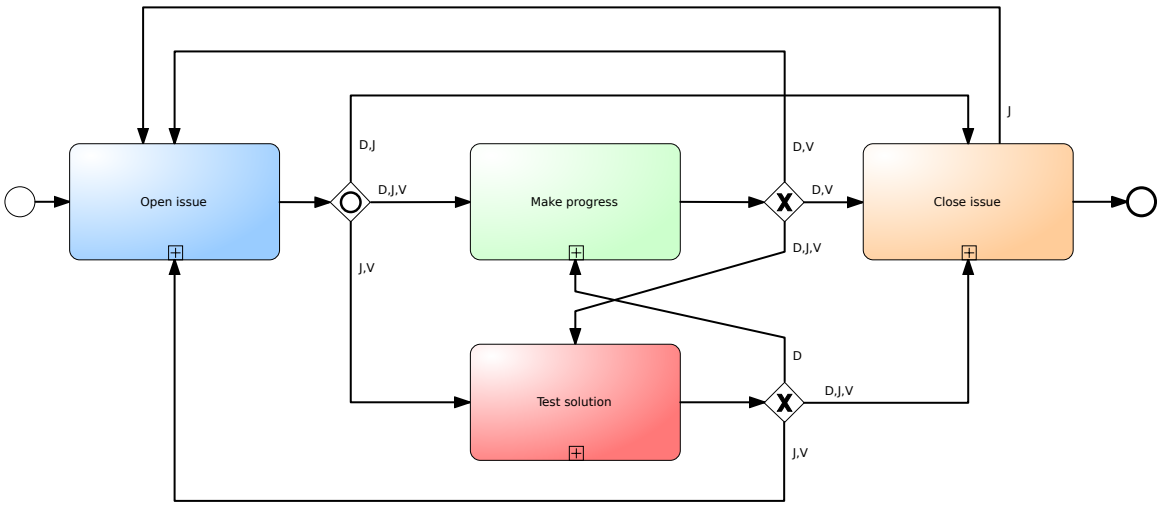


Figure 8. Result of the proposed algorithm (after configuration)

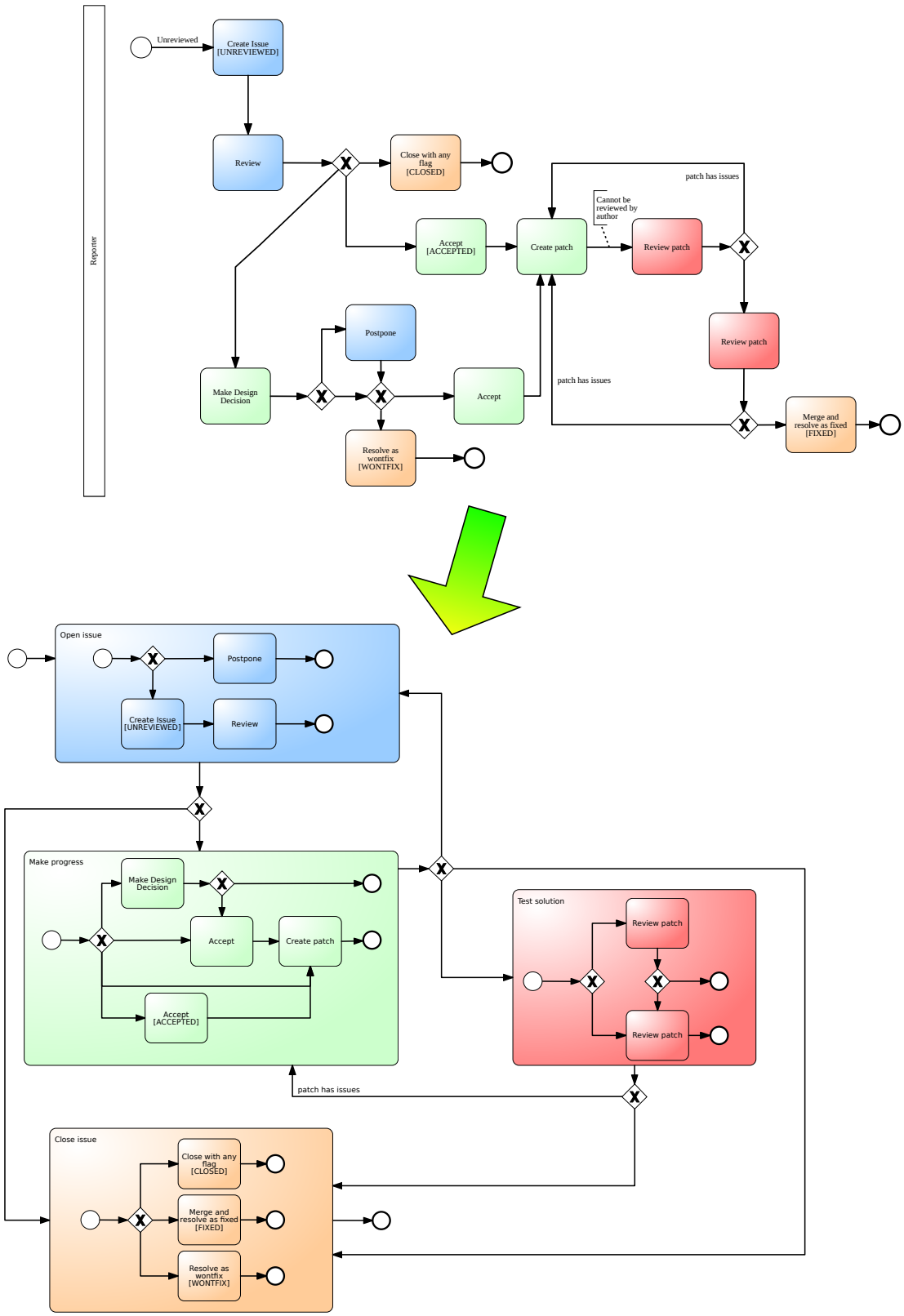


Figure 9. Comparison of initial diagram and its hierarchical version

ACKNOWLEDGMENT

The paper is supported by the AGH UST Grant.

REFERENCES

[1] K. Kluza, K. Kaczor, and G. J. Nalepa, "Enriching business processes with rules using the Oryx BPMN editor," in *Artificial Intelligence and Soft Computing: 11th International Conference, ICAISC 2012: Zakopane, Poland, April 29–May 3, 2012*, ser. Lecture Notes in Artificial Intelligence, L. Rutkowski and [et al.], Eds., vol. 7268. Springer, 2012, pp. 573–581.

[2] K. Kaczor, G. J. Nalepa, Ł. Łysik, and K. Kluza, "Visual design of Drools rule bases using the XTT2 method," in *Semantic Methods for Knowledge Management and Communication*, ser. Studies in Computational Intelligence, R. Katarzyniak, T.-F. Chiu, C.-F. Hong, and N. Nguyen, Eds. Springer-Verlag, 2011, vol. 381, pp. 57–66, DOI: 10.1007/978-3-642-23418-7. [Online]. Available: http://www.springerlink.com/content/h544g4238716m320/

[3] M. Szpyrka, "Exclusion rule-based systems – case study," in *International Multiconference on Computer Science and Information Technology*, vol. 3, Wisła, Poland, October 20-22 2008, pp. 237–242.

[4] OMG, "Business Process Model and Notation (BPMN): Version 2.0 specification," Object Management Group, Tech. Rep. formal/2011-01-03, January 2011.

[5] T. Allweyer, *BPMN 2.0. Introduction to the Standard for Business Process Modeling*. Norderstedt: BoD, 2010.

[6] G. J. Nalepa and K. Kluza, "UML representation for rule-based application models with XTT2-based business rules," *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, vol. 22, no. 4, pp. 485–524, 2012.

[7] B. Silver, *BPMN Method and Style*. Cody-Cassidy Press, 2009.

[8] H. Reijers, J. Mendling, and R. Dijkman, "Human and automatic modularizations of process models to enhance their comprehension," *Information Systems*, vol. 36, no. 5, pp. 881–897, 2011.

[9] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven process modeling guidelines (7pmg)," *Information & Software Technology*, vol. 52, no. 2, pp. 127–136, Feb 2010.

[10] Z. Yan, R. Dijkman, and P. Grefen, "Business process model repositories – framework and survey," *Information and Software Technology*, vol. 54, no. 4, pp. 380–395, 2012.

[11] R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, and J. Mendling, "Similarity of business process models: Metrics and evaluation," *Information Systems*, vol. 36, no. 2, pp. 498–516, Apr 2011.

[12] D. V. Nuffel and M. D. Backer, "Multi-abstraction layered business process modeling," *Computers in Industry*, vol. 63, no. 2, pp. 131–147, 2012.

[13] F. Pittke, H. Leopold, J. Mendling, and G. Tamm, "Enabling reuse of process models through the detection of similar process parts," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, M. Rosa and P. Soffer, Eds. Springer Berlin Heidelberg, 2013, vol. 132, pp. 586–597.

[14] B. Weber, M. Reichert, J. Mendling, and H. A. Reijers, "Refactoring large process model repositories," *Computers in Industry*, vol. 62, no. 5, pp. 467–486, 2011.

[15] M. La Rosa, P. Wohed, J. Mendling, A. ter Hofstede, H. Reijers, and W. M. P. Van der Aalst, "Managing process model complexity via abstract syntax modifications," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 4, pp. 614–629, 2011.

[16] C. Cappelli, J. C. Leite, T. Batista, and L. Silva, "An aspect-oriented approach to business process modeling," in *Proceedings of the 15th workshop on Early aspects*, ser. EA '09. New York, NY, USA: ACM, 2009, pp. 7–12.

[17] R. Uba, M. Dumas, L. García-Bañuelos, and M. Rosa, "Clone detection in repositories of business process models," in *Business Process Management*, ser. Lecture Notes in Computer Science, S. Rinderle-Ma, F. Toumani, and K. Wolf, Eds. Springer Berlin Heidelberg, 2011, vol. 6896, pp. 248–264.

[18] M. Dumas, L. García-Bañuelos, M. L. Rosa, and R. Uba, "Fast detection of exact clones in business process model repositories," *Information Systems*, vol. 38, no. 4, pp. 619–633, 2013.

[19] N. Zaaboub Haddar, L. Makni, and H. Ben Abdallah, "Literature review of reuse in business process modeling," *Software & Systems Modeling*, pp. 1–15, 2012.

[20] J. Mendling, G. Neumann, and W. Aalst, "Understanding the occurrence of errors in process models based on metrics," in *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*, ser. Lecture Notes in Computer Science, R. Meersman and Z. Tari, Eds. Springer Berlin Heidelberg, 2007, vol. 4803, pp. 113–130.

[21] M. Kunze and M. Weske, "Metric trees for efficient similarity search in large process model repositories," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, M. Muehlen and J. Su, Eds. Springer Berlin Heidelberg, 2011, vol. 66, pp. 535–546.

[22] R. Dijkman, M. L. Rosa, and H. A. Reijers, "Managing large collections of business process models – current techniques and challenges," *Computers in Industry*, vol. 63, no. 2, pp. 91–97, 2012.

[23] W. M. van der Aalst, "Business process management: A comprehensive survey," *ISRN Software Engineering*, vol. 2013, 2013.

[24] M. L. Rosa, M. Dumas, A. H. ter Hofstede, and J. Mendling, "Configurable multi-perspective business process models," *Information Systems*, vol. 36, no. 2, pp. 313 – 340, 2011.

[25] M. Rosemann and W. M. P. van der Aalst, "A configurable reference modelling language," *Inf. Syst.*, vol. 32, no. 1, pp. 1–23, Mar. 2007.

[26] F. Puhlmann, A. Schnieders, J. Weiland, and M. Weske, "Variability Mechanisms for Process Models. PESOA-Report TR 17/2005, Process Family Engineering in Service-Oriented Applications (pesoa). BMBF-Project."

[27] A. Schnieders and F. Puhlmann, "Variability mechanisms in e-business process families," in *Proc. International Conference on Business Information Systems (BIS 2006*, 2006, pp. 583–601.

[28] M. L. Rosa, M. Dumas, R. Uba, and R. M. Dijkman, "Business process model merging : An approach to business process consolidation," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 22, no. 2, 2013.

[29] G. J. Nalepa, "Proposal of business process and rules modeling with the XTT method," in *Symbolic and numeric algorithms for scientific computing, 2007. SYNASC Ninth international symposium. September 26–29*, V. Negru and et al., Eds., IEEE Computer Society. Los Alamitos, California ; Washington ; Tokyo: IEEE, CPS Conference Publishing Service, september 2007, pp. 500–506.

[30] G. J. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[31] A. Ligęza and G. J. Nalepa, "A study of methodological issues in design and development of rule-based systems: proposal of a new approach," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 117–137, 2011.

[32] G. Nalepa, S. Bobek, A. Ligęza, and K. Kaczor, "Algorithms for rule inference in modularized rule bases," in *Rule-Based Reasoning, Programming, and Applications*, ser. Lecture Notes in Computer Science, N. Bassiliades, G. Governatori, and A. Paschke, Eds., vol. 6826. Springer Berlin / Heidelberg, 2011, pp. 305–312.

[33] K. Kluza, T. Maślanka, G. J. Nalepa, and A. Ligęza, "Proposal of representing BPMN diagrams with XTT2-based business rules," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 243–248.

[34] M. Szpyrka, G. J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 249–255.

[35] F. Coenen *et al*, "Validation and verification of knowledge-based systems: report on eurovav99," *The Knowledge Engineering Review*, vol. 15, no. 2, pp. 187–196, 2000.

[36] K. Kluza and G. J. Nalepa, "Proposal of square metrics for measuring business process model complexity," in *Proceedings of the Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 919–922.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 923–926

Simulation driven design of the German toll system – profiling simulation performance

Tommy Baumann*, Bernd Pfitzinger†‡ , Thomas Jestädt†
*Andato GmbH & Co. KG, Ehrenbergstraße 11, 98693 Ilmenau, Germany. Email: tommy.baumann@andato.com
†Toll Collect GmbH, Linkstraße 4, 10785 Berlin, Germany. Email: {bernd.pfitzinger|thomas.jestaedt}@toll-collect.de
‡FOM Hochschule für Oekonomie & Management, Bismarckstraße 107, 10625 Berlin, Germany

***Abstract*—Taking an existing large-scale simulation model of the German toll system we identify the typical workload by profiling the runtime behavior. Crucial performance hot spots are identified and related to the real-world application to analyze and evaluate the observed efficiency. In a benchmark approach we compare the observed performance to different simulation frameworks.**

I. Introduction

As technology advances, systems and processes with higher complexity, interconnectedness and heterogeneity can be developed. Simultaneously, user requirements are constantly increasing: Software evolution is a fact of life. Modeling and simulation techniques are applied to design, analyze, evaluate, validate, and optimize such systems. The article analyzes the performance of a large-scale Discrete Event Simulation (DES, [1]) simulation model of the German toll system implemented in MSArchitect [2] – similar to and larger than existing simulation models [3, 4].

The next section gives an overview of the automatic German toll system, the corresponding simulation model and typical simulation results. Section III introduces the simulation framework architecture and performance measurement techniques. Section IV analyzes and evaluates the simulation performance of several DES kernels using small test models. Section V analyses the performance within the application domain followed by a summary in section VI.

II. Executable model of the German toll system

For the application domain we use an existing simulation model of the German toll system [5–7], a large-scale autonomous toll system [8] operated by Toll Collect GmbH. The tolls for heavy-goods vehicles (HGVs) driving on federal motorways – a total of 4.36 bn € in 2012 [9] – is collected by the toll system, more than 90% fully automatic using the more than 750 000 on-board units (OBUs) deployed in the HGVs. The simulation model includes all subsystems necessary for the automatic tolling processes (fig. 1) and for delivering updates to the OBU software, geo and tariff data as well as a model of the user interaction [10].

From the process perspective the simulation model covers business and system processes differing at least 7 orders of magnitude in time: All major technical processes with durations of one second and longer are included in the model aiming to predict the dynamic system behavior of fleet-wide

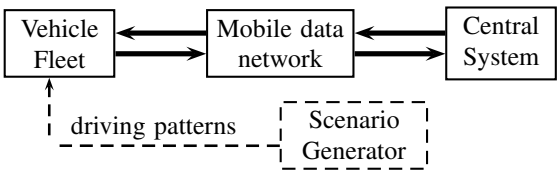


Fig. 1. High-Level simulation model of the Toll Collect system (upper half) and the model for the user interaction (scenario generator, lower half).

updates (taking weeks to months). In fact, the model includes some processes with higher temporal resolution (down to 50 ms for the connection handling by the firewalls). Using the Pearson correlation as metric to compare the simulation results with the observed update rates between 04/2012 and 01/2013 we find the correlation to be above (better than) 0,994.

Even on the application level the user interaction (precalculated driving patterns) creates a large number of events to be processed by the simulation logic. On average each OBU will be powered-on for 16% of the time and process tolls for 32 000 km annually ([11], one toll event per 4.2 km on average [12]) spread across 1 300 power cycles (including three times as many periods of loss of access to the mobile data network). Of course, many more events are created from within the application logic, e.g. to forward tolls to the central systems or to run error recovery protocols in the case of network unavailability.

III. Simulator architecture and performance measurement

This section describes the architecture of the simulator and different possibilities to measure and evaluate simulation performance. MSArchitect distinguishes between atomic models and composite models (as most actor-oriented DES/PDES tools) to capture the behavior and structure of systems and processes. Atomics interact with the simulation kernel and contain the whole behavioral description, including event consumption and creation, time advance of events, and manipulation of data entities. They are typically written in C++ (all common programming languages work as well) and compiled into binary code for execution. The composite models provide the structural composition of models, but do not themselves contribute to the execution semantics. Combining both modeling types results in a hierarchical model tree with atomic models as leafs and composite models as nodes, as shown in fig. 2.

978-83-60810-53-8/$25.00

TABLE I
TECHNICAL REPRESENTATION LAYERS USED IN MSARCHITECT

Technical layer	Vocabulary	Performance factors
DES Composition	interconnection of ports, states and, parameters	model hierarchy, summable model structure, data exchange via ports, forks and merges, sharing of states, port multiplicities
DES Atomic Interface	ports, states, parameters, lifecycle methods, inheritance, model/data types	data type conversion, reuse of data objects, memory organization
Code	C++ types, variables, instructions, method calls, control/data-flow, inheritance	kernel event scheduling, C++ type resolution, utilization of external code, memory allocation, caching, code granularity/distribution
Machine	Object code, register, memory, instructions (arithmetic, jump, call)	CPU capabilities (instruction set, latencies, cache), memory, code structure

The definition of atomic models and composite models relies on a well-defined application programming interface (API) to the simulation kernel: Typically consisting of ports (communication interfaces), states, parameters and methods [13].

Based on the simulation kernel API definition and the possibilities for model description four technical representation layers can be differentiated in MSArchitect (tab. I, in principle applicable to all DES tools) with different factors impacting the performance on each layer. The *DES Composition Layer* describes the interconnection of models within a composite model and allows analyzing structural dependencies and properties [13]. On this layer, much of the performance depends on the application level behavior and its implementation as abstract simulation model. Starting with the *DES Atomic Interface Layer* the performance becomes independent of the application domain and is determined by the simulation toolset ([14], which describes the interface of atomic models as well as restrictions on features available for the functional description). The *Code Layer* contains the functional implementation of all atomics in the form of lifecycle methods and custom code sections. The *Machine Layer* contains preprocessed and compiled code for model execution and references to external runtime libraries to be executed on a given CPU architecture.

The layers define the information available for profiling the runtime behavior, collected either by the simulation kernel, simulation logs or (external) performance profiling tools. This analysis repeats the kernel benchmarks presented in [7] and expands the performance analysis to the application level taking the example of a real-world simulation in section V.

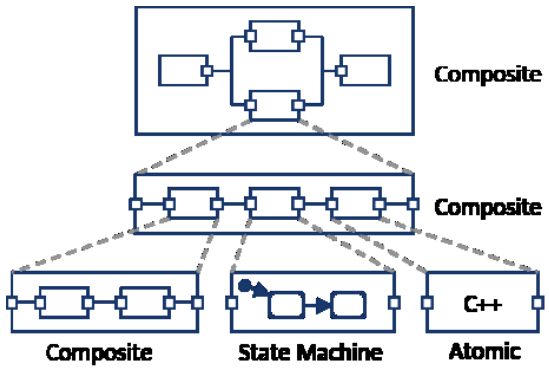


Fig. 2. Model hierarchy: Composite models and Atomics.

IV. BENCHMARK OF SIMULATION KERNEL PERFORMANCE

In this section we perform a kernel benchmark using the five test models introduced in [7]. With regards to the technical representation layers in tab. I the test models are defined on the *DES Composition Layer* and the *DES Atomic Interface Layer*. We included the most important performance influencing factors in our tests: The Future Event List (FEL) management, memory and data type management, pseudo-random number generator performance, and arithmetic operations performance [15].

Each test model is simulated with a set of simulation parameters using different system design tools. We selected six system design tools for evaluation: Ptolemy II, Omnet++, AnyLogic, MLDesigner, SimEvents and MSArchitect. All tools were run in serial mode on an Intel Core i7 X990 at 3.47 GHz with 24 GiByte RAM using either Windows 7 Enterprise (64 bit) or openSuse 11.4 (32 bit, kernel 2.6.37.6).

Fig. 3 gives the results of the event processing performance of the Runtime Scaling test. The tests show that neither the event processing performance nor the memory consumption is affected by increasing the simulation runtime and only OMNeT++ is sensitive to the additional hierarchy levels. However, from the test results it is already obvious that the different tools vary in event processing performance by an order of magnitude: MSArchitect provides the highest speed. MLDesigner, AnyLogic, and OMNeT++ provide 25% of the speed (compared to [7] the MSArchitect performance improved by more than 30%). Ptolemy II is twenty times slower. Looking at the memory usage during the simulation the difference between the tools is again more than an order

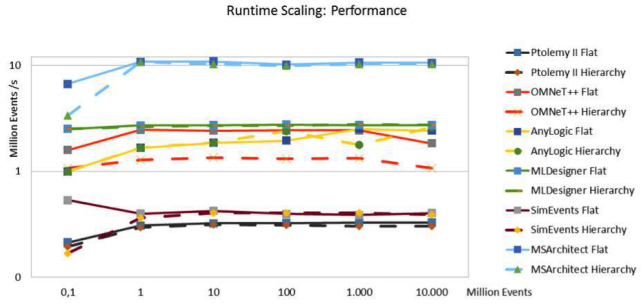


Fig. 3. Runtime performance (top) and memory usage (bottom) scaling of different DES tools.

of magnitude – the slowest tool using the most memory and the fastest tool using the least.

Repeating the FEL Size Scaling test we observe – as expected – a systematic performance reduction with increasing FEL size, due to the increasing overhead for FEL management. With increasing FEL size three of the five tools develop drastic performance degradation coinciding with a rapid grow of memory consumption. In absolute numbers, MSArchitect has the best test performance and the lowest memory usage until FEL size 10^6. Subsequently the memory usage of MLDesigner is lower since MSArchitect runs in 64 bit mode only. However, in our tests MLDesigner stopped to work for FEL sizes above $15 \cdot 10^7$. We tested MSArchitect successfully with a FEL size of 10^8.

The comparison of the DES tools using the FEL Adaption, Data Type Management and Random Number Generation tests yield the same findings as presented previously ([7], with some improvements in MSArchitects' RNG performance): Both, MSArchitect and OMNeT++ show the best performance in some categories. MSArchitect is the fastest simulation kernel in most categories and requires least memory for data handling.

V. Profiling of the simulation model

To evaluate the application-level simulation performance of our model of the German toll system, we use both the kernel logging capabilities of MSArchitect and an external profiling application (Intel VTune). Kernel logging allows to count the number of calls of atomic models as well as the total number of samples (corresponding to a processor cycle). The external profiler allows measuring the time and space complexity as well as chip-level details on the instruction execution.

To profile the simulation model we take the simulation scenario used to verify the simulation model of ther automatic German toll system against real-world data (see section II) using pre-calculated user interaction. The simulation model reads and parses the driving patterns (containing the events for power cycles, tolling and network unavailability) and feeds the events into the simulation model. The simulation model in turn creates many more events (e.g. for scheduling timeouts) to be processed by the simulation kernel. Using a single CPU core the simulation run encompassing a fleet of 700 000 OBUs and a simulated time period of 45 weeks takes less than 10 hours to compute.

As mentioned above, in a first step we apply the kernel logging capabilities of MSArchitect resulting in a file with profiling information at the end of the simulation run. Tab. II shows an excerpt of the file, containing the top-10 (out of 65) atomic blocks by runtime.

First of all, the atomic block "AccessSessionStateSwitch" is striking, consuming a large amount of time with a high number of calls. The block is responsible for switching OBU data structures in response to its state to one of the output ports. As the block switches between 34 states, 539 samples per call are acceptable. Nevertheless the number of calls could be reduced for performance improvement by changing the model architecture – especially once the model is ported to the parallel DES core.

TABLE II
KERNEL LOGGING RESULTS: TOP 10 ATOMIC BLOCKS BY RUNTIME.

Atomic Block	Calls [M]	Samples [G]	Time [%]	Samples per Call
AccessSessionStateSwitch	19 980	10 760	10,89	539
ExternDStxt	0,0007	9 565	9,68	13 665 M
StaHandling	482	6 503	6,58	13 483
EinzelbuchungsHandling	4 660	6 442	6,52	1 382
IpAutomat	7 323	5 749	5,82	785
Delay (Standard)	8 874	5 522	5,59	622
CheckComponentState	7 363	3 859	3,91	524
NetzverlustHandling	3 020	3 479	3,52	1 152
AccessSessionStateWrite	5 841	3 215	3,26	551
MfbSwitch	5 525	3 196	3,24	579

The next conspicuous atomic block is "ExternDStxt", reading the pre-generated files provided by the scenario generator model as ASCII file. The block consumes 9,681% of the runtime for 13 665 M samples/call and is rarely executed (twice per simulated day). In order to reduce the load, scenarios should be computed on the fly. The atomic block "StaHandling" is responsible for generating and controlling status requests, which may result in update processes. The block consumes 6,581% of simulation time. We see potential for improvements in changing the implementation (e.g. conversion of formulas to save operations, replacing divisions by multiplications with reciprocal and using of compare functions from standard libraries).

With 4 660 M calls "EinzelbuchungsHandling" is a frequently executed atomic block. After analyzing the implementation we find 1 382 samples/call acceptable. The block depends on the random number generator and would benefit from faster random number generation algorithms. The atomic block "SimOutObuVersions" cyclically writes the software, region, and tariff version of all OBUs to an output file. In our scenario we simulate 50 weeks and write data every 30 minutes, resulting in 16801 calls. 89 M samples/call seems to be quite costly and offers room for improvement.

In summary the simulation of the scenario took 98 811 263 M calls. Of these, the model components consumed 84,51% and the simulation kernel (logical processor) 15,49%.

In the second step we apply the Intel VTune [16] profiler, which operates at functional level resp. Code Layer (see table III). The external profiler catches the activities of both the simulation kernel and the simulation model (denoted as "K" or "M" in tab. III).

Most of the CPU time is consumed by kernel functions responsible for data transport. These functions are grouped by component (resp. namespace msa.sim.core, denoted as "K" in the first column of tab. III), as Port.send, EventManager.enqueueEvent, LogicalProccesor.mainLoopFast, and EventManager.dequeueEvent. In sum the functions consume 61,1% of the CPU time. Conspicuous is the relative high last level cache miss rate of function EventManager.enqueueEvent with 3,2% and the needed instructions per call of function LogicalProcessor.mainLoopFast with 2 379. However, the

TABLE III
VTUNE PROFILING RESULTS FOR SIMULATION KERNEL (K) AND MODEL (M): TOP 10 FUNCTIONS BY RUNTIME

Shown are the CPU instructions retired (IR), estimated call count (eCC), instructions per call (IPC) and last level cache miss rate (MR).

	Function	Time [%]	IR [G]	eCC [M]	IPC	MR [%]
K	Port.send	9,0	44	689	65	0,4
K	EventManager.enqueueEvent	7,7	21	92	237	3,2
K	LogicalProcessor.mainLoopFast	7,0	17	7	2 379	0,3
K	EventManager.dequeueEvent	6,3	104	2 517	41	1,1
K	big._mul<unsigned int>	5,0	103	2 611	40	0,3
M	StaHandling.Dice	4,8	12	11	1 097	0,1
K	EventManager.scheduleEvent	3,5	53	1 286	42	0,2
K	Any.extractToken	3,0	70	1 805	39	1,7
K	Pin.popFrontToken	2,8	49	1 234	40	0,2
K	EventManager.bucketOf	2,7	17	327	55	0,0

number of calls depends on the dispatch of data within atomic model components, which are grouped in form of user libraries. In our model we have two user libraries: GPRSSimulation (GPRSSimulation.Components.Atomics, denoted as "M" in the first column of tab. III) and Standard (msa.Standard.Control). The latter is a support library included in MSArchitect. Combined they are responsible for 20,1% of CPU time consumption. Performance critical and starting point for improvement is function StaHandling.Dice with 1 097 instructions per call and a CPU time consumption of 4,80%.

Both, kernel logging and profiling showed that most of the resources are utilized by functions responsible for data input/output and functions responsible for transmission and processing of tolling information. Relating the resource utilization of model components to real-word applications we could recognize a weak correlation: The application-level performance is determined by the real-world behavior of the simulated system.

VI. SUMMARY

Extending [7] we have shown how to analyze the performance of DES simulations: Generic benchmark test-cases allow a simple and direct comparison of different simulation tools. Not surprisingly the tools differ vastly as to their time and memory consumption. However, the benchmark results cannot be transferred to the application domain: The workload generated by a given simulation model determines in large part its performance. Taking an existing simulation model of a large-scale technical system we performed an in-depth performance analysis for one simulation tool using both the performance analysis methods provided by the simulation kernel and an external profiler with access to the CPU hardware profiling support.

Both profilers immediately identify the same bottleneck: Reading the ASCII-formatted pre-calculated driving patterns from disk. Consequently the simulation model is now integrated with the scenario generator. This in turn will allow implementing an optimization algorithm to fit the driving patterns to the observed system behavior – a feature that we expect to drastically improve the accuracy of the simulation results for the short-term behavior [10].

The hardware profiler catches both the application-level methods as well as the atomics provided by the simulation kernel (with or without access to its source code). Looking e.g. at the cache miss rate we find some simulation kernel routines and several application-level methods with a considerable probability of needing access to the main memory. We take this as starting point for future improvements.

MSArchitect, the simulation kernel used in the application benchmark, is currently extended to allow the automatic model reduction and (semi-) automatic parallelization of simulation runs. The single-core benchmark performed here will be the baseline to measure the improvements against.

REFERENCES

[1] E. A. Lee and D. G. Messerschmitt, "Static scheduling of synchronous data flow programs for digital signal processing," *IEEE Transactions on Computers*, vol. 100, no. 1, pp. 24–35, 1987.

[2] Andato GmbH & Co. KG, "MSArchitect," [accessed 12-May-2013]. [Online]. Available: http://www.andato.com/

[3] K. Lunde and L. Kieble, "Simulating communication within a satellite-based automated toll collection system," *Proceedings of the 55th International Scientific Colloquium*, pp. 318 – 323, 2010.

[4] K. Lunde, L. Kieble, and M.-A. Funk, "Prediction strategies in a service level granting prefetching cache for version-controlled gis data," *ISAST Transactions on Computers and Intelligent Systems*, vol. 2, no. 2, pp. 46–51, 2010.

[5] B. Pfitzinger, T. Baumann, and T. Jestädt, "Analysis and evaluation of the german toll system using a holistic executable specification," *45th Hawaii International Conference on System Sciences (HICSS)*, vol. 0, pp. 5632–5638, 2012.

[6] ——, "Network resource usage of the german toll system: Lessons from a realistic simulation model," in *46th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2013, pp. 5115–5122.

[7] T. Baumann, B. Pfitzinger, and T. Jestädt, "Simulation driven design of the German toll system – evaluation and enhancement of simulation performance," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*. IEEE, 2012, pp. 901–909.

[8] CEN , "ISO/TS 17575-1:2010 electronic fee collection - application interface definition for autonomous systems - part 1: Charging," 2010.

[9] Bundesministerium der Finanzen, "Sollbericht 2013," *Monatsbericht des BMF*, vol. 2, pp. 6–57, Feb. 2013. [Online]. Available: http://www.bundesfinanzministerium.de/Content/DE/Monatsberichte/2013/02/Downloads/monatsbericht_2013_02_deutsch.pdf?__blob=publicationFile&v=4

[10] B. Pfitzinger, T. Jestädt, and T. Baumann, "Simulating the German toll system: Choosing 'good enough' driving patterns," in *Proceedings of the mobil.TUM 2013 – International conference on mobility and transport*, Lehrstuhl für Verkehrstechnik, Ed. Technische Universität München, 2013.

[11] Bundesamt fùr Gùterverkehr, "Maut-Jahresstatistik 2011/2010," 2012, [accessed 10-March-2012]. [Online]. Available: http://www.bag.bund.de/SharedDocs/Downloads/DE/Statistik/Lkw-Maut/Jahrestab_11_10.pdf?__blob=publicationFile

[12] M. Dettmar, F. Rottinger, and T. Jestädt, "Achieving excellence in GNSS based tolling using the example of the german HGV tolling system," in *Proceedings of the 9th ITS Europe Congress*, Jun. 2013.

[13] Y. Zhou and E. A. Lee, "Causality interfaces for actor networks," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 7, no. 3, p. 29, 2008.

[14] A. Pacholik, T. Baumann, W. Fengler, and D. Grüner, "Discrete event simulation performance – benchmarking simulators," in *International Simulation Multi-Conference (SummerSim)*, Genoa, Italy, 2012.

[15] G. S. Fishman, *Discrete-Event Simulation: Modeling, Programming and Analysis*. Berlin: Springer, 2001.

[16] Intel, "Intel VTune Amplifier," [accessed 12-May-2013]. [Online]. Available: http://software.intel.com/en-us/intel-vtune-amplifier-xe

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 927–930

Moving Trend Based Filters Design in Frequency Domain

Jan Tadeusz Duda
AGH University of Science
and Technology,
Al. A. Mickiewicza 30,
30-059 Krakow, Poland
Email: jdu@agh.edu.pl

Tomasz Pełech-Pilichowski
AGH University of Science
and Technology,
Al. A. Mickiewicza 30,
30-059 Krakow, Poland
Email: tomek@agh.edu.pl

***Abstract*—An original approach to digital moving trend based filters (MTF) design, based on Bode plots analysis is proposed, aimed at seasonal time series decomposition and prediction. A number of polynomials of different range are discussed to be used in the MTF as the LS approximation formula. The Bode plots of the MTF are shown, and the best filter is selected. Results of a seasonal time series decomposition and prediction with the best MTF is presented and compared to the classical MTF calculations (involving the linear LS approximation).**

I. Introduction

The nonstationary time series filtering with moving trends is the well known approach to a nonparametric long term trend extraction from the series, aimed at further processing of stationary residuals and the series prediction [1], [2]. The classical moving trend filter (MTF) is based on rolling approximation of the series with the least-square (LS) linear approximation in a moving window [1]. The window width affects the extracted trend smoothness and cyclic components separation effectiveness. However typically, it is adjusted by a trial method, to reach the appropriately smooth nonparametric trend. This paper shows that much better smoothing properties and cyclic component extraction may be reached by using in MTF a higher order polynomial approximations and by specification of the required filter properties in frequency domain. Hence, the MTF design is proposed by analysis of Bode plots [4] of a number of the filter variants. The MTFs designed in this way were successfully applied to analysis of hydrogeological data [5] and to financial time series prediction [6]. Smoothing and prediction of a step change and a cyclic signal with the studied MTF was shown to illustrate their properties.

II. Moving Trend Based Filters – Formal Basis And Properties

Nonstationary time series $y(t)$ may be viewed as the sum of an aperiodic trend function $f(t)$, a cyclic component $C(t)$ of time period T, and a higher frequency zero-average noise $z(t)$ [3], [7]:

$$y(t)=f(t)+C(t)+z(t) \tag{1}$$

The periodic component can be written in the form of the harmonic series [4]:

$$C(t)=\sum_{k=1}^{K} A_k \sin\left(\omega_T i_k (t-\tau_k)\right),\ \omega_T=\frac{2\pi}{T} \tag{2}$$

where i_k , k=1,..., K denote the set of harmonics indices of the consecutive components k=1,...K (e.g. i_k=1, 2, 10), A_k – the amplitude of i_k-th harmonic, τ_k is the delay of the k-th component.

The nonparametric trend $f(t)$ may be calculated for each time step t_n by a low-pass digital filter designed in such a way to remove the ω_T and higher frequency components from the original series $y(t)$. The cyclic component $C(t)$ can be extracted from the filtering residuals by the Least Square (LS) approximation with the regression model of the form (2), and then, the regression residuals $z(t)$ may be viewed as a high-frequency stochastic process and treated with ARMA approach [2] (if its homoscedasticity can be assumed) or with GARCH models in case of its heteroscedasticity [3].

One of the techniques recommended to calculate the nonparametric trend $f(t_n)$ is a rolling approximation of the series $y(t_n)$ with the LS linear approximation in a window containing M samples, and then averaging of the approximates $y_F(i,t_n)$ obtained for each t_n [1]. It is referred to as moving trend based smoothing/filtering, which may be further used to h-samples ahead prediction of the series main component $f(t_n+h)$ by its extrapolation with a h-samples increment $\Delta_h f$ averaged with harmonic weights [1]:

$$\begin{gathered} f(t_{n+h})=f(t_n)+\Delta_h f, \\ \Delta_h f=\sum_{i=1}^{n-h} C_i\left(f(t_{i+h})-f(t_i)\right) \\ C_0=0, C_i=C_{i-1}+\frac{1}{(n-h)(n-h-i+1)} \end{gathered} \tag{3}$$

Hereby we propose a generalization of the moving trend smoothing algorithm, by employing higher order approximating polynomials, with appropriately designed properties. Let us consider the polynomial of the form (4) in the time interval of M samples, with the time counted from $-M$+1 to 0:

978-83-60810-53-8/$25.00

$$y_F(t_i) \stackrel{def.}{=} b_0 + b_1 t_i + b_2 t_i^2 + b_3 t_i^3 + b_4 t_i^4, \quad t_i = \{-M+1, \ldots, -1, 0\} \tag{4}$$

The derivatives of y_{F0} at the interval end ($t_i = 0$) can be easily shaped by fixing selected coefficients b_k at zero values, which implies different profiles of the LS approximates y_F, as shown in figure 1.

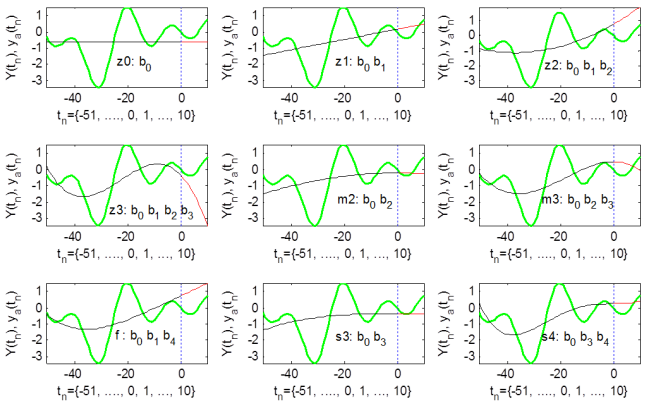


Fig 1. Properties of the approximating polynomials (4) considered to be used in the moving trend based filters; filter codes z0, z1, ..., s3, s4 used in sequel and corresponding nonzero coefficients are listed; vertical dotted line shows the interval end; shadow line: $y(t_n)=\sin(0.5\omega_T t_n)- \sin(\omega_T t_n)+ \sin(2\omega_T t_n)- \sin(3\omega_T t_n)$, $T=M=52$

In the moving trend algorithms the series splits into three sections. The first (starting *s*) section begins at the first (oldest) sample and finishes with the (*M*-1)-th one, the filtering window width enlarges from *M* to 2*M*-2, and the number L_n of the approximates $y_F(i,t_n)$ to be averaged increases from 1 to the *M*-1. The second (central *c*) section ranges from the *M*-th to *n*–*M*+1 samples, the filtering window width is 2*M*-1 (constant), and the number of approximates is *M*. The third (final *f*) section contains the samples from *n*–*M*+2 to *n* (the newest one), the window width reduces from 2*M*-2 to *M*, and the number of approximates $y_F(i,t_n)$ to be averaged decreases from *M*-1 to 1.

The calculations in the sections $\{s, c, f\}$ may be expressed in the FIR filtering form [3], [5]:

for $i = 1,\ldots,M-1$:

$$\begin{aligned} f(t_i) &= \sum_{k=1}^{i+M-1} g_s(i,k)\cdot y(t_{i+M-k}) \\ &= \sum_{k=1}^{n} G_s(k,i)\cdot y(t_{n-k+1}), \\ G_s(k,i) &\stackrel{def.}{=} [g_s(i,k), 0_{(n-i-M+1)}]^T, \end{aligned} \tag{5}$$

for $i = M,\ldots,n-M+1$:

$$\begin{aligned} f(t_i) &= \sum_{k=1}^{2M-1} g_c(k)\cdot y(t_{i+M-k}) \\ &= \sum_{k=1}^{n} G_c(k,i)\cdot y(t_{n-k+1}), \\ G_c(k,i) &\stackrel{def.}{=} [0_{(i-M+1)}, g_c, 0_{(n-i-M)}]^T, \end{aligned} \tag{6}$$

and for the final section, $i = n-M+2,\ldots,n$; $j \stackrel{def.}{=} i-n+M-1 = 1,\ldots,M-1$:

$$\begin{aligned} f(t_i) &= \sum_{k=1}^{2M-j-1} g_f(j,k)\cdot y(t_{i+M-j-k}) \\ &= \sum_{k=1}^{n} G_f(k,i)\cdot y(t_{n-k+1}), \\ G_f(k,i) &\stackrel{def.}{=} [0_{(i-M)}, g_f(j,k)]^T \end{aligned} \tag{7}$$

where g_s, g_c, g_f denote the impulse response vectors of filters in the sections ***s***, ***c***, ***f***, written also as the columns G_s, G_c and G_f of the unified smoothing filter matrix $G_{nx(n-1)}$, and the proper filter vector $G(k,n)$.

Similarly, the prediction formula (3) may be written in the following convolution form:

$$\begin{aligned} f(t_{n+h}) &= \sum_{k=1}^{n} P_h(k)\cdot y(t_{n-k+1}), \\ P_h(k) &\stackrel{def.}{=} \\ G_f(k, M-1) &+ \sum_{i=1}^{n-h} C_i\cdot(G(k,i+h) - G(k,i)), \\ G &\stackrel{def.}{=} [G_s, G_c, G_f] \end{aligned} \tag{8}$$

Notice that P_h are strongly affected by properties of the proper (the worst) filter $G_f(k,M-1)=G(k,n)$.

By making the Fourier Transform of the filters g_s, g_c, g_f and P_h involving different approximating polynomial types {z0....s4} with different *M* (see fig. 1), one may examine their properties in frequency domain, and select a filtering variant (type, *M*) suitable for smoothing and/or prediction demands, usually related to ω_T viewed as the cut-off frequency of the designed low-pass filters. The Bode plots of the examined filters are shown in figures 2-6. We have stated that the approximation window width *M* affects directly g_s, g_f and P_h delays, but it is of almost no effect on a shape of all the filters gain. Hence *M* may be taken as the lowest value producing gains close to 1 for $\omega<\omega_T$, near zero for $\omega=\omega_T$ and close to 0 for $\omega>\omega_T$.

Figure 2 shows the central smoothing filters are much better than 1st order recursive ones.

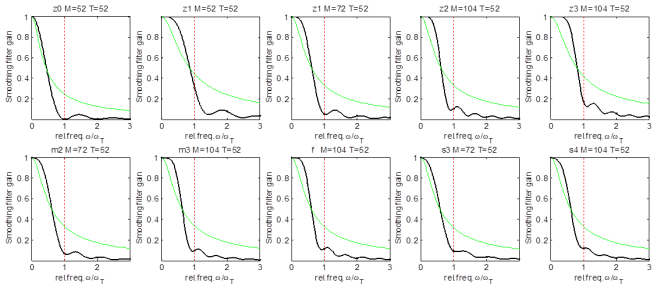


Fig 2. Gain diagrams (vs. ω/ω_T) for the best smoothing filter (central section); vertical point lines show ω_T, shadow solid lines – gain diagram for the 1st order recursive filter of the same half-gain frequency.

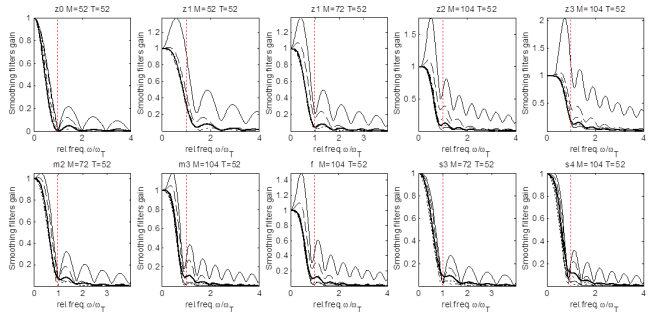


Fig 3. Gain diagrams (vs. ω/ω_T) for the smoothing filters: central section - bold lines $h<-M$, the final section for h=0 solid lines (proper filter), h=-20 dotted lines, h=-40 point-dotted lines.

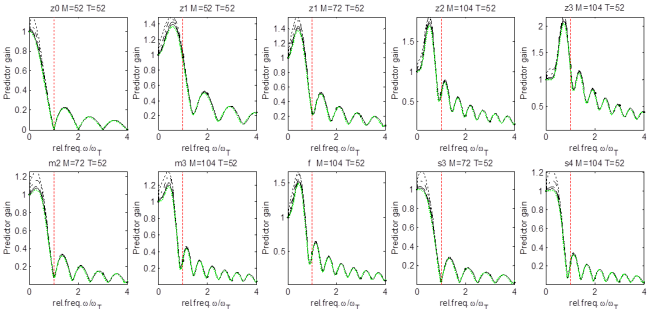


Fig 4. Gain diagrams for the moving trend based predictors: shadow bold line – final filter h=0; h=5 solid lines, h=10 dotted lines, h=26 point-dotted lines, h=T=52 point lines.

Gain properties of all the smoothing filters in the central section (fig. 2) are similar. When assuming M=1.38*T=72, the classical filter z1 seems to be the best due to the pass- and attenuation band properties as well as cut-off frequency gain, although z3 pass-band and attenuation of z0 look better. However a view on figures 3 and 4 gives evidence that only z0, s3 and s4 might be accepted from the perspective of final section smoothing (fig. 3) and prediction (fig. 4) properties. In particular, very bad pass and attenuation properties (excessive gain) of the classical filter z1 are clearly seen. Having in mind numerical problems (ill-conditioning) which can be met in s4 for larger M, one may take that the filter s3 with M=72 (1.38T) is the best choice (its pass-band is noticeably better than that of z0). The same conclusion may be drawn on a basis of delay properties shown in figures 5 and 6. In the pass-band a close to uniform and small delay is required (minimum delay distortion of the trend). It is satisfied only by z0 filters, but s4 delay distortion is acceptable and significantly lower than for the classical filter z1. The delay of predictors is larger than that of the final filter (h=0) by prediction horizon (see fig. 7). It means that the MTF prediction (eq. 3) does not differ essentially from Zero Order Hold of the $f(t_n)$.

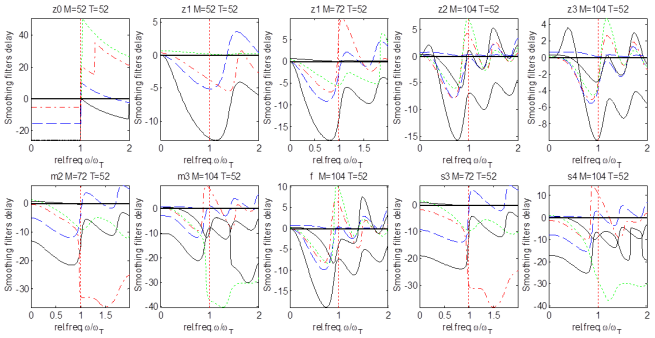


Fig 5. Delay of the final section smoothing filters: h=0 solid lines, h= -20 dotted lines, h=-40 point-dotted lines, h=-60 point lines, h=-70 solid lines close to the zero-delay, bold-line 0 delay of central section filter.

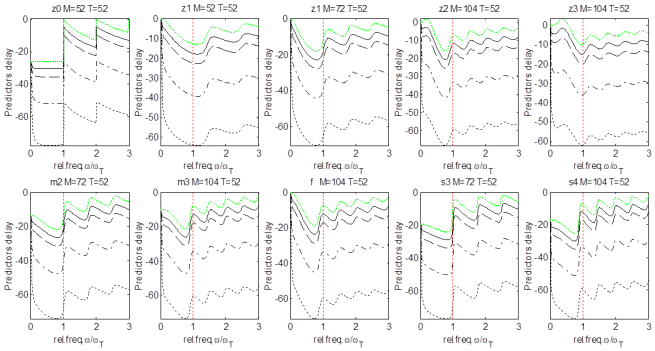


Fig 6. Predictors delay: h=0 shadow point lines, h= 5 solid lined, h=10 dotted lines, h=26 point-dotted lines, h=T=52 point lines.

The frequency properties presented above are visible in time domain responses – see figures 8, 9. Step change distortions shown in figure 8 are the larger, the greater irregularities of the pass-band gain and delay.

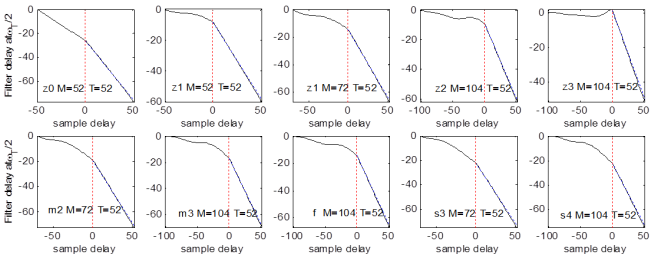


Fig 7. Delay of smoothing filters and predictors for $\omega_T/2$ versus the sample delay.

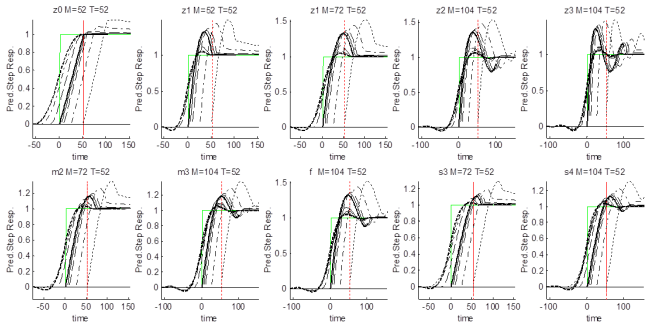


Fig 8. Signal step-wise change smoothing and prediction: shadow bold line – signal; bold dotted line – central segment filter response (h<-M); dotted point line smoothing with h=-40, h=-20 dotted line; bold line final filter response (h=0); prediction with h=5 solid lines, h=10 dotted lines, h=26 point-dotted lines, h=52 point lines.

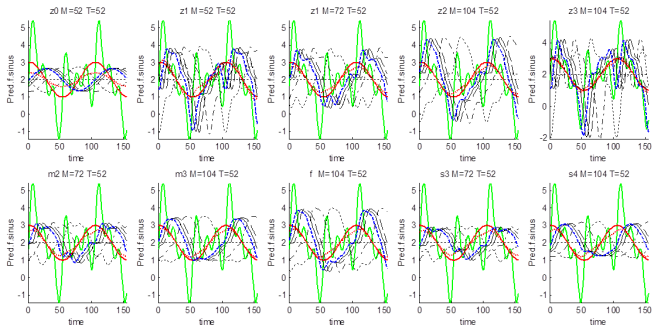


Fig 9. Periodic signal processing with the studied filters: shadow bold line – the signal $y(t_n)$ (see fig.1); bold line – the main harmonic of y, ω=0.5ω_Γ (to be extracted), bold point lines – the main harmonic reconstruction by the central filter response (h=-M), the main harmonic reconstruction with the final filter – bold dotted lines, and prediction with h=5 solid lines, h=10 dotted lines, h=26 point-dotted lines, h=52 point lines.

Figure 9 illustrates effects of filtering and prediction of a periodic signal (used also as the example in fig. 1). The harmonic of ω=ω_Γ/2 has to be extracted (reconstructed), but the signal contains strong components in the filters attenuation band. Hence the attenuation gain profile is of significant effect on the extracted signal shape. The best reconstruction is reached with z0 filters. The classical (z2) filters produce highly distorted responses, both in smoothing and prediction cases, while s3 and s4 yield acceptable results. All predictions are similar in shape to the proper filter response (h=0) and additionally delayed by the prediction horizon – see fig. 7 (i.e. they do not differ noticeably from ZOH predictions).

The extracted signal distortion in the starting and final sections is significant, hence separation of the filtering residuals into periodic $C(t)$ and stochastic $z(t)$ components, by fitting the regression model (2), should be performed with the central section data only. Then the periodic component $C(t)$ should extrapolated on the full data interval and subtracted from the filtering residuals to get $z(t)$.

III. Conclusion

The classical moving trend smoothing algorithm (based on linear approximates) is of low efficiency, when applied to series prediction. Much better smoothing and prediction properties may be reached by employing the 3.th order polynomial (s3) including only a constant and 3.th order monomial (only b_0, b_3 are to be tuned by LS method). The approximation window width M may be easily adjusted by examination the Gain Plots of the moving trend based filters in frequency domain. The recommended filter s3 enables for very effective separation of the series onto low frequency (ω<ω_Γ) and high frequency (ω≥ω_Γ) components, by taking the approximation window width M=1.38*T.

Smoothing (reconstruction of low frequency components) is the most effective (with no delay) in the central segment of the series. In the final section the low frequency signal distortion is significant, mainly due to varying delay of the consecutive final segment filters, which decreases prediction quality. The distortion produced by the recommended filter s3 is much weaker than that of the classical moving trend smoothing.

The periodic component $C(t)$ may be extracted from the filtering residuals by a regression method applied to residuals in the central section of the processed series.

References

[1] M. Cieślak (red.), *Prognozowanie gospodarcze. Metody i zastosowanie*, PWN Warszawa, 2002

[2] G.E.P. Box, G.M. Jenkins, G.C. Reinsnel, *Time Series Analysis, Forecasting and Control*. 3rd ed. Prentice Hall, Englewood Cliffs, NJ, 1994

[3] M.V. Askom, S. Chenouri, A.K. Mahmoodabadi: A*RCH and GARCH models*. Department of Statistics & Actuarial Sciences, University of Waterloo, 2001

[4] R.K. Otnes, L. Enochson, *Digital Time Series Analysis*, New York: John Wiley, 1972

[5] J.T. Duda, T. Pełech-Pilichowski, A. Augustynek, *Wykorzystanie trendu pełzającego do analizy i prognozowania szeregów finansowych*. [W:] Współczesne problemy zarządzania przedsiębiorstwami w gospodarce rynkowej (red. H. Howaniec, W. Waszkielewicz), Wyd. ATH, Bielsko-Biała, 2013 (in print)

[6] J.T. Duda, T. Pełech-Pilichowski., *Opracowywanie prognoz sytuacji hydrogeologicznej i ostrzeżeń przed niebezpiecznymi zjawiskami zachodzącymi w strefach zasilania lub poboru wód podziemnych*. Research Report, AGH UST, Faculty of Management, Kraków, 2012

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 931–938

Incorporating Text Analysis into Evolution of Social Groups in Blogosphere

Bogdan Gliwa, Anna Zygmunt, Stanisław Podgórski
Department of Computer Science
AGH University of Science and Technology
30-059 Kraków, Poland
{bgliwa,azygmunt}@agh.edu.pl, pstanisl@student.agh.edu.pl

***Abstract*—Data reflecting social and business relations has often form of network of connections between entities (called social network). In such network important and influential users can be identified as well as groups of strongly connected users. Finding such groups and observing their evolution becomes an increasingly important research problem. One of the significant problems is to develop method incorporating not only information about connections between entities but also information obtained from text written by the users. Method presented in this paper combine social network analysis and text mining in order to understand groups evolution.**

I. Introduction

NOWADAYS, elements of our everyday life move increasingly to the virtual reality: we write blogs or comment on someone else's posts, participate in discussions on forums, we exchange our opinions on fanpages of telecommunications companies and banks whose products we use. Everywhere we leave traces of our activity, which can be analyzed and ably combined with each other. Trading companies and banks might be interested in finding active or influential people in their environment and to offer them a new product in hope that it will be proposed by them to many others. Identification on time disgruntled people on banks or telecommunications companies fanpages will allow to respond quickly and prevent the spread of discontent.

The data about different types of dependencies can be modeled as a network of relationships and its structure can be analyzed using Social Network Analysis methods (e.g. for finding important nodes). Such a network, however, is not homogeneous and one can distinguish groups of people, for example, who more often exchange opinions. Such groups frequently are formed around important individuals and, for various reasons, the groups continue to exist or not, grow, shrink or can be joined with other groups. To understand causes of such events, which significantly affect the behavior of groups, it is important to include information that we can extract from the content of opinions or comments that are left behind. If the group talk about the same topics, does it affect its longer duration? Or, perhaps has a variety of discussed topics stronger impact on the duration of groups? How the themes of discussion are changing in a group? Is a small group with a strong leader more durable than a large one with few strong individuals?

Such knowledge derived from open sources can be combined, for example, with information about the history of bank transfers or loans as well as data about phone calls. Methods and algorithms proposed in the paper have been tested on one of the largest and highly dynamic polish blogosphere: salon24.pl, in which the main topic of discussion are political issues. However, they can be applied to other social media such as, for example, Twitter.

II. Related work

A. Blogosphere and Social network analysis

Internet social media (e.g. blogs, forums, media sharing systems, microblogging, social networking, wikis) has revolutionized the Internet and the way of communication between people. Among them, blogs play a special role in creating opinions and information propagation. Author gives opinions on some themes or describes interesting events and readers comment on these posts. Posts can be categorized by tags. A very important element of blogs is the possibility of adding comments, which allows discussions. Basic interactions between bloggers are writing comments in relation to posts or other comments. The relationships between bloggers are very dynamic and temporal: lifetime of a post is very short [1].

Based on blogs, posts and comments, we can build network, which can be analysed by Social Network Analysis (SNA) methods [2]. The SNA approach provides measures (SNA centrality measures) which make it possible to determine the most important or influential nodes (bloggers) in the network. Around such bloggers, the groups are forming, sharing similar interests.

B. Groups in social networks

Groups (or communities) are sets of nodes that are relatively densely connected to each other but sparsely connected to other nodes in network [3]. Many methods of finding groups exist in literature - one of the most popular ones is the CPM method (Clique Percolation Method) [4], which allows to extract overlapping groups i.e. groups that can have shared nodes with other groups.

Considering the dynamic nature of various social media, a growing interest in developing algorithms for extracting communities that take into account the dynamic aspect of the network has been observed.

978-83-60810-53-8/$25.00

A method of tracking groups over time was proposed in [5]. First, a division into time steps is carried out. At each step, the graph is created and groups are extracted. Groups from consecutive time steps are matched using the Jaccard index (value of this measure above predefined threshold means a continuation for analysed group). Palla et al. in [6] identified basic events that may occur in the life cycle of the group: growth, merging, birth, construction, splitting and death.

For further analysis, different characteristics, describing the communities and their transformation in time [7], are calculated, which concerns the comparison of the strength of internal relations of group members with their external connections with nodes outside the group, density of connections in the group or stability of the membership in time.

C. Methods of text mining

Text classification is one of major goals of Text Mining [8]. It involves extracting similar documents, inferring text topic and searching documents based on topic criteria.

Most text mining methods focus on text preprocessing (eg. stop words removal, words stemming and lemmatization) and converting input into structural representation [9]. Each word is represented as a separate entity assigned with a weight of the word importance, and thus it also allows to easily extract keywords. Algorithm TF-IDF (Term Frequency - Inverted Document Frequency) is one of the most popular weighting method [9]. It is based on the assumption that the importance of a word is proportional to number of occurrences of this word in a document, and inversely proportional to number of documents in which the word occurred. However, using only keywords to classify texts fails to find connection between semantically convergent documents that utilize different vocabularies, and more complex methods need to be applied such as Topic Modeling [10].

Topic Modeling [11] is a statistical technique that uncovers abstract "topics" that occur in a collection of documents. "Topic" is a set of words that tend to co-occur in multiple documents, and, therefore, they are assumed to have similar semantics. Main benefit of this model is that instead of using words from pattern to search for similar documents, words from topic are used, and therefore similar texts can be discovered even if they use different vocabulary.

Entirely different approach to uncovering documents semantics involves human input and it is called tagging [12]. Tags can be assigned either by author or by community in a process called crowdsourcing [13]. Number of tags assigned to a document may be large, and, therefore, it is imperative that a proper grouping and selection mechanism is implemented.

D. Text mining in the context of social network analysis

Existing research utilizing both SNA and Text Mining are mainly focused on very narrow cases. Aggarwal and Wang in [14] provided broad overview of text mining methods useful for social networks analysis. Tuulos and Tirri in [15] analysed IRC (Internet Relay Chat) communication network to discover and verify chat channels topics. Agrawal et al. in [16] used text mining methods to split social group into protagonists and antagonists. Caverlee and Webb in [17] used automatic classification methods based on keywords extraction and Topic Modeling to confirm personal information provided by Myspace users.

III. Analysis topics of groups and their impact on group behaviour

In this section we provide the concept of methods used to further analysis. The social network from whole data range is divided into series of time slots and each time slot contains static snapshot of network from defined period of time. In every time slot we extract groups and then find their dynamics in time. Irrespectively, we also discover topics in texts of comments and posts. Afterwards, we try to match topics for groups based on topics of comments and posts written by members of groups between themselves. Next, we analyse relations between topics of groups and behaviour of groups.

A. Groups in dynamic social network

Groups in each time slot were detected using the CPM [4] method (directed version of CPM from CFinder [1]). Groups from neighbouring time slots can be matched in order to find continuation of groups from different time. For this purpose, the SGCI (Stable Group Changes Identification) [18] method was employed. The algorihm consists of four main steps: identification of short-lived groups in each separated time slot; identification of group continuation (using modified Jaccard measure), separation of the stable groups (lasting for a certain time interval) and the identification of types of group changes (transition between the states of the stable group). The SGCI method identifies following event types:

- **split**, occurs when group divides into several groups in next time slot,
- **deletion**, similar to split, but it happens when small group detaches from significantly bigger one,
- **merge**, when several groups in the previous time slot join together and create larger group,
- **addition**, similar to merge, but it takes place when small group attaches to significantly bigger group,
- **split_merge**, when for the predecessor group the event is split and for the succesor group of given transition the event is merge in the same time,
- **decay**, the total disintegration of the group - the group does not exist in the next time slot,
- **constancy** means simple transition without significant change of the group size,
- **change_size** - simple transition with the change of the group size.

More detailed description of this method is provided in [18].

B. Finding topics of groups

For texts of posts and comments we employed methods of text mining in order to discover topics. Topics were extracted using 3 different methods:

[1] http://www.cfinder.org/

- TF-IDF keywords - words with the highest TF-IDF scores,
- Topic Modeling - topics extracted with LDA algorithm,
- Tags provided by post authors.

Keywords set for Topic Model is assumed to be a set of the most significant words for topics inferenced for messages.

We compared these methods between themselves using *similarity* measure:

$$similarity(S_1, S_2) = \frac{|S_1 \cup S_2|}{\min(|S_1|, |S_2|)}$$

where: S – keywords set, $|S|$ – number of elements in S.

For each group we can also assign set of topics discussed by its members. The topics are inferred based on posts and comments written by members of groups. We focused mostly on topic modelling as this method provides the highest level of abstraction from presented methods. Only topics that were present in more than 5% messages for groups were taken into consideration.

We defined *topic exploitation* for given topic and group as a ratio between number of group messages on certain topic and all messages for this group:

$$topicExploitation_k = \frac{|T_k|}{\sum_{i=1}^{n} |T_i|}$$

where: T_k – set of messages (posts and comments) for which topic with number k was inferenced, n – number of all topics, $|T_x|$ – amount of elements in T_x.

C. Topics changes in groups

To describe topics changes during transition between groups, we introduced following metrics:

- *Change in topic exploitation* for m-th group after transition t from time slot n to $n+1$ is calculated as:

  $$c_{m,n,t} = \sum_i \sum_k [g_{m,n,i} - g_{k,n+1,i} \cdot f(m,n,k,t)]$$

  where: i is a number of topic, $g_{m,n,i}$ is the topic exploitation of i-th Topic for m-th group in n-th time slot, f is function returning 1 if k-th group in slot $n+1$ is a continuation of m-th group from slot n and this transition has event type t.
- *Maximal positive change of single topic* (how much a topic gained) for m-th group after transition t from time slot n to $n+1$ is defined as:

  $$mpc_{m,n,t} = \max_i \sum_k [\epsilon(g_{m,n,i} - g_{k,n+1,i} \cdot f(m,n,k,t))]$$

  where: ϵ is a function returning the argument when the argument is **negative**, otherwise 0; other symbols were explained for *Change in topic exploitation* measure.
- *Maximal negative change of single topic* (how much a topic lost) for m-th group after transition from time slot n to $n+1$ was calculated as:

  $$mnc_{m,n,t} = \max_i \sum_k [\theta(g_{m,n,i} - g_{k,n+1,i} \cdot f(m,n,k,t))]$$

  where: θ is a function returning the argument when the argument is **positive**, otherwise 0; other symbols were explained for *Change in topic exploitation* measure.

Using above metrics we can analyse influence of different evolution types on topics change. Therefore, for each evolution type the average values of above defined measures for all groups are evaluated and we refer to them as *Average overall change in topic exploitation*, *Average maximal positive change of single topic* and *Average maximal negative change of single topic* respectively.

For above metrics, evolution events were taken into consideration only if there were at least 10 such events in selected time period.

D. Migrations of users depending on topics

To analyse difference in topics between given user and given group, we defined *topic divergence*, which has the following form:

$$m_t = t_{group} - t_{user} = \sum_{i=1}^{n} |(topic_{i,user} - topic_{i,group})|$$

where: n is a number of all topics in model (350), t_{group} is set of weights of each topic for given group, $topic_{i,group}$ – weight of i-th topic for given group, t_{user} is set of weights of each topic for given user, $topic_{i,user}$ is weight of i-th topic for given user.

It's worth noting that minimal value of m_t is 0.0 when user and a group has identical weight for every topic and maximal value is 2.0 when they are totally different. Maximum value of 2.0 is connected with the fact that group might cover topic X in 100% and user might cover topic Y in 100%, and therefore difference between group and user on topic X is 100% and on topic Y is also 100% which adds up to 200%.

Using this measure, we are trying to investigate relations between *topic divergence* and migrations of users (leaving and joining to groups). For this purpose the following measures are utilized:

- *Probability of leaving the group*. We assumed that potentially any member can leave the group. This value is calculated as:

  $$P_l(m) = \frac{|leavers_m \cap candidates_m|}{|candidates_m|}$$

  where: $leavers_m$ are users that in fact left any group and had the value of *topic divergence* measure equals m; $candidates_m$ are members of groups that have *topic divergence* = m.
- *Probability of joining the group*. When considering topic measure we assumed that candidates for joining are all users that were active in previous time slot. This value is calculated as:

  $$P_j(m) = \frac{|joiners_m \cap candidates_m|}{|candidates_m|}$$

  where: $joiners_m$ are users that in fact joined any group and had the value of *topic divergence* measure equals m;

$candidates_m$ – users active in previous time slot with *topic divergence* = m.

While calculating joiners and leavers sets we considered all group continuations to be a single group. The reason for that is to prevent *deletion* event to distort results - if a group splits into multiple small groups and we are assuming that anyone from the group can leave, then we will get very high accuracy from each event when huge group changes into small group.

It is worth noting that only both values - probability and histogram with migrations can provide us with complete information. Probability alone strongly depends on test case - if only 1 user had measure value = X and this user migrates then probability of migration for measure=X will be 100%, even if 100 different users migrated but they all had measure value=Y just as rest 10000 users, and thus probability for measure value=Y will be 1%. Without histogram we could not tell if any of those cases are marginal.

Analogically histogram itself can tell us only for which value there are the most migrations.

IV. Description of experiments

A. Data set

The analysed data about blogs was retrieved from the portal www.salon24.pl, which is dedicated especially to political discussions, but also subjects from other domains may be brought up. The data consists of 26 722 users (11 084 of them have their own blog), 285 532 posts and 4 173 457 comments within the period 1.01.2008 - 31.03.2012. Presented results were conducted on whole dataset - from 1.01.2008 to 31.03.2012. The analyzed period was divided into time slots, each lasting 7 days and neighboring slots overlap each other by 1 days. In the examined period there are 259 time slots. In each slot we used the comments model, introduced in [19] - the users are nodes and relations between them are built in the following way: from user who wrote the comment to the user who was commented on or if the user whose comment was commented on is not explicitly referenced in the comment (by using @ and name of author of comment) the target of the relation is the author of post.

B. Number of groups

The number of communities, with given size, for different value of *k* (parameter for CPM algorithm) is presented in table I. The *k* parameter determines the minimum group size (e.g. *k* equals 3 means that groups should consist of 3 or more members). The larger value of *k*, the smaller size of the biggest group. As we can notice, small groups outnumber other ones for each *k*. Furthermore, for *k* equals 6 the quantity of groups is much lower than for other values of *k* parameter.

C. Evolution events

Table II contains number of different evolution events in dataset for different values of *k*. We can observe for *k* equal 4 or 5 that the most popular events are *addition* and *deletion*, but for *k* equal 6, the most frequent events are *merge* and *split* (events similar to *addition* and *deletion*). The reason is that

TABLE I
NUMBERS OF GROUPS WITH DEFINED SIZE.

size	k=4	k=5	k=6
< 5	1596	0	0
5 – 6	384	2372	0
6 – 7	207	632	584
7 – 8	113	255	149
8 – 9	88	139	86
9 – 10	50	63	39
10 – 50	289	332	199
50 – 100	25	54	30
100 – 200	59	96	6
> 200	172	17	0

for *k* equal 4 or 5, there is a lot of small groups and there are also very huge groups (which not happens for *k* equal 6). In further analysis, we focus on groups extracted for parameter *k* equal 5 from the CPM method.

TABLE II
NUMBERS OF EVOLUTION EVENTS.

type	k=4	k=5	k=6
change_size	699	470	195
constancy	257	100	42
merge	428	439	434
split	323	409	397
addition	1091	2070	197
deletion	1115	2040	188

D. Convergence of different message topic extraction methods

This experiment covered comparison of different messages topic inference methods: TF-IDF keywords, topic modelling and tags provided by users.

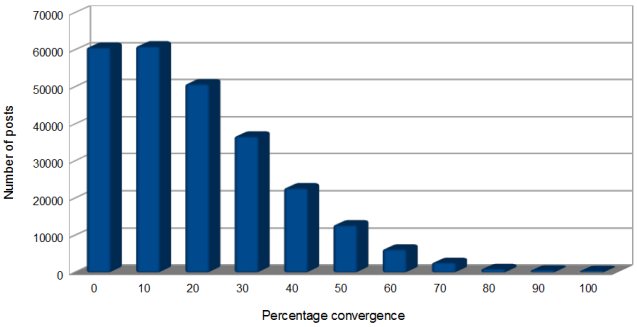


Fig. 1. Convergence level of TF-IDF keywords and most significant Topic Model words for inferenced topics

Fig. 1 presents comparison between TF-IDF keywords and Topic Modelling. As it can be seen, for about 20% of documents there was not even a single matching word. Convergence rate above 50% was achieved by merely 6000 posts which is around 5% of all input data.

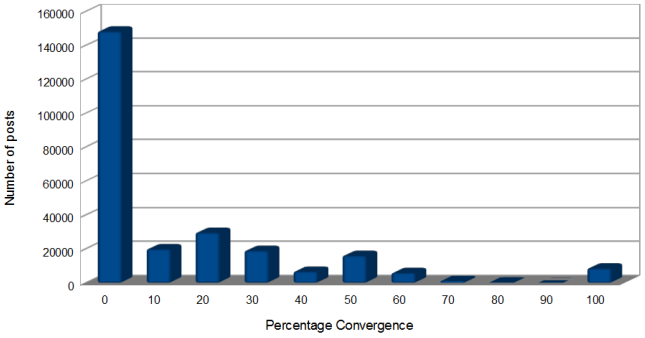


Fig. 2. Convergence level of TF-IDF keywords and user provided Tags

In fig. 2 TF-IDF keywords are compared with user tags. One can see that for huge part of documents achieved convergence rate was 0%. There are also local maxima at 20, 30, 50 and 100% and their origin is connected with number of tags user provides, which in most cases is between 1 and 5 (when matched 1/5, 1/4, 1/3, 1/2 and 1/1 of keywords).

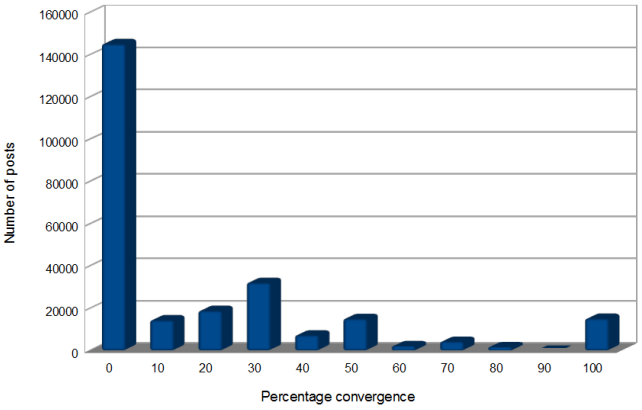


Fig. 3. Convergence level of user provided Tags and most significant Topic Model words for inferenced topics

Fig. 3 displays user tags in comparison with topic modelling. This histogram shows that, similarly to fig. 2, very huge part of tested posts don't have even a single word in common. We can also notice similar maxima due to the same reasons.

According to presented data, convergence rate seems to be very low, and therefore it would be imperative to check which of presented methods gives correct results. However, in depth analysis of our results uncovered that in fact all three methods are properly describing posts semantic. The problem is with vocabulary that is being used to describe it. The reason for different vocabulary is connected with level of generalisation that is utilised by presented methods:

- keywords provides very specific and detailed description of given document,
- tags provides general summary of the document and are far less specific, but error-prone of misspelling (they are provided by post authors),
- topics provides very general and abstract idea behind the document.

For these reasons, we are focusing more on topic modelling.

E. Topics coverage by groups

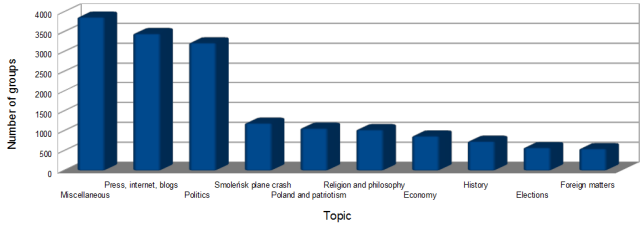


Fig. 4. Number of groups exploiting given topic

Figure 4 shows number of groups that exploit given topic in more than 5% of total group members messages. This figure presents ten most popular topics. We can notice that most popular topic is *Miscellaneous* which is very general and in fact is a mix of many themes.

F. Influence of group size on covered topics

In this experiment we aimed to check if there is a correlation between groups size and topics this groups covers.

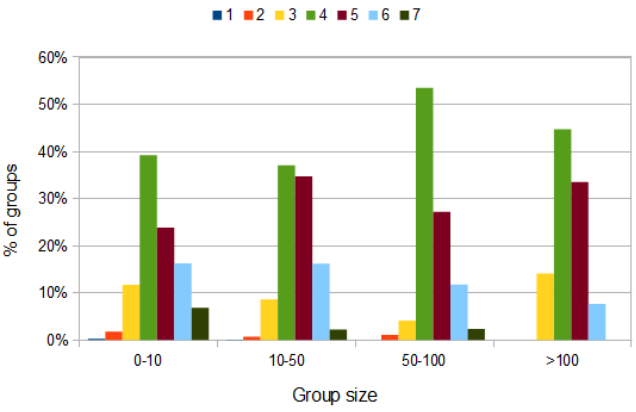


Fig. 5. Number of topics covered by groups with certain size, k=5

Fig. 5 shows number of Topics covered by groups with certain sizes. Considering 5% threshold of topic importance we used to remove noise. Maximum value any group could achieve was twenty topics, however no group covered more than seven topics and most groups cover four or five topics. In fig. 5 we can observe that three to seven topics are in groups of any size, but there are some small and medium-size groups that discuss only about one or two topics, which not happens in large groups.

Fig. 6 presents that for some topics *topic exploitation* is very similar regardless of group size (e.g. *Press, internet, blogging* topic), but there are some specific topics that are discussed to a greater extent among members in small group (e.g. topic related with science) or among members in larger ones (e.g. politics).

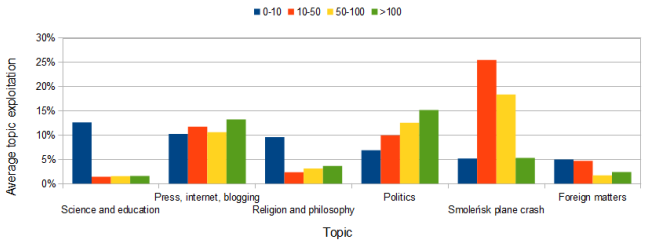


Fig. 6. Percentage topics exploitation for groups with certain size, k=5

G. Influence of duration time on covered topics

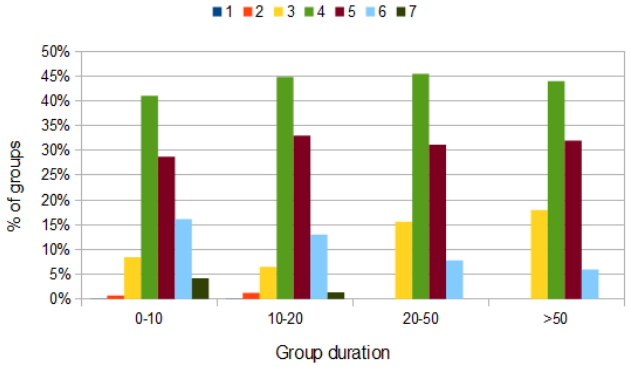


Fig. 7. Number of topics covered by groups with duration time, k=5

We can notice some regularities on fig. 7 :

- only very short living groups covers one and two topics,
- as previously, most groups cover four and five topics.

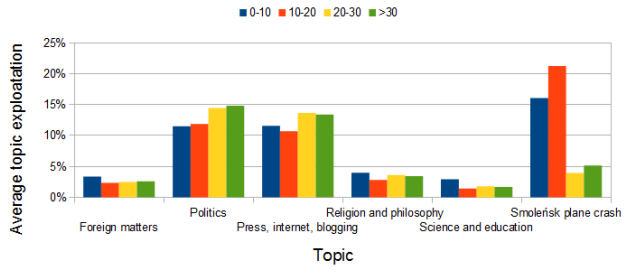


Fig. 8. Percentage topics exploitation for groups with certain duration time, k=5

In fig. 8 one can see that most topics achieved similar scores, with an exception - Polish Air Force Tu-154 crash was, to a large degree, discussed by short-term groups.

H. Connection between topics change and groups evolution

This experiment aimed to verify if there is any apparent connection between group evolution type and change in topics coverage for this group. Experiment was based on three measures (described earlier in section III-C):

- average overall change in topics coverage after evolution,
- average maximal positive change of single topic,
- average maximal negative change of single topic.

Evolution events were taken into consideration only if there were at least ten such events in selected time period. Presented results were collected for groups with $k = 5$ and for periods of length 360 days. There are two evolution types that are not present on the chart - *split_merge* that did not occur and *decay* that was omitted. *Decay* event means that group ceased to exist, and therefore we cannot calculate how topics of this group changed, because there is no continuation of the group.

Average overall change in topics. Figure 9 presents some regularities:

- *addition* event has clearly the highest overall topic change,
- *split* and *change_size* are connected with the lowest topic change,
- *merge* and *deletion* are in between,
- quite surprisingly *constancy* seems to vary between periods, even though one could expect that it will be connected with very small topic change.

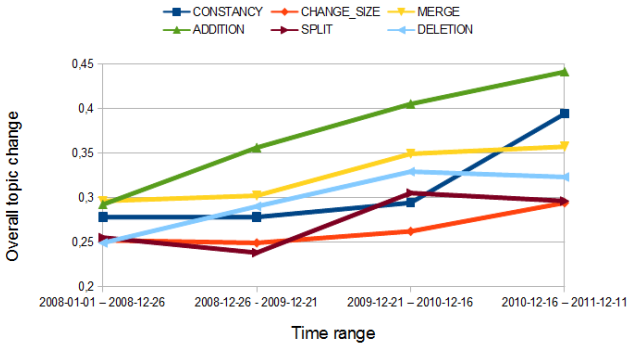


Fig. 9. Average overall change in topics coverage for every evolution event type - period 360 days

Average maximal positive change of single topic. Figure 10 presents that:

- *addition* has lowest average maximum single topic change. It means that on average after *addition* even topic that gained the most, gained very little.
- *deletion* and *split* caused highest positive change.
- *merge* and change_size were in between.

Average maximal negative change of single topic. Figure 11 shows that:

- *addition* is connected with highest drop for a single topic. It means that after *addition* there is a topic that significantly loses popularity.
- *change_size*, *split*, *deletion* and *merge* has very low average drop in topic popularity - even topics that lose popularity after such events lose very little.

Summary for different types of topic change. *Addition* is connected with: highest overall change in topics, highest negative change and lowest positive change. Therefore, we can deduce that when multiple small groups are forming a single

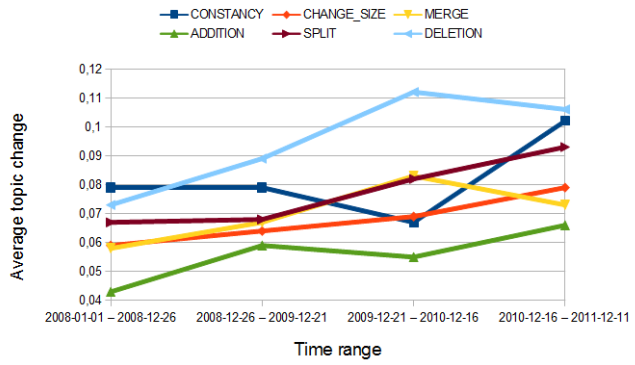


Fig. 10. Average maximal positive change of single topic - period 360 days

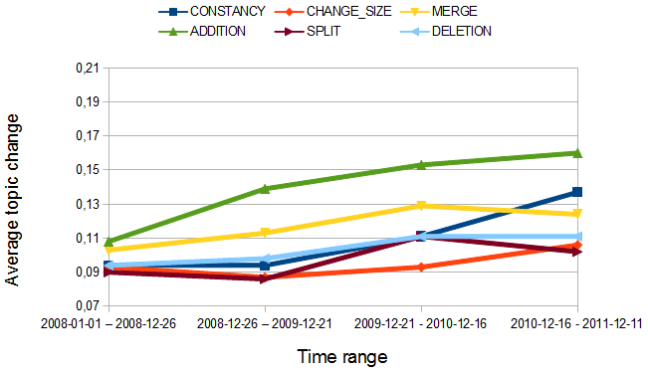


Fig. 11. Average maximal negative change of single topic - period 360 days

large group it is usually connected with significant drop of popularity of the main topic of each of the groups, and small rise in popularity of different topics - presumably of main topics of other groups.

Deletion and *split* cause small overall topic change, small negative change and large positive change. It means that splitting a group causes a rise of popularity of a single topic at the expense of all the others.

Change_size has very small changes in topics. We could expect that *constancy* behave the same way, however it does not.

Merge event causes medium rise of single topic popularity at the expense of all the others, meaning that joining groups are very similar and after join the leading topic emerges.

I. Migrations between groups

This experiment was conducted to check if migration of users between groups could be predicted using information about topic preferences of users and groups.

Important: currently analysed user does not have to be a member of the group. In fact, when calculating probability of joining a group we consider only users that are not yet members.

As a measure connected with users and groups topics we used *topic divergence* between user and group (details in section III-D).

Based on this measure, we tested their influence on:

- probability of leaving the group (candidates for leaving group are all members of group),
- probability of joining the group (candidates for joining are all users that were active in the previous time slot).

While calculating joiners and leavers sets we considered all group continuations to be a single group. The reason for that is to prevent *deletion* event to distort results - if a group splits into multiple small groups and we are assuming that anyone from the group can leave, then we will get very high accuracy from each event when huge group changes into many small groups.

Joining groups. Figure 12 shows that there is high probability of joining for users that are high convergent with the group, however, when we look at figure 13, we can notice that some of them (most convergent users with groups) are marginal cases (very few migrations). Moreover, we can notice that probability of joining groups is rather constant regardless the value of *topic divergence*, except the smallest values of this measure.

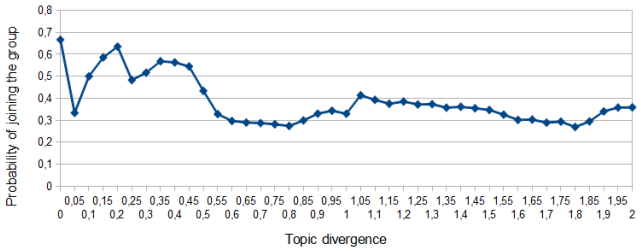


Fig. 12. Probability of user joining a group based on topic divergence.

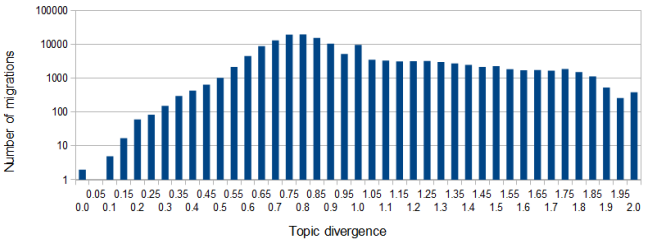


Fig. 13. Number of users that did join the group based on topic divergence.

Most migrations occur between 50% and 100% divergence and we can also see a rise in probability there, due to more cases. There is an interesting local extreme in figure 13 around 100% divergence. It would seem that there is a large portion of joiners that in previous time slot wrote all their posts and comments on one topic, or more likely wrote only a single post or comment, and then joined the group where this topic was not relevant. It is worth noting that our candidates were only a fraction of real joiners. It seems that about half joiners were people that were inactive in previous time slot (151 819 joiners were inactive user and 154 977 real joiners were from the candidate set).

Leaving groups. As can be deduced from figure 14, for low divergence values probability of leaving is low and is rising along with divergence up until 50% divergence and, further, the probability of leaving is rather constant. Alhough, it drops down for very high divergence, however, figure 15 tells us that it is a marginal case.

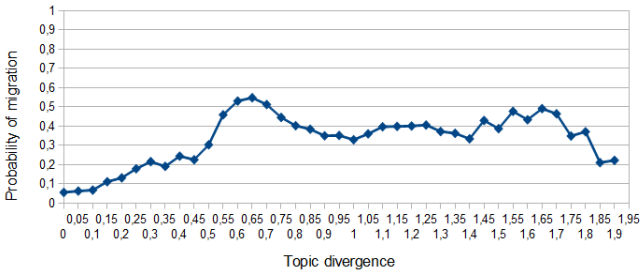


Fig. 14. Probability of user leaving a group based on topic divergence.

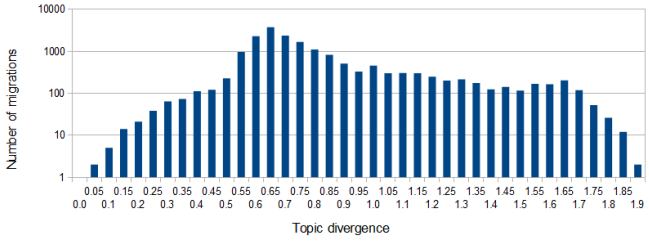


Fig. 15. Number of users that did leave the group based on topic divergence.

V. Conclusion

In this paper the analysis of topics for communities detected in real-world data from blogosphere is presented. We conducted experiments concerning relations between discussed topics by members of groups and some aspects of groups such as their duration time or their size. Furthermore, we also analysed influence of group evolution events on changes of topics and investigated the impact of topic divergence on users behaviour such as joining or leaving group.

Presented results seem promising and they reveal new insights into behaviour of groups and individuals. Analysis of topics discussed inside communities can be useful tool enabling better understanding of processes inside social network.

In future we are planning to use information about topics to improve our method of prediction of group behaviour [20]. Moreover, we intend to carry out similar experiments (related to analysis of topics in communities) on other datasets - we want to use also blogs in English language and datasets from different kinds of social media e.g. microblogs. Another interesting direction of further research is the analysis of key persons in groups in terms of topics they discuss and such analysis could lead to enhance our method of defining user roles and finding most influential people [21].

References

[1] N. Agarwal and H. Liu, *Modeling and Data Mining in Blogosphere*. Moegan & Claypool Publishers, 2009.

[2] P. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.

[3] S. Fortunato, "Community detection in graphs," in *Phys. Rep.*, 2010, ch. 486.

[4] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, pp. 814–818, 2005.

[5] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *Proc. International Conference on Advances in Social Networks Analysis and Mining (ASONAM'10)*. IEEE, 2010.

[6] G. Palla, A. lászló Barabási, T. Vicsek, and B. Hungary, "Quantifying social group evolution," *Nature*, vol. 446, p. 2007, 2007.

[7] J. Xu, B. Marshall, S. Kaza, and H. Chen, "Analyzing and visualizing criminal network dynamics: A case study," in *IEEE Conference on Intelligence and Security Informatics*, Tuczon, 2004.

[8] C. Aggarwal and C. Zhai, "A survey of text classification algorithms," in *Mining Text Data*, C. Aggarwal and C. Zhai, Eds. Springer, 2012, pp. 163–222.

[9] H. L. X. Hu, "Text analytics in social media," in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. Springer, 2012, pp. 385–414.

[10] S. Crain, K. Zhou, S. Yang, and H. Zha, "Dimensionality reduction and topic modelling: from latent semantic indexing to latent dirichlet allocation and beyond," in *Mining Text Data*, C. Aggarwal and C. Zhai, Eds. Springer, 2012, pp. 129–162.

[11] Y. Huang, "Support vector machines for text categorization based on latent semantic indexing," Electrical and Computer Engineering Department, The Johns Hopkins University, Tech. Rep., 2003.

[12] Y. Hassan-Montero and V. Herrero-Solana, "Improving tag-clouds as visual information retrieval interfaces," in *Proceedings of the InScit2006, International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.

[13] C. M. R. McCreadie and I. Ounis, "Crowdsourcing a news query classification dataset," in *Proceedings of CSE'10*, Geneva, Switzerland, 2010.

[14] H. W. C. Aggarwal, "Text mining in social networks," in *Social Network Data Analytics*, C. Aggarwal, Ed. Springer, 2011, pp. 353–378.

[15] V. Tuulos and H. Tirri, "Combining topic models and social networks for chat data mining," in *WI '04 Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, Washington DC, USA, September 2004, pp. 206–213.

[16] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *Proceedings of the 12th international conference on World Wide Web*, Budapest, Hungary, May 2003, pp. 529–535.

[17] J. Caverlee and S. Webb, "A large-scale study of myspace: Observations and implications for online social networks," in *Proceedings of the 2nd International Conference on Weblogs and Social Media (AAAI)*, Seattle, USA, March 2008.

[18] B. Gliwa, S. Saganowski, A. Zygmunt, P. Bródka, P. Kazienko, and J. Kozlak, "Identification of group changes in blogosphere," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012 Istanbul, Turkey, 26-29 August 2012*. IEEE Computer Society, 2012.

[19] B. Gliwa, J. Kozlak, A. Zygmunt, and K. Cetnarowicz, "Models of social groups in blogosphere based on information about comment addressees and sentiments," in *Social Informatics - 4th Int. Conf., SocInfo, Lausanne, Switzerland*, ser. Lecture Notes in Computer Science, vol. 7710. Springer, 2012, pp. 475–488.

[20] B. Gliwa, P. Bródka, A. Zygmunt, S. Saganowski, P. Kazienko, and J. Kozlak, "Different approaches to community evolution prediction in blogosphere," in *ASONAM 2013: IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining: Niagara Falls, Turkey*, 2013, accepted for printing.

[21] B. Gliwa, A. Zygmunt, and J. Kozlak, "Analysis of roles and groups in blogosphere," in *CORES - 8th International Conference on Computer Recognition Systems*, ser. Advances in Intelligent and Soft Computing, vol. 226. Springer, 2013, pp. 299–308.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 939–946

Towards Rule-oriented Business Process Model Generation

Krzysztof Kluza and Grzegorz J. Nalepa
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {kluza,gjn}@agh.edu.pl

***Abstract*—Attribute-Relationship Diagrams (ARD) aim at capturing relations, especially dependency relation, between the specified attributes. This paper describes work-in-progress research concerning process and rules integration, which takes advantage of the ARD method and allows for generating executable models. The paper examines the possibility of generating the rule-oriented BPMN model and enriching process models with rules from the ARD diagram.**

***Index Terms*—BPMN, Business Processes, Business Rules**

I. Introduction

BUSINESS Process (BP) models constitute a graphical representation of processes in an organization. Such a process is composed of related tasks that produce a specific service or product for a particular customer [1]. When it comes to practical modeling, Business Process Model and Notation (BPMN) [2], [3] constitutes a standard for this purpose.

The current version of the BPMN notation allows for modeling many aspects of business; nonetheless, it is not suitable for modeling some aspects of the enterprise, especially decision rules or constraints [4]. Recently, the Business Rules (BR) approach has been proposed as a new way of capturing the functional requirements and modeling system logic in a designer-friendly fashion. Moreover, the BR approach is a solution originated from Rule-Based Systems, that are a mature and well established technology. As processes and rules are related to each other, BR are often, but not limited to, used for the specification of task logic in process models.

BR can be acquired based on some data using machine learning techniques [5] or generated from natural language specification [6]; however, they often have to be modeled manually based on the knowledge collected from the domain experts, as usually their knowledge is not written down anywhere. Similarly, processes are designed manually as well. However, the simplified process models can be generated using process mining tools [7] or acquired from natural language description using NLP techniques [8]. Other method of acquiring BPMN models is to transform the existing process models in other languages to the BPMN notation. From a researcher's point of view, this can be a challenge in the case that languages are of different paradigm or presents a different aspect of the system, e.g. UML use-case diagrams [9].

The paper is supported by the HiBuProBuRul Project funded from NCN (National Science Centre) resources for science (no. DEC-2011/03/N/ST6/00909).

In this paper, we present work-in-progress research which is a part of our research concerning process and rules integration [10], [11], [12]. We examine the possibility of generation of the rule-oriented BPMN model as well as the possibility of enriching BP processes with rules from the ARD diagram.

As the ARD method allows a domain expert for gradual identification of the properties of a system being designed, we argue that having the system properties identified and described in terms of attributes, it is possible to generate executable BPMN model with the corresponding BR tasks as well as enrich the BP model with such BR tasks. Such an approach would allow for generating business processes and rule schemas for logic task specification at the same time. The generated rule prototypes comply with the XTT rule representation [13], [14], [15], [16] from the Semantic Knowledge Engineering approach [17].

The paper is organized as follows. In Section II we present the motivation for our research. Section III provides a short overview of the related approaches. Section IV presents the details of the ARD method. In Section V, we give an overview of the proposed method for process model generation and enriching the BP model with BR tasks based on a design example. The hybrid execution environment is presented in Section VI. Section VII summarizes the paper.

II. Motivation

The complexity of software has been constantly increasing for the last decades. To deal with this growth, new design methods and advanced modeling solutions are required [18]. For this purpose, modern applications use business processes and business rules as business logic specification [19].

According to the BPMN 2.0 specification [2], the notation is not suitable for modeling such concepts as rules. Therefore, this reveals the challenges in modeling and executing processes with rules. It is so because processes and rules are developed or modeled separately and are not well matched.

Our research aims at developing the integrated method for modeling BP with BR to provide a consistent method for modeling business systems. Such a method will allow for modeling processes with rules in a straightforward way, and then for executing such a developed model.

The main contribution of this paper is a presentation of the possibility of generation of the rule-oriented BPMN model, which can be used for enriching the BP models with BR tasks.

978-83-60810-53-8/$25.00

III. Related work

Several approaches can be considered as related to the method presented in this paper. As our method proposes automatic generation of a BPMN model, the approach can be compared with such approaches as: process mining [7], generating processes from text in natural language (based on NLP methods) [8], or finally transforming process from other notations to BPMN, especially from the notations that are not process-oriented, e.g. the UML use case diagrams [20].

The process mining methods [7] allow for very flexible process models generation, and in some cases this technique does not require any human activity. However, the result of the method is a very general process that is not suitable for direct execution. In order to be an executable BPMN model, it has to be significantly enhanced and refactored. In the case of our method, it is not as much flexible as process mining technique, but it produces a BPMN model which is executable and provides support for Business Rule tasks.

Generating processes from text description provided in natural language [8] can have practical results and allows for generating a high quality BPMN model. High quality models can also be obtained through translation from other representations, such as the UML use case diagrams [21], [22]. Unfortunately, a method based on the natural language description has to be supported by an advanced NLP system, thus practical applications of this method is very complex. Translation from other representations, in turn, requires process models designed using such representations, which often do not exist. In our approach, a process model is generated based on the carefully prepared ARD diagram. Although this requires the ARD diagram, it is very simple model and in some cases it can be obtained from text description using some mining technique to acquire attributes. This requires additional research yet. However, there has been trials of mining such attributes from text in natural language [23].

There are also other approaches, such as generating business process models from Bill Of Materials (BOM) [24], Product Based Workflow Design (PBWD) [25], [26], based on Product Data Model (PDM), or Decision Dependency Design (D3) [27], [28]. These are more similar to the method proposed in this paper. However, in the case of our approach, apart from the fact that it generates an executable BPMN model, it supports the rule prototypes generation for Business Rule tasks, what makes the BPMN model data consistent with the rule engine requirements for data. Therefore, we claim that our approach is partially rule-based [29].

It is important to mention that the presented approach can be used either for generating a new rule-oriented BPMN model or for enriching the existing process model with rules based on the corresponding ARD diagram.

Although the work presented in this paper is work-in-progress research, the method overview presented in the paper reveals significant differences from the techniques mentioned above, especially in the case of method simplicity and support for rules in process models.

IV. Attribute Relationship Diagrams

Attribute Relationship Diagram (ARD) [30] constitute a method which allows a user (especially a domain expert) for gradual identification of the system properties during design.

The goal of this method is to capture functional dependencies between attributes. The attributes are expressed in terms of Attributive Logic [31], [17], [32] and denote particular system properties identified by the domain expert. The identified dependencies form a directed graph in which properties are represented as nodes and dependencies are represented as transitions. In the following definitions, we present more formal description of ARD.

A typical atomic formula (or fact) takes the following form

$$a(p) = d$$

where a is an attribute, p is a property and d is the current value of a for p. More complex descriptions take usually the form of conjunctions of such atoms and are omnipresent in the AI literature [33].

An **attribute** $a_i \in A$ is a function (or partial function) of the form:

$$a_i \colon P \to \mathbb{D}_i$$

where

- P is a set of property symbols,
- A is a set of attribute names,
- $\mathbb{D}$ is a set of attribute values (the domains).

An example of an attribute can be the `carAge`, which denotes the age of a car, and the attribute value is within the domain $\mathbb{D}_{carAge} = [0, \inf]$.

A **generalized attribute** $a_j \in A$ is a function (or partial function) of the form:

$$a_j \colon P \to 2^{\mathbb{D}_j}$$

where $2^{\mathbb{D}_j}$ is the family of all the subsets of $\mathbb{D}_{\mathbb{j}}$.

An example of a generalized attribute can be the `ownedInsurances`, which is a set of the customer insurances, and the attribute value is a subset of the domain $\mathbb{D}_{ownedInsurances}$, which consists of the possible insurances that a particular customer can posses.

In the case of abstraction level, the ARD attributes and generalized attributes can be described either as conceptual or physical ones.

A **conceptual attribute** $c \in C$ is an attribute describing some general, abstract aspect of the system.
Conceptual attribute names are capitalized, e.g.: `BaseRate`. During the design process, conceptual attributes are being *finalized* into, possibly multiple, physical attributes.

A **physical attribute** $a \in A$ is an attribute describing a specific well-defined, atomic aspect of the system.
Names of physical attributes are not capitalized, e.g. `payment`. A physical attribute origins from one or more (indirectly) conceptual attributes and can not be further *finalized*.

A **simple property** $p_s \in P$ is a property described by a single attribute.

A **complex property** $p_c \in P$ is a property described by multiple attributes.

A **dependency** $d \in D$ is an ordered pair of properties (f, t), where $f \in P$ is the **independent property** and $t \in P$ is the **dependent property** that depends on f. For simplicity $d = (f, t) \in D$ will be presented as: $dep(f, t)$.

An **ARD diagram** R is a pair (P, D), where P is a set of properties, and D is a set of dependencies, and between two properties only a single dependency is allowed.

To illustrate the ARD concepts, an exemplary ARD diagram with properties and the dependency between them is presented in Figure 1. The diagram should be interpreted in the following way: `payment` depends somehow on `carCapacity` and `baseCharge`.

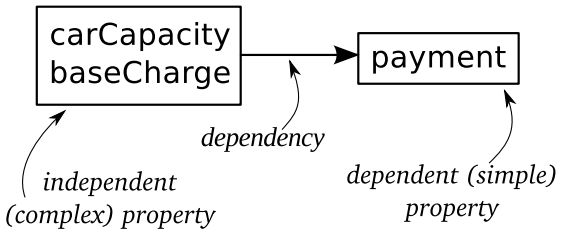


Figure 1. An example of the ARD diagram

The core aspects of the ARD method are diagram transformations, which regard properties and serve as a tool for diagram specification and development. Transformations are required to specify additional dependencies or introduce new attributes for the system. For the transformation of the diagram R_1 into the diagram R_2, the R_2 is more specific than the R_1.

Finalization *final* is a function of the form:

$$final : p_1 \rightarrow p_2$$

that transforms a simple property $p_1 \in P$ described by a conceptual attribute into a property $p_2 \in P$, where the attribute describing p_1 is substituted by one or more conceptual or physical attributes describing p_2, which are more detailed than the attribute describing a property p_1.

In Figure 2, an exemplary finalization transformation is presented. It shows that the simple property `BaseRate` (described by a single conceptual attribute) is finalized into a new complex property described by two physical attributes `carCapacity` and `baseCharge`.

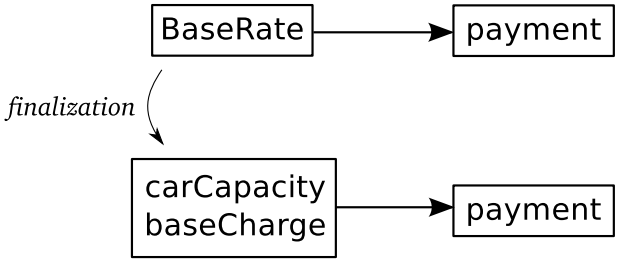


Figure 2. An example of the ARD finalization transformation

Split *split* is a function of the form:

$$split : p_c \rightarrow \{p^1, p^2, \ldots, p^n\}$$

where a complex property p_c is replaced by n properties, each of them described by one or more attributes originally describing p_c. Since p_c may depend on some other properties $p_o^1 \ldots p_o^n$, dependencies between these properties and $p^1 \ldots p^n$ have to be stated.

To illustrate this transformation, Figure 3 shows the complex property described by two physical attributes (`carCapacity` and `baseCharge`), which is split into two simple properties described by these attributes.

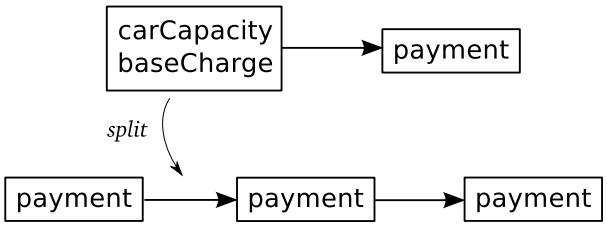


Figure 3. An example of the ARD split transformation

Upon splitting and finalization, the ARD model is more and more specific (see Figure 4). The consecutive levels of ARD forms a hierarchy of progressively detailed diagrams, which constitutes Transformation Process History (TPH) [34]. The implementation of this hierarchical model is provided through storing the lowest available, most detailed diagram level at any time, and additional information needed to recreate all of the higher levels. Such model captures information about changes made to properties at consecutive diagram levels.

A. Polish Liability Insurance Case Study

Let us now present an illustrative example of the Polish Liability Insurance (PLLI) case study. The example was developed as a benchmark case for the Semantic Knowledge Engineering (SKE) approach for rule-based systems [17]. Based on this simple case study, we will then present how ARD can be used for BPMN model generation.

In the PLLI case study, the price for the liability insurance for protecting against third party claims is to be calculated.

The price is calculated based on various reasons, which can be obtained from the domain expert. The main factors in calculating the liability insurance premium are data about the vehicle, such as the car engine capacity, the car age, etc. Additionally, the impact on the insurance price have such data as the driver's age, the period of holding the license, the number of accidents in the last year, etc. Moreover, in the calculation, the insurance premium can be increased or decreased because of number of payment installments, other insurances, continuity of insurance or the number of cars insured. All these pieces of data can be specified using the ARD method and presented using the ARD diagram (see Figure 5). As specification of ARD is an iterative process, the corresponding TPH diagram, presenting split and finalization transformations, can be easily depicted, as shown in Figure 6.

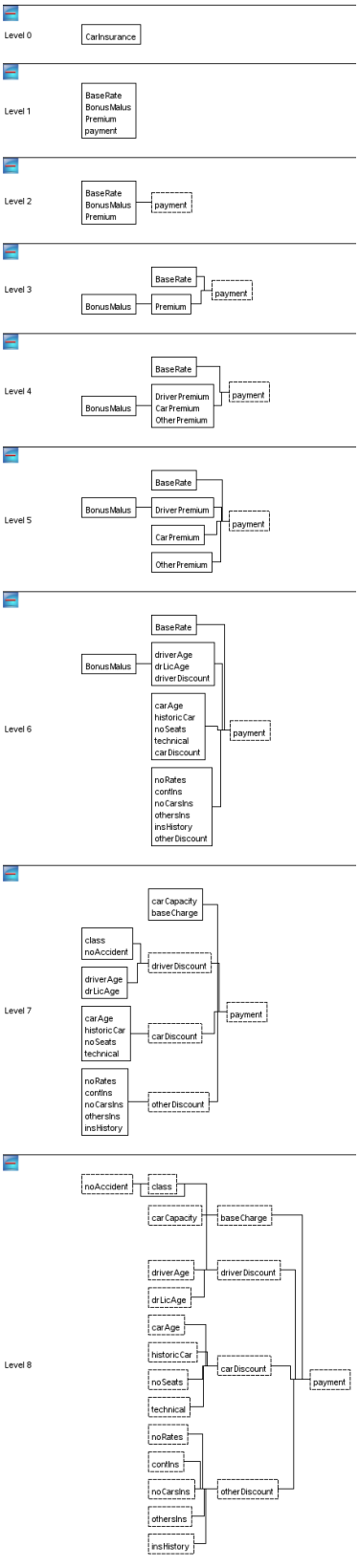


Figure 4. The ARD levels for the PLLI case study

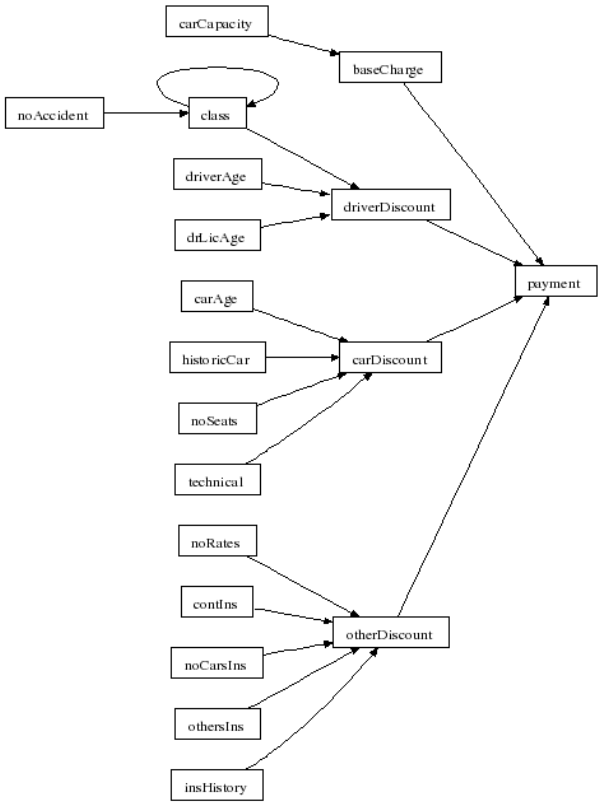


Figure 5. A complete ARD model for the PPLI example

B. Advantages of ARD

There are several advantages of using the ARD method to specify the system. Firstly, this method describes the system in an attribute-oriented way, and thus it is easy to comprehend and generates a simple model.

ARD can be used even if there are not many pieces of information available. It is so because this method does not require anything apart from the specification of dependencies between attributes. It is important to mention that we do not specify the detail semantics of the dependency relationship; thus it is only claimed that one property depends on other property. Although this limitation of ARD can be seen as a drawback, the main focus of this method is on simplicity.

The ARD method can also be extended, e.g. there can be used some mining technique to acquire attributes and dependencies among them. However, this requires additional research tasks yet. There has been trials of mining such attributes from text in natural language [23].

Applying ARD as a design process allows a domain expert to identify attributes of the modeled system and refine them gradually, as well as generates rule prototypes based on the identified attributes. Thanks to storing the history of transformations, it is possible to refactor such a system [35].

In the following section, we give a short overview of the proposed method of process model generation and enriching the BP model with BR tasks based on a design example.

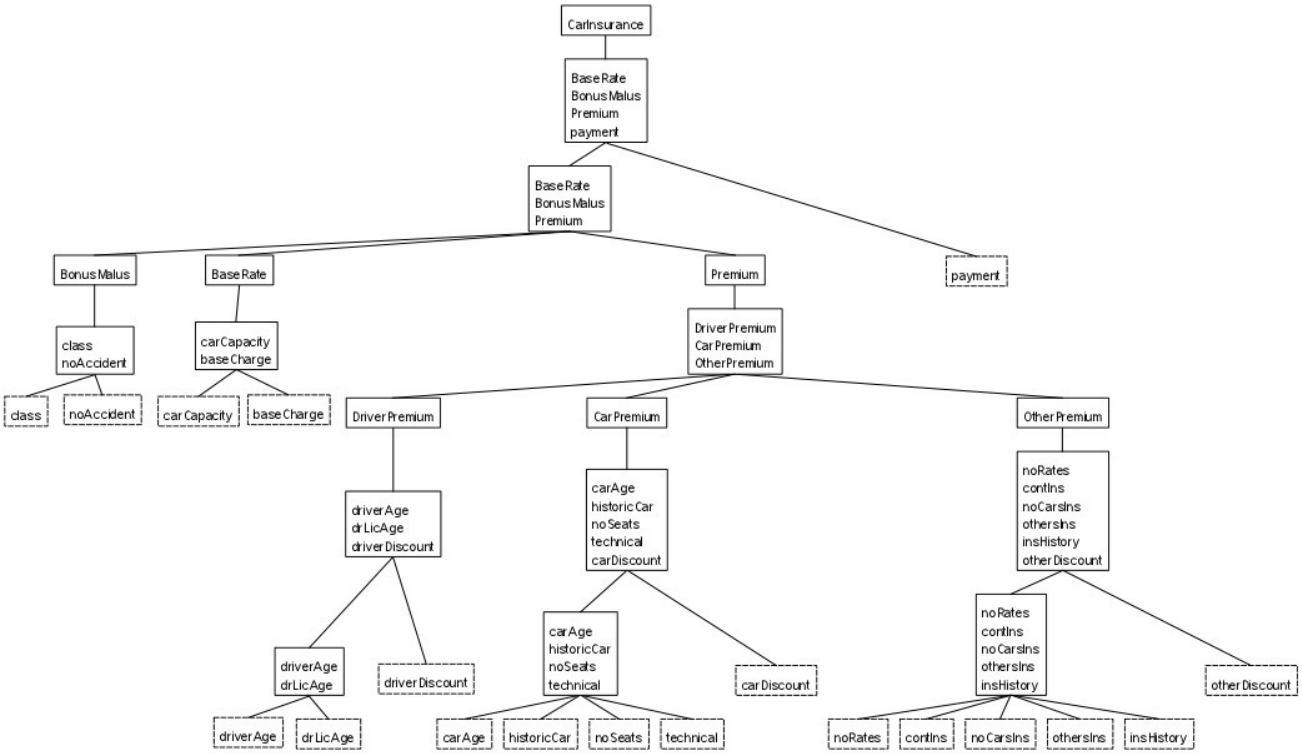


Figure 6. An example of the TPH diagram, corresponding to the ARD diagram presented in Figure 5

V. Rule-oriented BPMN Model Generation

In the proposed rule-oriented approach for BPMN model generation, we consider:

1) **Generating the whole BPMN model based on ARD**, presented in the previous section. Such an approach requires specific input and generates particular output:

   Input: Attribute-Relationship Diagram (ARD), and additionally Transformation Process History (TPH).
   The input is the most detailed ARD+ diagram, that has all of the physical attributes identified (in fact, this can also be applied to higher level diagrams, generating rules for some parts of the system being designed).

   Output: A rule-oriented BPMN process model.
   The output is a process model in the BPMN notation with User tasks, BR tasks and additional elements of the control flow objects. BR tasks contain rule prototypes in a very general format:

   ```
   rule:
   condition attributes | decision attributes
   ```

   Goal: The goal of this approach is to automatically build a rule-oriented BPMN process model on the basis of the ARD diagram (optionally supported by the TPH diagram). The algorithm will generate both User Tasks with form attributes for entering particular pieces of information and Business Rule Tasks with prototypes of decision tables.

   Sketch of the algorithm:

   1. Generate BR tasks from ARD based on the modified version of the algorithm for generating the XTT2 representation from ARD (detailed description of this part is presented below).
   2. Generate proper User tasks which acquire necessary information from the user.
   3. Generate proper User/Mail tasks to communicate process results to the user.
   4. Complete the diagram using control flow with additional flow objects, such as start and end events, and gateways.

2) **Enriching the existing BPMN model with BR tasks based on ARD** (either developed parallely to BP model or generated based on the process description).

   Input: BPMN process model, Attribute-Relationship Diagram (ARD) corresponding to the BPMN model, and additionally Transformation Process History (TPH).

   Output: A rule-oriented BPMN process model.

   Goal: The goal of this approach is to automatically enrich a BPMN process model with rule tasks on the basis of the ARD diagram (optionally supported by the TPH diagram). The algorithm will support refactoring of the process model to rule-oriented way by proposing new BR tasks for the process model.

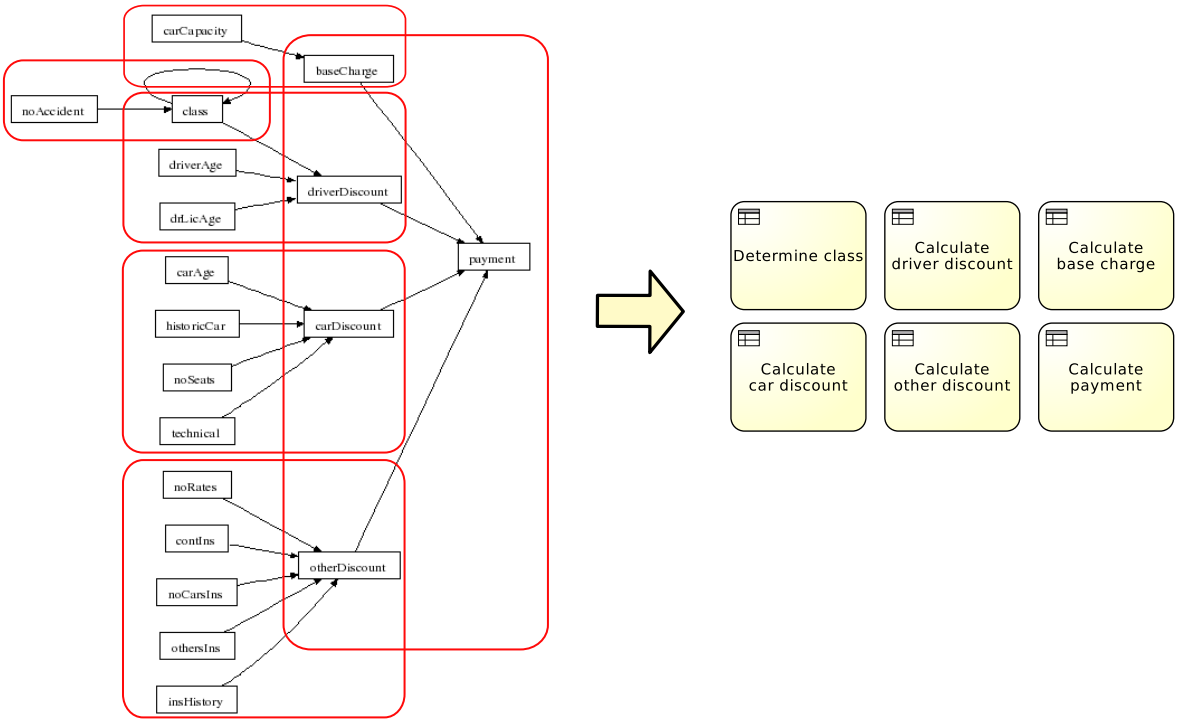


Figure 7. BR tasks generated from the ARD diagram

In the aforementioned cases, the most important aspect is to generate Business Rule tasks with rule prototypes for a process model. This can be done using the modified version of the algorithm for generating the XTT2 representation from ARD (described in [36], [34]). The result of application of this algorithm is presented in Figure 7. A draft of the algorithm for generating Business Rule tasks with rule prototypes for a process model is as follows:

1) Prepare data:
   a) Choose a dependency $d \in D : dep(f,t), f \neq t$, where D is a set of dependencies in the ARD diagram.
   b) Select all independent properties (other than f) that t depends on. Let $F_t = \{f_t^i : dep(f_t^i, t), f_t^i \neq f\}$. Remove the considered dependencies from the set: $D := D \setminus \{d_{f_t^i,t}\}$.
   c) Select all dependent properties (other than t) that depend only on f. Let $T_f = \{t_f^i : dep(f, t_f^i), t_f^i \neq t, \nexists f_x : (dep(f_x, t_f^i), f_x \neq f)\}$. Remove the considered dependencies from the set: $D := D \setminus \{d_{f,t_f^i}\}$.
2) Create BR tasks based on F_t and T_f:
   a) if $F_t = \emptyset, T_f = \emptyset$, create a BR task determining the value of the t attribute and associate the task with the following decision table schema: $f \mid t$.
   b) if $F_t \neq \emptyset, T_f = \emptyset$, create a BR task determining the value of the t attribute and associate it with the following decision table schema: f, f_t^1, f_t^2, ... $\mid t$.
   c) if $F_t = \emptyset, T_f \neq \emptyset$, create a BR task determining the value of the $T_f \cup \{t\}$) attributes and associate it with the decision table schema: $f \mid t$, t_f^1, t_f^2, ...
   d) if $F_t \neq \emptyset, T_f \neq \emptyset$, create two BR tasks determining the value of the T_f and t attributes and associate them with the following decision table schemas respectively: f, f_t^1, f_t^2, ... $\mid t$ and $f \mid t_f^1$, t_f^2, ...
3) Go to step 1 if there are any dependencies left ($D \neq \emptyset$).

The result of application of the BR task generation for the Polish Liability Insurance case is presented in Figure 7. Next, User tasks which acquire necessary information from the user and User/Mail tasks to communicate process results to the user have to be generated (see Figure 8 and 9).

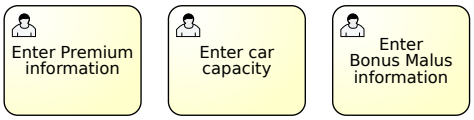


Figure 8. User tasks generated from the ARD diagram

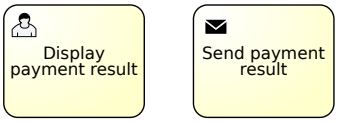


Figure 9. User/Mail tasks generated from the ARD diagram

Finally, the model have to be completed using control flow with additional flow objects, such as start and end events, and gateways. The resulting diagram can be observed in the Activiti-based environment presented in Figure 10.

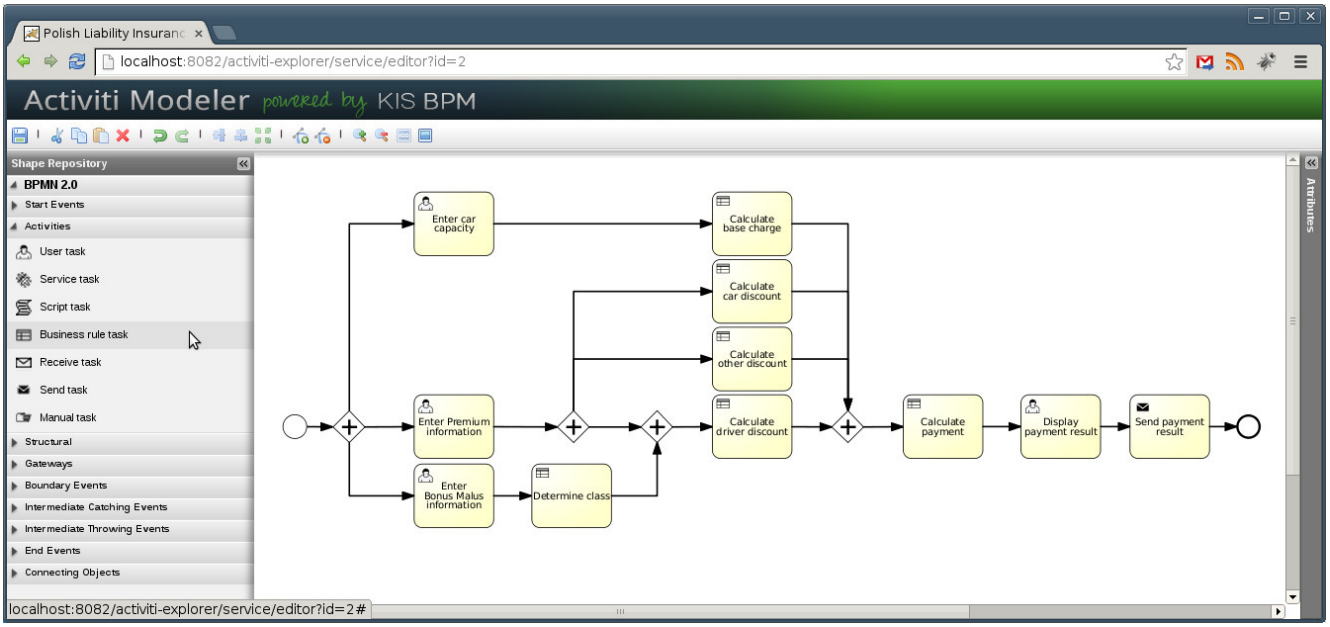


Figure 10. A prototype Activiti-based environment for modeling and executing processes with rules

VI. Hybrid Execution Environment

As a BPMN model generated from ARD constitutes an executable specification of a process, it can be executed in the process runtime environment. However, for complete execution of the model, i.e. execution of the Business Rule task logic, a process engine, such as jBPM [37] or Activiti [38], has to delegate rule execution to the business rule engine. As decision table schemas are generated automatically, the created decision tables have to be complemented with rules. Decision table can be filled in with rules using a dedicated editor [13] or a dedicated plugin for the process modeler [11]. Then, our prototype hybrid execution environment [12], [39], can serve as a specific execution example for this approach.

VII. Concluding Remarks

The aim of this paper is to examine the possibility of generating the rule-oriented BPMN model and enriching process models with rules based on the ARD diagram. We give an overview of the method for process model generation and present a first draft of the algorithm for automatic generation of rule-oriented BPMN process models from Attribute Relationship Diagram. In the algorithm, BR tasks with corresponding decision table schemas are generated and the resulting model can be executed in the hybrid execution environment.

The presented approach can be used either to generate the whole BPMN model based on the existing ARD diagram or to enrich the existing BPMN model with BR tasks based on ARD developed parallely to BP model or generated based on the process description. As the generated rule schemas are complementary to the process model, the solution addresses the two mentioned challenges: separation between processes and rules in the modeling phase and the problem of the execution of such separated data, which usually requires some additional integration or configuration in the execution environment.

As this paper presents a work-in-progress research, our future work will consist in refining and formalizing the presented approach. Then, we plan to extend the approach with new patterns and some optimization elements. We consider also enriching the ARD diagram with selected relations from the similar methods [24], [25], [27], [28], [26] and integrate the method with automatic verification features [40], [41].

References

[1] A. Lindsay, D. Dawns, and K. Lunn, "Business processes – attempts to find a definition," *Information and Software Technology*, vol. 45, no. 15, pp. 1015–1019, December 2003, elsevier.

[2] OMG, "Business Process Model and Notation (BPMN): Version 2.0 specification," Object Management Group, Tech. Rep. formal/2011-01-03, January 2011.

[3] T. Allweyer, *BPMN 2.0. Introduction to the Standard for Business Process Modeling*. Norderstedt: BoD, 2010.

[4] B. Silver, *BPMN Method and Style*. Cody-Cassidy Press, 2009.

[5] T. M. Mitchell, *Machine Learning*. MIT Press and The McGraw-Hill companies, Inc., 1997.

[6] I. S. Bajwa, M. G. Lee, and B. Bordbar, "SBVR Business Rules Generation from Natural Language Specification," in *AAAI Spring Symposium: AI for Business Agility*. AAAI, 2011. [Online]. Available: http://www.aaai.org/Library/Symposia/Spring/ss11-03.php

[7] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer Publishing Company, Incorporated, 2011.

[8] F. Friedrich, J. Mendling, and F. Puhlmann, "Process model generation from natural language text," in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, H. Mouratidis and C. Rolland, Eds. Springer Berlin Heidelberg, 2011, vol. 6741, pp. 482–496.

[9] A. Sinha and A. Paradkar, "Use cases to process specifications in business process modeling notation," in *Web Services (ICWS), 2010 IEEE International Conference on*, 2010, pp. 473–480.

[10] G. J. Nalepa, K. Kluza, and S. Ernst, "Modeling and analysis of business processes with business rules," in *Business Process Modeling: Software Engineering, Analysis and Applications*, ser. Business Issues, Competition and Entrepreneurship, J. Beckmann, Ed. Nova Science Publishers, 2011, pp. 135–156.

[11] K. Kluza, K. Kaczor, and G. J. Nalepa, "Enriching business processes with rules using the Oryx BPMN editor," in *Artificial Intelligence and Soft Computing: 11th International Conference, ICAISC 2012: Zakopane, Poland, April 29–May 3, 2012*, ser. Lecture Notes in Artificial Intelligence, L. Rutkowski and [et al.], Eds., vol. 7268. Springer, 2012, pp. 573–581. [Online]. Available: http://www.springerlink.com/content/u654r0m56882np77/

[12] G. J. Nalepa, K. Kluza, and K. Kaczor, "Proposal of an inference engine architecture for business rules and processes," in *Artificial Intelligence and Soft Computing: 12th International Conference, ICAISC 2013: Zakopane, Poland, June 9–13, 2013*, ser. Lecture Notes in Artificial Intelligence, L. Rutkowski and [et al.], Eds., vol. 7895. Springer, 2013, pp. 453–464. [Online]. Available: http://www.springer.com/computer/ai/book/978-3-642-38609-1

[13] G. J. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[14] A. Ligęza and G. J. Nalepa, "A study of methodological issues in design and development of rule-based systems: proposal of a new approach," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 117–137, 2011.

[15] M. Szpyrka, "Exclusion rule-based systems – case study," in *International Multiconference on Computer Science and Information Technology*, vol. 3, Wisła, Poland, October 20-22 2008, pp. 237–242.

[16] A. Ligęza and M. Szpyrka, "Reduction of tabular systems," in *Artificial Intelligence and Soft Computing - ICAISC 2004*, ser. Lecture Notes in Computer Science, L. Rutkowski, J. Siekmann, R. Tadeusiewicz, and L. Zadeh, Eds. Springer Berlin / Heidelberg, 2004, vol. 3070, pp. 903–908.

[17] G. J. Nalepa, *Semantic Knowledge Engineering. A Rule-Based Approach*. Kraków: Wydawnictwa AGH, 2011.

[18] G. J. Nalepa and K. Kluza, "UML representation for rule-based application models with XTT2-based business rules," *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, vol. 22, no. 4, pp. 485–524, 2012. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S021819401250012X

[19] G. J. Nalepa, "Proposal of business process and rules modeling with the XTT method," in *Symbolic and numeric algorithms for scientific computing, 2007. SYNASC Ninth international symposium. September 26–29*, V. Negru and et al., Eds., IEEE Computer Society. Los Alamitos, California ; Washington ; Tokyo: IEEE, CPS Conference Publishing Service, september 2007, pp. 500–506.

[20] OMG, "Unified Modeling Language (OMG UML) version 2.2. superstructure," Object Management Group, Tech. Rep. formal/2009-02-02, February 2009.

[21] J. R. Nawrocki, T. Nedza, M. Ochodek, and L. Olek, "Describing business processes with use cases," in *BIS*, 2006, pp. 13–27.

[22] D. Lubke, K. Schneider, and M. Weidlich, "Visualizing use case sets as bpmn processes," in *Requirements Engineering Visualization, 2008. REV '08.*, 2008, pp. 21–25.

[23] M. Atzmueller and G. J. Nalepa, "A textual subgroup mining approach for rapid ARD+ model capture," in *FLAIRS-22: Proceedings of the twenty-second international Florida Artificial Intelligence Research Society conference: 19–21 May 2009, Sanibel Island, Florida, USA*, H. C. Lane and H. W. Guesgen, Eds., FLAIRS. Menlo Park, California: AAAI Press, 2009, pp. 414–415, to be published.

[24] W. van der Aalst, "On the automatic generation of workflow processes based on product structures," *Computers in Industry*, vol. 39, no. 2, pp. 97–111, 1999.

[25] I. Vanderfeesten, H. Reijers, and W. Aalst, "Case handling systems as product based workflow design support," in *Enterprise Information Systems*, ser. Lecture Notes in Business Information Processing, J. Filipe, J. Cordeiro, and J. Cardoso, Eds. Springer Berlin Heidelberg, 2009, vol. 12, pp. 187–198.

[26] I. Vanderfeesten, H. Reijers, W. Aalst, and J. Vogelaar, "Automatic support for product based workflow design: Generation of process models from a product data model," in *On the Move to Meaningful Internet Systems: OTM 2010 Workshops*, ser. Lecture Notes in Computer Science, R. Meersman, T. Dillon, and P. Herrero, Eds. Springer Berlin Heidelberg, 2010, vol. 6428, pp. 665–674.

[27] F. Wu, L. Priscilla, M. Gao, F. Caron, W. Roover, and J. Vanthienen, "Modeling decision structures and dependencies," in *On the Move to Meaningful Internet Systems: OTM 2012 Workshops*, ser. Lecture Notes in Computer Science, P. Herrero, H. Panetto, R. Meersman, and T. Dillon, Eds. Springer Berlin Heidelberg, 2012, vol. 7567, pp. 525–533.

[28] W. Roover and J. Vanthienen, "On the relation between decision structures, tables and processes," in *On the Move to Meaningful Internet Systems: OTM 2011 Workshops*, ser. Lecture Notes in Computer Science, R. Meersman, T. Dillon, and P. Herrero, Eds. Springer Berlin Heidelberg, 2011, vol. 7046, pp. 591–598.

[29] S. Goedertier and J. Vanthienen, "Rule-based business process modeling and execution," in *In: Proceedings of the IEEE EDOC Workshop on Vocabularies Ontologies and Rules for The Enterprise (VORTE 2005). CTIT Workshop Proceeding Series (ISSN 0929-0672*, 2005, pp. 67–74.

[30] G. J. Nalepa and I. Wojnicki, "Towards formalization of ARD+ conceptual design and refinement method," in *FLAIRS-21: Proceedings of the twenty-first international Florida Artificial Intelligence Research Society conference: 15–17 May 2008, Coconut Grove, Florida, USA*, D. C. Wilson and H. C. Lane, Eds. Menlo Park, California: AAAI Press, 2008, pp. 353–358, accepted.

[31] A. Ligęza, *Logical Foundations for Rule-Based Systems*. Berlin, Heidelberg: Springer-Verlag, 2006.

[32] A. Ligęza and G. J. Nalepa, "Knowledge representation with granular attributive logic for XTT-based expert systems," in *FLAIRS-20: Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference: Key West, Florida, May 7-9, 2007*, D. C. Wilson, G. C. J. Sutcliffe, and FLAIRS, Eds., Florida Artificial Intelligence Research Society. Menlo Park, California: AAAI Press, may 2007, pp. 530–535.

[33] A. A. Hopgood, *Intelligent Systems for Engineers and Scientists*, 2nd ed. Boca Raton London New York Washington, D.C.: CRC Press, 2001.

[34] G. J. Nalepa and I. Wojnicki, "ARD+ a prototyping method for decision rules. method overview, tools, and the thermostat case study," AGH University of Science and Technology, Tech. Rep. CSLTR 01/2009, June 2009.

[35] G. J. Nalepa and I. Wojnicki, "VARDA rule design and visualization tool-chain," in *KI 2008: Advances in Artificial Intelligence: 31st Annual German Conference on AI, KI 2008: Kaiserslautern, Germany, September 23–26, 2008*, ser. Lecture Notes in Artificial Intelligence, A. R. Dengel and et al., Eds., vol. 5243. Berlin; Heidelberg: Springer Verlag, 2008, pp. 395–396, to be published.

[36] G. J. Nalepa and A. Ligęza, *Software engineering: evolution and emerging technologies*, ser. Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press, 2005, vol. 130, ch. Conceptual modelling and automated implementation of rule-based systems, pp. 330–340.

[37] *jBPM User Guide*, 5th ed., The jBPM team of JBoss Community, Dec 2011, online: http://docs.jboss.org/jbpm/v5.2/userguide/.

[38] T. Rademakers, T. Baeyens, and J. Barrez, *Activiti in Action: Executable Business Processes in BPMN 2.0*, ser. Manning Pubs Co Series. Manning Publications Company, 2012.

[39] K. Kaczor, K. Kluza, and G. J. Nalepa, "Towards rule interoperability: Design of Drools rule bases using the XTT2 method," *Transactions on Computational Collective Intelligence XI*, vol. 8065, pp. 155–175, 2013.

[40] K. Kluza, T. Maślanka, G. J. Nalepa, and A. Ligęza, "Proposal of representing BPMN diagrams with XTT2-based business rules," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 243–248. [Online]. Available: http://www.springerlink.com/content/d44n334p05772263/

[41] M. Szpyrka, G. J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 249–255. [Online]. Available: http://www.springerlink.com/content/m181144037q67271/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 947–949

The Set of Time Structures for Economic Phenomena Description

Maria Mach-Król
University of Economics
Dept. of Knowledge Engineering
Bogucicka 3, 40-287 Katowice,
POLAND
Email:
maria.mach-krol@ue.katowice.pl

***Abstract*—The paper describes some possible time structures for describing and analyzing economic phenomena. Since a simple linear time structure is not enough for this task, use of more complex structures are proposed.**

I. Introduction

INTRODUCING the notion of time allows to perform inference about changing domains, including the economic one. It also allows a computer to simulate human inference, because people infer about time and change [3]. In particular, such notions as change, causality or actions are described in the context of time, therefore the proper representation of time, and proper temporal reasoning is so important in the field of e.g. artificial intelligence [9, 10].

In artificial intelligence domain, which is concerned about knowledge, the temporal aspects of this knowledge are particularly important, because they are connected with both representation and inference. For an intelligent system to simulate intelligent behavior, to adapt to changes in the environment, or to verify its beliefs it has to be able to acquire new knowledge, and to maintain knowledge in an up-to-date form. Knowledge changes due to two main reasons. The first reason is the passing of time, the second one is new information on objects in the knowledge, having temporal characteristics [2, 11].

In order to represent properly temporal phenomena (or temporal knowledge about them) it is necessary to establish – prior to choosing the temporal formalism – a proper time structure. It is so because time structure determines the representation. For example, if one chooses dense time, it is not possible to represent knowledge in the situation calculus, because it is for only discrete phenomena [8]. In turn, the structure of time depends on the characteristics of the domain to be modeled. As Kania writes [5] p. 62, determining the structure of time is necessary to properly operate on time values, to operate on database etc.

The structure of time has to depend on a phenomena being investigated and should enable the best way to model it.

II. Time Structures to Describe Economic Phenomena – Basic Assumptions

We are convinced that for a temporal analysis of enterprise's environment it should be assumed that time is: discrete, branching into the future, finite in the past, but infinite in the future. We have chosen this structure because of several reasons:

a) Discrete time – there are several elements of the environment that change in a continuous manner, but some of the elements (e.g. barriers to entry) change discretely. From a practical point of view, it is not possible to provide information to the temporal intelligent system continuously. Changes have to be registered discretely. Moreover, assuming continuous time would be linked with introducing a second order axiomatization [1] p. 36.

b) Time branching into the future – the enterprise's environment is nondeterministic. Linear time assumes deterministic domain, while time branching into the future assumes a nondeterministic one. Also, introducing time branching into the future, when present actions may develop into several future ones, would allow to deepen the analysis of the environment, allowing e.g. for „what-if" analyses. It is not the only possible structure. For example, if we take into account the differences of temporal aspects of different markets we may think of a parallel time structure, which enables e.g. analyzing different markets simultaneously. Also a right linear time structure (time branching into the past) could be adopted – in order to determine, which changes in the past on the markets are responsible for the present situation of an enterprise. Using nonlinear time structures for analyzing economic environment is surely an interesting research area.

c) Time unbounded in the future – this assumption seems obvious: in a given moment, an enterprise is not able to define for how long it will be operating, therefore it is not possible to determine a moment in time, when the analysis will not be needed any more. However, as managerial practice shows, a time horizon longer than 5 years is not needed. For example, investment plans for more than 5 years are practically unreal. But it should be pointed out here, that a time point named „now" is different every day, moving into the future, and so moves the time horizon, even striving for infinity. Also obvious is assuming time bounded in the past: nor an enter-

978-83-60810-53-8/$25.00

prise, nor a temporal intelligent systems operate from „always". Moreover, an analysis getting far into the past would not be useful, because the environment is changing and turbulent. It should be assumed a certain „past time horizon of analysis", therefore bounding time in the past is justified.

Formally speaking, a time structure for modeling economic realm is a structure fulfilling the following conditions::

a) $\forall t_1, t_2 \in T$
   - $t_1 < t_2 \Rightarrow \exists t_3: (t_1 < t_3) \wedge \neg(\exists t_4: t_1 < t_4 \wedge t_4 < t_3)$,
   - $t_2 < t_1 \Rightarrow \exists t_3: (t_3 < t_1) \wedge \neg(\exists t_4: t_3 < t_4 \wedge t_4 < t_1)$,

b) $\forall t_1, t_2, t_3 \in T\ (t_2 < t_1 \wedge t_3 < t_1) \Rightarrow (t_2 < t_3) \vee (t_2 = t_3) \vee (t_3 < t_2)$

c) $\neg(\forall t_1 \exists t_2: t_2 < t_1)$.

It is a general assumption, but in specific situations the model may be broadened, because – as it has been already pointed out – the time structure has to be adjusted to the phenomenon being analyzed. In the next section different time structures are presented and discussed.

III. A SET OF TIME STRUCTURES

The most commonly adopted model is the one of linear time, which can be graphically depicted as a straight line, while formally a time structure T is linear, if [6] p. 20:

$$\forall t_1, t_2 \in T: (t_1 < t_2) \vee (t_1 = t_2) \vee (t_2 < t_1).$$

The models of nonlinear time are: time branching into the future, time branching into the past, time branching in both directions (parallel time), cyclic time. A motivation for adopting the branching time structure is as follows: many different pasts („routes") may have led to the present time, and from „now" may arise many different „routes" into the future. The formal definitions are: ([6] p. 21 and next):

A time structure T is branching into the future (left-linear), if

$\forall t_1, t_2, t_3 \in T\ (t_2 < t_1 \wedge t_3 < t_1) \Rightarrow (t_2 < t_3) \vee (t_2 = t_3) \vee (t_3 < t_2)$.

A time structure T is branching into the past (right-linear) if

$\forall t_1, t_2, t_3 \in T\ (t_1 < t_2 \wedge t_1 < t_3) \Rightarrow (t_2 < t_3) \vee (t_2 = t_3) \vee (t_3 < t_2)$.

A time structure T is parallel, if it is left- and right-linear, that is branching into both directions.

One more structure, discussed rarely in the literature, but interesting, is a cyclic time structure. A metric point time structure T is an ordered tuple $\langle T, C, <, <^*, \delta, S\rangle$, where: T – set of time points, C – set of distances between points, $<$ - a global order over T, $<^*$ - local order over T, δ – metrics over T, S – length of a semicircle.

For each time point $x \in T$ there exists exactly one point $x^* \in T$ that $\delta(x, x^*) = S$. These two points divide the time circle into two semicircles. The characteristics of a cyclic time structure are as follows (after [4], p. 30):

- completeness: $\forall x, y\ (x < y)$,
- local antisymmetry: $\forall x, y\ (x <^* y \rightarrow \neg(y <^* x))$,
- local linearity: $\forall x, y\ ((x \neq y\ \&\ x \neq y^*) \rightarrow (x <^* y \vee y <^* x))$,
- local transivity: $\forall x, y, z\ ((\delta(x, y) + \delta(y, z) < S) \rightarrow (x <^* y\ \&\ y <^* z \rightarrow x <^* z))$,
- coherence: $\forall x, y, z\ ((\delta(x, y) + \delta(y, z) < S) \rightarrow (x <^* y\ \&\ y <^* z \rightarrow \delta(x, y) + \delta(y, z) = \delta(x, z)))$.

One may imagine such situations, in which classic time structures would be not enough. Consider a situation, when two enterprises join (perform a fusion), operate as one enterprise for a certain period of time, and divide again into two enterprises, but cooperating together (so it is justified to analyze them together, but not on one time axis). In this case we deal subsequently with time structures: right-linear one, linear one, and left-linear one. Formally this situation may be written as:

- $(t_1, t_2, t_3 < t_F) \Rightarrow (\forall t_1, t_2, t_3 \in T: (t_1 < t_2 \wedge t_1 < t_3) \Rightarrow (t_2 < t_3) \vee (t_2 = t_3) \vee (t_3 < t_2))$,
- $((t_1, t_2 > t_F) \wedge (t_1, t_2 < t_P)) \Rightarrow (\forall t_1, t_2 \in T: (t_1 < t_2) \vee (t_1 = t_2) \vee (t_2 < t_1))$,
- $(t_1, t_2, t_3 > t_P) \Rightarrow (\forall t_1, t_2, t_3 \in T\ (t_2 < t_1 \wedge t_3 < t_1) \Rightarrow (t_2 < t_3) \vee (t_2 = t_3) \vee (t_3 < t_2))$,

where: t_F – the moment of fusion, t_P – the moment in which enterprise divides again.

This structure has the following properties:

- transivity: $\forall x, y\ (x < y\ \&\ y < z \rightarrow x < z)$,
- anti-reflexivity: $\forall x\ \neg(x < x)$,
- antisymmetry: $\forall x, y\ (x < y \rightarrow \neg(x < y))$,
- discreteness: $\forall x, y\ (x < y\ \rightarrow \exists z(x < z\ \&\ \neg\exists u(x < u\ \&\ u < z)))$

  $\forall x, y\ (x < y\ \rightarrow \exists z(z < y\ \&\ \neg\exists u(u < u\ \&\ u < y)))$.

Of course, different structures may be combined together in different ways, according to the analytical needs. It seems that the easiest is combining branching structures with linear one. The combinations similar to the one shown above may be numerous, one can imagine e.g. chains composed of branching and linear time structures. On the other hand, combining the cyclic time structure with branching and/or linear ones seems difficult, or even impossible, because the cyclic structure is a closed one.

It is necessary to discuss, how many time structures can arise from basic ones? Before we answer this question, we have to make some assumptions:

- for simplicity, we consider only a basic structure $\langle T, <\rangle$, other axioms of this structure are omitted;

- we assume representation in 1st order predicate calculus; in the calculus generally the linearity axiom is manipulated, therefore we deal with a finite set: linearity axiom, left-linearity axiom, right-linearity axiom;

- we omit the question of time metrics, because this does not affect the linear or non-linear property of time structure.

Having the above assumptions, we may say that the set of possible time structures arising from combining basic time structures is a combination with repetitions of those struc-

tures. As it is commonly known, the number of *k*-element combinations with repetitions of a *n*-element set is given by a formula:

$$C_n^k = \frac{(k+n-1)!}{k!\,(n-1)!}$$

Where

k – number of sequence elements,

n – number of set elements.

In the considered case, k = 1, 2, 3 or 4, and n = 4 (linear structure, left-linear structure, right-linear structure, parallel structure).

Therefore, the set of possible time structures is computed as:

$$S = 1 + \frac{(2+4-1)!}{2!\,(4-1)!} + \frac{(3+4-1)!}{3!\,(4-1)!} + \frac{(4+4-1)!}{4!\,(4-1)!} = 66$$

It should be noted here that in case of using multiple time structures for economic realm description and in case of using a temporal intelligent system, we have to deal with a heterogenic time structure in the knowledge base of the system. Thus, we face a problem of unifying the structure. Is it necessary for performing reasoning by an intelligent system? This problem is similar to the one of the heterogeneous knowledge in a temporal intelligent system, described in detail in [7], but it is beyond the scope of this paper.

IV. Concluding Remarks

In the paper the motivation for temporal representation of economic knowledge has been presented, the possible time structures have been pointed out, and the possible combinations of them for economic realm description have been shown. The main conclusion stresses out the variety of possibilities given by only 4 time structures, combined in different ways. Leaving linear time axiomatization, in order to take into account richer structures, will enable to better depict the economic realm, e.g. for building a knowledge base of a temporal intelligent system.

The problem of time structures heterogeneity in a temporal knowledge base arises while using more than one structure. It has to be discussed and checked, whether to perform reasoning in a temporal intelligent system it is necessary to unify these structures, and if so, how it should be done. This will be the topic of future research studies.

References

[1] Bennett B., Galton A. P., *A unifying semantics for time and events*. "Artificial Intelligence", Vol. 153, 2004, pp. 13-48.

[2] Benthem van, J., *Temporal Logic*, in: Gabbay D. M., Hogger C. J., Robinson J. A. (Eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4: *Epistemic and Temporal Reasoning*. Clarendon Press, Oxford 1995.

[3] Galton A., *Time and Change for AI*, in: Gabbay D. M., Hogger C. J., Robinson J. A. (Eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4: *Epistemic and Temporal Reasoning*. Clarendon Press, Oxford 1995.

[4] Hajnicz E., *Time Structures. Formal Description and Algorithmic Representation*. LNAI 1047, Springer, 1996.

[5] Kania K., *Temporalne bazy danych w systemach informatycznych zarządzania [Temporal databases in computer management systems]*. Prace Naukowe Akademii Ekonomicznej im. Karola Adamieckiego w Katowicach, Wydawnictwo AE Katowice, 2004.

[6] Klimek R., *Wprowadzenie do logiki temporalnej[Introduction to temporal logic]*. Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 1999.

[7] Mach M. A., *Temporalna analiza otoczenia przedsiębiorstwa. Techniki i narzędzia inteligentne [Temporal analysis of enterprise environment. Intelligent techniques and tools]*. Wydawnictwo AE Wrocław, 2007.

[8] McCarthy J., Hayes P., *Some philosophical problems from the standpoint of artificial intelligence*. "Machine Intelligence" Vol. 4, American Elsevier 1969, pp. 463-502.

[9] Vila L., *A Survey on Temporal Reasoning in Artificial Intelligence*. "AI Communications", 7(1), 1994, pp. 4-28.

[10] Saracco, Cynthia M., Matthias Nicola, and Lenisha Gandhi. "A Matter of Time: Temporal Data Management in DB2 for z/OS." (2012). ftp://170.225.15.26/software/data/sw-library/db2/papers/A_Matter_of_Time_-_DB2_zOS_Temporal_Tables_-_White_Paper_v1.4.1.pdf (access: May 10th, 2013).

[11] Caron, Filip, Jan Vanthienen, and Bart Baesens. "Rule-Based Business Process Mining: Applications for Management." *Management Intelligent Systems*. Springer Berlin Heidelberg, 2012. 273-282.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 951–958

Assessment of Business Intelligence Maturity in the Selected Organizations

Celina M. Olszak
University of Economics in Katowice
ul. Bogucicka 3b,
40-287 Katowice, Poland
Email: celina.olszak@ue.katowice.pl

Abstract—**The main purpose of this paper is to assess the level of Business Intelligence (BI) maturity in organizations. The research questions I ask in this study are: (1) what possibilities offer BI systems for different organizations, (2) how to measure and evaluate the BI maturity in organizations? The study was based on: (1) a critical analysis of literature, (2) a observation of different BI initiatives undertaken in various organizations, as well as on (3) semi-structured interviews conducted in polish organizations in 2012. Some interviews, conducted in 20 polish enterprises, were held with executives, senior members of staff, and ICT specialists. The reminder of my paper is organized as follows. Firstly, the idea of BI is described. Next, the issue of BI maturity models is recognized. Finally, Garter's Maturity Model for Business Intelligence and Performance Management is used to assess the level of BI in surveyed organizations.**

I. Introduction

ALTHOUGH Business Intelligence (BI) has been developing for over 20 years, many organizations are not able to make from it the effective tool for decision making or creating the competitive advantage [1], [2]. One of the reason of this fact is that they do not know a theoretical foundation how to diagnose (measure) BI using in organizations. So, the systematic and deliberate study on possibilities that offer BI for organizations and the ways of its assessment is crucial. The research questions I ask in this study are: (1) what possibilities offer BI systems for different organizations, (2) how to measure and evaluate the BI maturity in organizations? The study was based on: (1) a critical analysis of literature, (2) a observation of different BI initiatives undertaken in various organizations, as well as on (3) semi-structured interviews conducted in polish organizations in 2012. Some interviews, conducted in 20 polish enterprises, were held with over 80 responders: executives, senior members of staff, and ICT specialists They represented the service sector: telecommunications (T)-4, consulting (C)-4, banking (B)-4, insurance (I)-4, marketing agencies (MA)-4.

The reminder of my paper is organized as follows. Firstly, the idea of BI is described. Next, the issue of BI maturity models is recognized. Finally, Garter's Model for Business Intelligence and Performance Management is used to assess the level of BI maturity in surveyed organizations.

The paper provides valuable information on the possibilities that offer BI systems for organizations and the ways of their evaluation. It is dedicated for decision-makers, managers and ICT specialists interested in using BI systems in organizations. The study makes useful contribution to the literature and theorists on the idea of BI, BI using in organizations, and the ways of its assessment.

II. Background and Related Works On Business Intelligence

A. The issue of Business Intelligence

From a historical perspective, Business Intelligence (BI) is a popularized, umbrella term introduced by Howard Dresner of the Gartner Group in 1989 to describe a set of concepts and methods to improve business decision making by using fact-based support systems [3]. BI involves collecting, storing and presenting data, and managing knowledge by means of employing different analytic tools. Intelligent data analysis is usually obtained by OLAP (On-Line Analytical Processing), data mining and data warehouses techniques [4].

With the passing of time, the term BI has been understood much more broadly, namely, as a connecting factor of different components of decision support infrastructure [5], and providing comprehensive information for policy makers [6]. Hence, many definitions of BI focus on the capability of an enterprise to improve business efficiency and to achieve higher business goals. It is said that BI provides a means to obtain crucial information to improve strategic decisions and, therefore, plays an important role in current decision support systems [7]. The term Business Intelligence (BI) is often used as a broad category of technologies, applications, and processes for gathering, storing, accessing, and analyzing data to help users make better decisions [8]. More generally, BI can be understood as a process providing better insight in a company and its chain of actions.

According to many authors there are distinguished 3 ages in the development of BI: BI 1.0, BI 2.0, BI 3.0.

The first age of BI, called BI 1.0. falls on seventies and eighties of XX century. It is closely related with the management information systems (MIS), executive information systems (EIS), and decision support systems (DSS). Generally, the first applications from this age were dedicated on

978-83-60810-53-8/$25.00

mainframes. They were able to process the simple tasks for operational and tactical management. They were characterized by production the simple reporting and represented simple, static applications. Individual reports were written by expert programmers. BI 1.0 was focused on "delivery to the consumer" and market leaders include: SAS, IBM [9].

The second age of BI (1990-2005) - BI 2.0 is the type of enterprise scale BI we see today. It means a concept and methodologies for improvement of business decisions using facts and information from supporting systems [33]. It is characterized by end-user friendlier client-based BI tools and centralized. Data warehouse configured to deliver preformatted information to specialists analysts and users within management. So, the role of BI 2.0 and its impact on organizations (compared to BI 1.0) has been changed. From simple, static analytical applications, BI 2.0 has evolved into solutions that can be used in strategic planning, predictive modeling, forecasting, monitoring operations, and studying the profitability of products [1],[10]. BI 2.0 is focused on "creation and delivery for consumers" and market leaders include: Business Objects, Cognos, Hyperion, Microsoft, Teradata, Oracle.

BI 3.0. presents a new era in the evolution of BI. Thanks to web and mobile technologies it appears intelligent business network for every one. There is a growing acceptance of the idea that analysis is a collaborative (not only singular) and social effort. It focuses on a collaborative workgroups (which are self-regulated) and on information outcomes within the confines of core business interaction with customers, employees, regulators etc. There is common sense that BI 3.0 should go beyond reliance on structured data available in internal sources but should use also external, mostly unstructured data in various formats (social media posts, free form web content, images, and video files) [11]. BI 3.0 is concentrated on "creation, delivery and management for consumers" [9]. According to Scott [12] there are 5 core attributes that support BI 3.0 philosophy: proactive, real-time, integrated with business processes, operational (available to line workers), and extended to reach beyond the boundaries of the organizations to improve information delivery and decision support functionality for all. It is indicated also that there is no reason to depreciate in BI 3.0 the functions (known from BI 2.0) like: reporting, OLAP, data mining. They have still their strong position. BI 3.0 philosophy is to raise the added value of BI tools' architecture by anchoring collaborative style of information search and analysis with intuitive and self-service user interface that delivers timely and highly relevant insights to anyone who is properly authorized and needs them [11].

According to Chatter [13] there are 3 prerequisites for software tools to be recognized as a BI 3.0 tools: be social, relevant (automatically delivers relevant insights that users really need according to their situation and user profile), and fully self-service (intuitiveness).

The analysis of different articles, papers and reports show that BI is mainly identified with:

- tools, technologies, and software products. BI is used to collect, integrate, analyze and make data available [14]. It includes: data warehouse, data mining and OLAP (On-line Analytical Processing). Data warehouse is a key technology, integrating heterogenic data from different information sources for analytical purposes [7]. Hence, it is assumed that the main tasks to be faced by BI include: intelligent exploration, integration, aggregation and a multidimensional analysis of data originating from various information resources [15];
- knowledge management. BI is the capability of the organization to explain, plan, predict, solve problems and learn in order to increase organizational knowledge [16]. BI is assumed to be solution that is responsible for transcription of data into information and knowledge [10];
- decision support systems. BI is considered as a new generation of decision supports systems. They differ from previous management information systems in, first of all, their wider thematic range, multivariate analysis, semi-structured data originating from different sources and multidimensional data presentation [17], [16], [6], [18];
- dashboards. Dashboards are the becoming the preferred method for delivering and displaying BI to users. They are more visual and intuitive, and typically provide linkages that enable immediate action to be taken [5];
- new working culture with information - BI constitutes an important upturn in techniques of working with information [4]. It means specific philosophy and methodology that would refer to working with information and knowledge, open communication and knowledge sharing [10]. The BI users must know more than just technology - business and soft skills are needed too;
- process. The process constitutes of activities to gather, select, aggregate, analyze, and distribute information [19]. Some of these activities are the responsibility of the BI staff, while others are the joint responsibility of the BI staff and the business units [8];
- analytics and advanced analyses. The term "analytics", introduced by Davenport and Harris[20], means "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, fact-based management to drive decisions and actions. Analytics are a subset of what has come to be called BI: a set of technologies and processes that use data to understand and analyze business performance";
- Competitive Business Intelligence (CI). Another subset of BI is CI. Its goal is to provide a balanced picture of the environment to the decision makers [15]. CI is the analytical process that transforms scattered information about competitors and customers into relevant, accurate

and usable strategic knowledge on market evolution, business opportunities and threats.

B. Business Intelligence in organizations

According to Goodhue, Wixom and Watson [8] there are three targets that organizations can aim for when implementing BI:

- single or a few applications. They are used in selected departments (marketing, sale, controlling etc.) to support effective marketing campaigns, to analyze profitability different products and to monitor the behaviors of customers;
- BI infrastructure. "The organizations create an infrastructure for BI by clearing up and defining their data, establishing efficient process to move data from source systems to a highly extensible data warehouse, implementing a variety of BI tools and applications, and investing in BI user training";
- organizational transformation. BI systems are used in order to introduce new business model oriented on change management, knowledge management and customer relationship management. BI aims to run company differently. In this case, some investments in huge corporate data warehouse are needed.

Many case studies confirm that BI may be utilized in an organization for [11], [21], [22];

- increasing the effectiveness of strategic, tactic and operational planning including first of all: (a) modelling different variants in the development of an organization; (b) informing about the realization of enterprise's strategy, mission, goals and tasks; (c) providing information on trends, results of introduced changes and realization of plans; (d) identifying problems and 'bottlenecks' to be tackled; (e) providing analyses of "the best" and "the worst" products, employees, regions; (f) providing analyses of deviations from the realization of plans for particular organizational units or individuals; (g) and providing information on the enterprise's environment;
- creating or improving relations with customers, mainly: (a) providing sales representatives with adequate knowledge about customers so that they could promptly meet their customers' needs; (b) following the level of customers' satisfaction together with efficiency of business practices; (c) and identifying market trends;
- analysing and improving business processes and operational efficiency of an organization particularly by means of: (a) providing knowledge and experience emerged while developing and launching new products onto the market; (b) providing knowledge on particular business processes; (c) exchanging of knowledge among research teams and corporate departments.

The most spectacular results, from using BI, have been observed while running promotional campaigns, anticipating sales and customer behaviors, creating loyalty policies and investigating anomalies and frauds [23]. The studies show that BI may also generates a wide variety of organizational benefits [8]. Some BI benefits are tangible and easy to measure (e.g., the reduction of software and hardware licenses and fees). Other benefits, such as improved quality and timeliness of information or improvement of business process and the enabling of new ways of doing business, are much more difficult to quantify, but they may generate a competitive advantage or open up new markets for the company. According to Howson [24], who examined 513 organizations, to the most significant measures of success of BI projects belong: improved business performance, better access to data, support of key stakeholders, user perception that it is mission critical, return on investment, percentage of active users, costs savings, defined users.

C. Business Intelligence maturity models

The effective development of BI in the organization should be based on scientific theories. It seems that theory of maturity models gives the good foundation [25]. The term of maturity describes a "state of being complete, perfect or ready. To reach this a desired state of maturity, an evolutionary transformation path from an initial to a target stage needs to be progressed" [26]. Maturity models are used to guide this transformation process. They help define and categorize the state of an organizational capability [27]. Maturity model for BI helps organization to answer for these questions: where in the organization is most of the reporting and business analysis done today?, who is using business reports, analysis and success indicators?, what drives BI in the organization?, which strategies for developing BI are in use today?, and what business value does BI bring? [28].

A high number of maturity models for BI has been proposed [26], [29], [27], [30].

One of the most popular is Gartner's Maturity Model for Business Intelligence and Performance Management. It describes a roadmap for organizations to find out where they are in their usage of BI. It provides a path for progress by which they can benefit from BI initiatives. The model recognizes five levels of maturity: unaware, tactical, focused, strategic, and pervasive. The assessment includes three key areas: people, processes, metrics and technology [28], [29], [30]. The first level is often described as "information anarchy". It means that data are incomplete, incorrect, inconsistent and organization does not have defined metrics. The uses of reporting tools are limited. The second level of BI maturity means that the organization starts to invest into BI. Metrics are usually used on the department level only. Most of the data, tools, and applications are in "silos". Users are often not skilled enough in order to take advantage of the BI system. At the third BI maturity level the organization achieves its first success and obtains some business benefits from BI, but it still applies to a limited part of the organization. Management dashboards are often requested at this level. At the strategic level, organizations have a clear business strategy for BI development. The application of BI is

TABLE I.
OVERVIEW OF BI MATURITY MODELS

Name of the model	Description
TDWI's Business Intelligence Model –Eckerson's Model Eckerson [30]	This model focuses mainly on the technical aspect for maturity assessment. It constitutes of 6 maturity levels and uses a metaphor of human evolution: prenatal, infant, child, teenager, adult and sage
Gartner's Maturity Model for BI and PM [31]	The model is a mean to assess the maturity of an organization's efforts in BI and PM and how mature these need to become to reach the business goals. The model recognizes 5 maturity levels: unaware, tactical, focused, strategic, pervasive
AMR Research's Business Intelligence/ Performance Management [29]	The model is described by 4 maturity levels: reacting (where have we been?), anticipating (where are we now?), collaborating (where are we going?), and orchestrating (are we all on the same page?). It is used to assess the organization in the area BI and PM
Business Information Maturity Model [28]	The model is characterized by 3 maturity levels. The first level answers the question „ what business users want to access", the second "why the information is needed", the third "how information put into business use"
Model of Analytical Competition [1]	The model describes the path that an organization can follow from having virtually no analytical capabilities to being a serious analytical competitor. It includes 5 stages of analytical competition: analytically impaired, localized analytics, analytical aspirations, analytical companies, and analytical competitors
Information Evolution Model, SAS [32]	The model supports organization in assessing how they use information to drive business, e.g., to outline how information is managed and utilizes as a corporate asset. It is characterized by 5 maturity levels: operate, consolidate, integrate, optimize, innovate
Model Business Intelligence Maturity Hierarchy [33]	The model was developed in knowledge management and constitutes of 4 maturity levels: data, information, knowledge and wisdom
Infrastructure Optimization Maturity Model [28]	The model enables a move from reactive to proactive service management. It aids in assessing different areas comprising the company infrastructure. The model is described by 4 maturity levels: basic, standardized, rationalized (advanced), and dynamic
Lauder of Business Intelligence (LOBI) [28]	The model describes levels of maturity in effectiveness and efficiency of decision making. IT, processes and people are assessed from the perspective of 6 levels: facts, data, information, knowledge, understanding, enabled intuition
Hawlett Package Business Maturity Model	The model aims at describing the path forward as companies work toward closer alignment of business an IT organizations. It includes 5 maturity levels: operation, improvement, alignment, empowerment, and transformation
Watson's Model [27]	The model is based on the stages of growth concept, a theory describing the observation that many things change over time in sequential, predictable ways. The maturity levels include: initiation, growth, and maturity
Teradata's BI and DW MM [26]	Maturity concept is process-centric, stressing the impact of BI on the business processes The model has 5 maturity levels: reporting (what happened?), analyzing (why did it happen?), predicting (what will happen?), operationalizing (what is happing?), and activating (make it happen).

often extended to customers and suppliers. It supports the tactical and strategic decision making. Sponsors come from the highest management. At the last BI maturity level, BI pays pervasive role for all areas of the business and corporate culture. BI provides flexibility for adapting to the fast business changes and information demand. The users have access to information and analysis needed for creating a business value and influence business performance. The usage of BI is available to customers, suppliers, and other business partners.

Moving from one maturity level to another requires changes in all of the characteristics that make up these stages (e.g., changes in management vision, founding, data management) [8].

III. RESEARCH METHODOLOGY

The aim of the survey was to assess the BI using in 20 purposefully selected organizations, and to determine the factors that allow the firms to achieve high competences in BI, and consequently various business benefits [2]. The research was of qualitative nature and used as a research tech-

TABLE II.
TYPES OF ASKED QUESTIONS AND ANSWERS

No	Asked questions during interviews	Answers (number of organizations)
1	How do you define BI?	Tools to manage information (9), data warehouse (5), analytical applications (4), new way of doing business (2)
2	What do you use BI for (reporting, ad-hoc reporting, analyzing, alerting, predictive modeling, operationalizing, optimization, activating, etc.) ?	Reporting (15), ad-hoc reporting (9), analyzing (12), alerting (2), predictive modeling (2), optimization (3), activating (2)
3	Does your organization have a defined BI strategy?	Comprehensive BI strategy (5), partly defined BI strategy (12), none (3)
4	Does your organization have defined business processes?	Defined basic processes (9), defined core business processes (6), not defined (5)
5	Does your organization/department have defined metrics?	Metrics for selected departments (13), metrics for the whole organization (4), none metrics (3)
6	Assess the quality of data used in your organization (complete, correct, consistent; high/medium/poor quality data, etc.)	High quality data (6), medium quality data (11), rather poor quality data (3)
7	Are you skilled enough in order to take advantage of BI systems?	Skilled enough (7), not skilled enough (8), poor skilled (5)
8	Do you use management dashboards?	Used management dashboards in limited scope (14), used management dashboards in whole organization (4), not used (2)
9	Is your BI (un)limited to the part/department of organization?	BI limited to the part of organization (15), unlimited (5)
10	Are you motivated to use BI (how)?	Users motivated by training (8), motivated by bonuses (6), not motivated (6)
11	Do you use BI for analyzing customers, suppliers, competitors and other business partners?	BI for analyzing customers (17), suppliers (14), competitors (5), other stakeholders (4)
12	Who is the sponsorship of BI in your organization?	CIO (3), senior management (6), business analyst (4), ICT specialists (7)
13	What kind of BI software do you use?	Regional data warehouse (9), centralized data warehouse (5), operational data bases (6)
14	Describe some successes/failures from using BI	Success: acquiring new customers (14), acquiring new suppliers (11), increase of sale (8), fraud detection (6), launching new channels of sale (3), launching new products (3). Failures: not trust in BI (4), gap between BI/ business (12), users do not recognize their own data after it is processed (7), decision-making skills absent (6), BI is expensive (5)
15	Indicate some benefits from using BI	Better access to data (13), better decisions (12), improvement of business process (9), improved business performance (8), costs saving (7), transparency of information (5), new way of doing business (2)

nique of an in-depth interview. Types of core interviews questions relevant to this paper are reflected in table 1.

The survey was conducted in 2012 among purposefully selected firms (in Poland) that are considered to be advanced in BI. They represented the service sector: telecommunications (T)-4, consulting (C)-4, banking (B)-4, insurance (I)-4, marketing agencies (MA)-4. Interviews were held with executives, senior members of staff and ICT specialists. Interviewees were selected on their involvement in BI or on their ability to offer an insight based on experience in BI and related decision support systems. Gartner's Maturity Model for Business Intelligence and Performance Management (described in the previous section), for the assessment of the BI-maturity level in selected organizations, was used.

IV. FINDINGS

My research confirmed that BI identified in the literature was also experienced in selected organizations. Table 2 presents the answers for asked questions. The BI maturity in surveyed organizations (mapping onto Gartner's Maturity Model for Business Intelligence and Performance Management) and factors that allow them to achieve the various business benefits with BI, are indicated in table 3.

V. DISCUSSION

The collected and processed data were mapped onto Gartner's Maturity Model for Business Intelligence and Performance Management. The obtained results allow to state that among 20 surveyed organizations two organizations fall into the category of "pervasive" level. These were telecommunication company and marketing agency. Their analytical and BI competences are aimed at business benefits, like: acquiring new customers, launching new products and new channels of sale.

BI competences are treated by these organizations as their core competences that help them to compete on the market. Organizations achieve significant economic benefits and use BI for marketing analyses (sales profitability, profit margins, meeting sales targets, time of orders), customer analyses (time of maintaining contacts with customers, customer

TABLE III.
OVERVIEW OF BI- MATURITY LEVELS IN SURVEYED ORGANIZATIONS

Level	People	Process	Metrics and technology	Scope of benefits
Unaware	Users do not know their own data requirements or how to use them	Users do not know business processes; data are poor quality	Lack of appropriate hardware and software; the metrics are not defined; the use of reporting is limited	Almost none
Tactical 2I, 2C, 1MA	The users take the first BI initiatives; low support from senior executives	Identification of basic business processes	Regional data warehouses are built; analyzing trends and past data; first interactive reporting tools; metrics are usually used on the department level only	Low benefits limited to small group of users; better access to data and static reporting
Success factors: support from senior management, appropriate BI tools, quality of data, defined business processes and metrics				
Focused 2T, 1MA, 2C, 2I, 2B	Users try to optimize the efficiency of individual departments by BI	Standardization of business processes and building best practices in BI	Management dashboards are used; a centralized data warehouse is built; ad-hoc reporting, query drilldown	Benefits limited to departments and business units; improvement of internal business processes and decision making on operational level
Success factors: developing corporate culture based on facts, stating clearly BI strategy, implementing training system on BI				
Strategic 1MA, 1T, 2B	Users have high BI capabilities, but often not aligned with right role	Business process management based on facts	High-quality data; have BI strategy; using more complex prediction and modeling tools; data mining	Benefits for the whole organization; integrated analysis for finance, logistics, production; improvement of decision making on all levels of management
Success factors: support from CEO, motivation of users for collecting, analyzing and using information				
Pervasive 1T, 1MA	Users have capabilities and time to use BI; skill training in BI; users are encouraged to collect, process analyze and share information; CEO passion and broad-based management commitment	Broadly supported, process-oriented culture based on facts, learning and sharing of knowledge	Enterprise-wide BI architecture largely implemented; customized reports; business and BI are aligned and cooperative	Benefits for the whole environment; competing in BI; new ways of doing business
Success factors: strong support of CEO, effective HRM and all user's trust in BI				

profitability, modeling customers' behavior and reactions, customer satisfaction), monitoring of competitors and current trends in the marketplace. The common analytical approach is used by the whole organization where broadly supported fact-based and learning culture is cultivated. The interviewees confirmed that the factors that help those organizations to stay at that high maturity level in BI, include strong support of CEO and all user's trust in BI.

An interesting group was made up of organizations classified at the fourth BI maturity level. Four organizations (1MA, 1T, 2B) in my study fit into the strategic level. They do not compete through BI, but they have high competences in using different BI analyses, like: financial analyses (reviewing of costs and revenues, calculation and comparative analyses of corporate income statements, analyses of corporate balance sheet and profitability), customer profitability, customer segmentation, improving marketing effectiveness. It seems that there is a very little to be done in order to use BI for making significant changes in running a business. Therefore, shifting these organizations from "strategic" to "pervasive" level requires more support from CEO and his/her real passion. The interviewees indicated the greater need for motivation of users for collecting, analyzing and using information.

The survey has shown that up to 9 organizations (2T, 1MA, 2C, 2I, 2B) use BI on the department level. Although they would be much more common in a random sample, and perhaps the largest group. BI in these organizations has not been playing a strategic role and benefits from it are limited. BI is used to perform ad hoc reporting and to answer questions related to departments' ongoing operations, up-to-date financial standing, sales and co-operation with suppliers and customers. BI and management are often not aligned. The observation and interviews with senior executives allow to state that the lack of appropriate knowledge about possibilities of BI among staff results in a relatively low use of it. Therefore, the main tasks for organizations include first of all: developing corporate culture based on facts and learning,

stating clearly BI strategy and implementing training system on BI.

I found in my study that 5 organizations (2I, 2C, 1MA) are at the position of "tactical" maturity level. They use a traditional approach to management, focused more on the performing the basic tasks of departments than on business processes. The knowledge about BI in these organizations is rather low and identified mainly with regional data warehouses or databases. Only basic business processes are recognized and basic metrics are used. The interviewees indicated that many users have some problems with recognizing their own data after processing. The users have rather low experience with other types of management information systems. Their intellectual resources are not adequate in order to develop complex BI infrastructure and to use it for improvement of business processes and decision-making.

To conclude the discussion on the use of BI in surveyed organizations, I wonder why organizations in a similar segment, with similar financial resources and comparable BI infrastructure, derive from BI such diverse benefits (e.g. in the studied case, telecommunications companies and advertising agencies). Seeking the answer to this question it should be noted that the organizations, that have been classified into the category of BI "pervasive" level, were highly determined to collect, process, analyze, and share information. Corporate culture based on facts and learning helped them to use chances offered by BI. The most important factors that decided on the success of BI initiatives refer not to the technology, but to the strong believe of all users in BI.

VI. Conclusions

The main conclusion of this study is that BI systems may be a trigger for making more effective decisions, improving business processes and business performance, as well as doing new business. Observation and conducted discussions with interviews let to state, that the factors that allow organizations to achieve business benefits with BI, include first of all: management leadership and support, corporate culture, expressed by effective information resources management, clearly stated strategy and objectives, and use of appropriate BI technologies. Additionally, the important factors were: clearly defined business processes, business performance measurement, incentive system to encourage collecting, analyzing information and knowledge sharing, appropriate resources (financial, intellectual), training and education on BI and knowledge management.

The research has made a theoretical contribution to our understanding of BI issue. The outcomes extend current theory on BI and provide useful information, which hopefully will help the organizations to understand the consequences of the different ways of BI using, as well as, to determine the factors on which they should give particular attention while building different BI applications.

References

[1] T. H. Davenport, J. G. Harris, and R. Morison, *Analytics at Work: Smarter Decisions, Better Results,* Harvard Business Press, Cambridge, 2010.

[2] C. M. Olszak, "The Business intelligence-based Organization- new chances and Possibilities", in *Proceedings of the International Conference on Management, Leadership and Governance*, Bangkok University, Thailand, 7-8 February 2013, Edited by Vincent Ribiere and Lugkana Worasinchai, Bangkok University, Published by Academic Conferences and Publishing International Limited Reading UK 44-118-972-4148, pp. 241-249.

[3] D. J. Power, *A brief history of Decision Support Systems*, [online], http://dssresources.com/history/dsshistory.html, 2007.

[4] B. Liautaud, and M. Hammond, *E-Business Intelligence. Turning Information into Knowledge into Profit*, McGraw-Hill, New York, 2002.

[5] C. Ballard, D.M Farrell, M. Gupta, C. Mazuela, and S. Vohnik, *Dimensional Modeling, in Business Intelligence Environment*, IBM, International Technical Support Organization, 2006.

[6] S. Negash, "Business Intelligence", *Communications of Association for Information Systems*, Vol. 13, 2004, pp. 177-195.

[7] W. H. Inmon, D. Strauss, and G. Neushloss, *DW 2.0: The Architecture for the Next Generation of Data Warehousing*, Elsevier Science, Amsterdam, 2008.

[8] B.H. Wixom, and H.J. Watson, "The BI-based organization", *International Journal of Business Intelligence Research*, Vol. 1, No. 1, 2010, pp 13-28.

[9] S. J. Gratton, "BI 3.0 The Journey to Business Intelligence. What does it mean?", http://www.capgemini.com.technology (retrieved October, 2012).

[10] S. Negash, and P. Gray, "Business Intelligence", in F. Burstein, and C.W. Holsapple (ed), *Decision Support Systems*, Springer, Berlin, 2008, pp 175-193.

[11] R. Nemec, "The Application of Business Intelligence 3.0 Concept in the Management of Small and Medium Enterprises", in M. Tvrdikova, J. Minster, P. Rozenhal (ed), *IT for Practice 2012*, Economicka Faculta, VSB-TU Ostrava, 2012.

[12] N. Scott, "The 3 ages of Business Intelligence: Gathering, Analysing and Putting it to Work*"*, http://excapite. blogspot-ages-of-business-ontelligence.html (retrieved January 2013).

[13] R. Chatter, "Decoding BI 3.0", http://searchbusinessintelligence. techtarget.in/answer/decoding-BI-30 (retrieved January 2013).

[14] J. Reinschmidt, and A. Francoise, *Business Intelligence Certification Guide*, IBM, International Technical Support Organization, 2000.

[15] V. L. Sauter, *Decision Support Systems for Business Intelligence*, Wiley, New Jersey, 2010.

[16] D. Wells, "Business analytics – getting the point", [online], http://b-eye-network.com/view/7133, 2008.

[17] J. A. O'Brien, and G. M. Marakas, *Introduction to Information Systems* (13th ed.), McGraw-Hill, New York, 2007.

[18] H. Baaras, and H.G. Kemper, "Management support with structured and unstructured data – an integrated Business Intelligence framework", *Information Systems Management*, Vol. *25,* No. 2, 2008, pp 132-148.

[19] Z. Jourdan, R. K. Rainer, and T. Marschall, "Business Intelligence: An Analysis of the Literature *", Information Systems Management*, Vol. 25, No. 2, 2007, pp. 121-131.

[20] T. H. Davenport, and J. G. Harris, *Competing on Analytics. The New Science on Winning,* Harvard Business School Press, Boston Massachusetts, 2007.

[21] P. Hawking, S. Foster, and A. Stein, "The Adoption and Use of Business Intelligence Solutions in Australia", *International Journal of Intelligent Systems Technologies and Applications*, Vol. 4, No. 1, 2008, pp 327-340.

[22] S. Chaudhary, "Management factors for strategic BI success", in , M.S. Raisinghani (ed.), *Business Intelligence in digital economy. Opportunities, limitations and risks*, Hershey: IGI Global, 2004, pp 191-206.

[23] C. M. Olszak, and E. Ziemba, "*Business intelligence systems as a new generation of decision support systems"*, in *Proceedings of PISTA, International Conference on Politics and Information Systems: Technologies and Applications,* Orlando: The International Institute of Informatics and Systemics, 2004.

[24] C. Howson, *Successful Business Intelligence: Secrets to Making BI a Killer Application,* McGraw-Hill, New York, 2008.

[25] C. M. Olszak, "Competing with Business Intelligence", in *IT for Practice*, Ekonomicka fakulta VSB-TU Ostrawa, 2012, p. 98-108.

[26] G. Lahrmann, F. Marx, R. Winter, and F. Wortmann, "Business Intelligence Maturity: Development and Evaluation of a Theoretical Model", *in Proceedings of the 44 Hawaii International Conference on System Science*, 2011.

[27] H. J. Watson, T. Ariyachandra, and R. J. Matyska, "Data Warehousing Stags of Growth*", Information Systems Management,* Vol. 18, No. 3, 2011, pp. 42-50.

[28] I. H. Hribar Rajteric, "Overview of Business Intelligence Maturity Models", *Management*, Vol. 15, No.1, 2010, pp. 47-67.

[29] J. Hagerty, "AMR Research's Business Intelligence/ Performance Management Maturity Model", http://www.eurim.org.uk/activities/ig/voi/AMR_Researchs_Business_Intelligence.pdf, retrieved September 2011, pp. 1.

[30] A. Schick, M. Frolick, and T. Ariyachandra, "Competing with BI and Analytics at Monster Worldwide", in *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011.

[31] N. Rayner, *Maturity Model of Overview for Business Intelligence and Performance Management*. Gartner Inc. Research , 2008, pp. 2.

[32] SAS, "Information Evaluation Model", http://www.sas.com/software/iem/, Retrieved September 2011.

[33] R. Deng, "Business Intelligence Maturity Hierarchy. A New Perspective from Knowledge Management", *Information Management,* http://www.information management.com/infodirect/20070323/1079089-1.html, retrieved September 2011, pp. 1.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 959–962

Towards a Better Understanding of Context-Aware Applications

Emilian Pascalau
Conservatoire National des Arts et Métiers
France
Email: emilian.pascalau@cnam.fr

Grzegorz J. Nalepa and Krzysztof Kluza
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
Email: {kluza,gjn}@agh.edu.pl

***Abstract*—With the new technological advances and strong move towards Future Internet and Internet as a Platform a new environment is emerging. This environment is generative, social, strongly interactive and collaborative, so users play a fundamental role in it. Business applications are simplifying, webifying and getting more user-centric. In this environment, context and context-awareness plays a fundamental role, as context gives meaning and accurately describes the situation of an user. This paper introduces the basis for a new research methodology that aims to address and visualize the topic of context and context-awareness from a holistic point of view, by means of text mining and text clustering.**

I. Introduction

THE concepts of context and context awareness have been studied for more than 20 years in the field of artificial intelligence, computer and cognitive science. However, it has been still identified by Gartner, alongside cloud computing, business impact of social computing, and pattern based strategy, as being one of the broad trends that will change IT and the economy in the next 10 years [1].

Moreover these IT and economical changes are reflecting themselves also onto business applications. Applications are simplifying, are becoming mobile, are moving to the cloud, are getting more social and user focused [2].

For instance smart phones although almost anywhere present nowadays, they present also a lot of downsides, i.e. they ring in movie theaters, conferences, meetings, air flights, they are the cause of car accidents. Theaters and even restaurants in some places have employed cell-phone jamming to limit cell-phone disruptions. This could be considered partly a social problem, but is also a technological problem. Through better applications, people can be helped to make better and safer use of their smart phones.

Hence we are faced with a series of new challenges in the context of developing future business apps. They need to:

- be highly adaptive;
- provide UIs that are user specific;
- provide means for users to modify them by themselves according to their needs, goals, context, while still keeping the underlying infrastructure in place;
- be interactive;
- be distributed, at the same time cloud computing ready;
- support both desktop and mobile environments, while providing a similar experience, finally
- developed with business users, and vendors and for customers, while hiding as much as possible all technical requirements.

To understand what is required to tackle these challenges a holistic and unified approach is needed. In this setting, the purpose of this paper is to introduces the basis of a new research methodology that aims to address and visualize the topic of context and context awareness from a holistic point of view, by means of text mining and text clustering.

It is worth emphasizing, that in this paper we do not explicitly introduce the term *"context"*. In fact, we aim at identifying and categorizing different uses of this term according to the existing literature.

II. Context and Context Awareness – Related Work in General

Over time, there have been written a large number of surveys on the topic of context and context-awareness addressing different directions of research. By far the most known and most cited (according to Google Scholar[1]: 3213 times) survey is the work by Dey and Abowd [3]. They argue that context is important for interactive applications and for applications where the context of users is changing rapidly. They discuss what a context-aware application means and give a definition for the term *context*.

Context and context-awareness in mobile computing is one of the oldest directions [4] of research when it comes to context-aware applications. This is because information such as user's location, nearby people and devices, time of day and user activity can be used to improve, for instance, latency time or boost up communication signal by using bandwidth from nearby devices and so forth.

Context modeling is part of any endeavor that tackles context-aware systems. Strang and Popien [5] address the problem of modeling of *context* in the context of ubiquitous computing paradigm (paradigm for which context is a driver). Several approaches are studied and discussed. Bolchini et al. [6] deal also with the problem of context modeling from a data-oriented perspective. A general framework for classifying context modeling approaches has been defined.

[1]http://scholar.google.com/scholar?cites=7671228831233541198&as_sdt=2005&sciodt=0,5&hl=en

978-83-60810-53-8/$25.00

Context-awareness has become lately of importance also for the development of web service-based systems [7]. These systems are more and more knowledge-based [8], e.g. using business rules and processes for a specification [9]–[11].

Although the issue of context has been well studied in the field of mobile systems and pervasive computing, the interest in the business-oriented community has been only recently growing. In fact with business applications becoming more social and user focused, cloud-based, massively distributed and mobile, both of these areas seem to finally meet. Therefore, studying the impact of context in business applications is of great conceptual and practical importance.

III. Research Methodology

To be able to investigate and to get a better understanding about the role of context and context awareness in relationship with business applications we used a particular research method which we are going to present in this section.

We are going to emphasize through out this section, that there is a huge amount of work that tackles the problem of context and context awareness in different fields and from different aspects. However, there is no unified view on the matter, nor – to the best of our knowledge – there is any approach that provides a holistic view on the subject. Therefore we propose a research methodology which takes advantage of the existing techniques for text clustering and text mining to get a broader view on the research that has been done on the side of context and context awareness.

The motivation to use text mining and clustering techniques is very simple. Too many papers that need to be organized, make the task almost impossible to fulfill. Moreover such an approach will provide an automatic way to extract related terms, topics and directions of research.

We present our methodology in a form of a simple workflow, modeled as a business process model, designed using the BPMN [12] notation and depicted in Figure 1. The model presents the steps that we took in our research approach and the ordering of those steps.

We have compiled a bibliography file which so far contains 94 carefully selected bibliographic entries that spans over a period of more than 20 years, starting 1991-2013. The quality of the papers is also an important factor. There are two ways to weight and asses the quality of the papers. One way is objective as it is given by the number of citation a paper has. We have extracted the number of citation, where this number existed, for a paper from digital libraries websites: CiteSeer[2], Google Scholar[3], ACM Digital Library[4], IEEE Xplore Digital library[5]. In the cases where there is no available citation number, we can not know for sure if a paper has been cited or not, therefore it is up to the researcher to read and asses the quality of a paper. This approach one could say that is rather subjective. Table I depicts only a small excerpt of this bibliography file, due to space limitation. However the entire file is available to download as a bibtex file or consult online at the following address: `http://geist.agh.edu.pl/pub:research:context_aware`.

The steps for compiling this bibliographic collections are depicted in Figure 1. We start by searching via Google for context related keywords i.e. *context*, *context-awareness*, *context-aware surveys*. A survey is a better entry point as it provides a wider view on a subject. These are just entry search terms. The more you search and read, the more terms can be further used. Besides the "random" search, we followed (searched) also concrete references that were indicated in the initial papers that we retrieved and read.

The next step in the process (See Figure 1) is to add bibliographic entries. We used for the clustering algorithms the abstract of each paper, if there was one. In consequence a bibliographic entry, if there is one, needs to have an abstract. Some of the papers also contained keywords. We have also used when available the keywords associated. These were combined with the abstract.

We used JabRef[6] to compile our bibliography. JabRef offers the functionality of an export layout, which we are using to export the bibliographic information into Carrot2[7] input format. Carrot2 as stated on the project website is an "Open Source Search Results Clustering Engine. It can automatically organize small collections of documents (search results but not only) into thematic categories". The reason for using Carrot2 over other tools (such as Lemur[8], Terrier[9]) is its *simplicity*. It was very simple to write an export layout from JabRef to Carrot2 XML input format. The export layout is also available online at the previosly given address. And also the results are by default given also in several visual formats.

Although Carrot2 provides several search algorithms we used Lingo and K-Means algorithms as they provided the best results. Unfortunately the free version of Carrot2 does not provide options to addresses issues such as synonyms in order to improve the results. Arthur and Vassilvitskii state in [13] that the K-Means method is a well known geometric clustering algorithm based on work by Lloyd [14]. Though the K-Means term has been first used by MacQueen [15]. According to [13] given a set of n data points, the algorithm uses a local search approach to partition the points into k clusters. Lingo [16] as described by the authors is able to capture thematic threads in a search result, that is discover groups of related documents and describe the subject of these groups in a way meaningful to a human.

Figures 2 and 3 depict the results of running the K-Means and respectively Lingo algorithms over our bibliographic collection. The results are visualized in a Foam representation. Results are similar but not the same. We can easily visualize directions of research and words related with the context concept. Having similar results it helps to verify the output of

[2]http://citeseerx.ist.psu.edu/
[3]http://scholar.google.com/
[4]http://dl.acm.org/
[5]http://ieeexplore.ieee.org/Xplore/home.jsp
[6]http://jabref.sourceforge.net/
[7]http://project.carrot2.org/
[8]http://www.lemurproject.org/
[9]http://terrier.org/

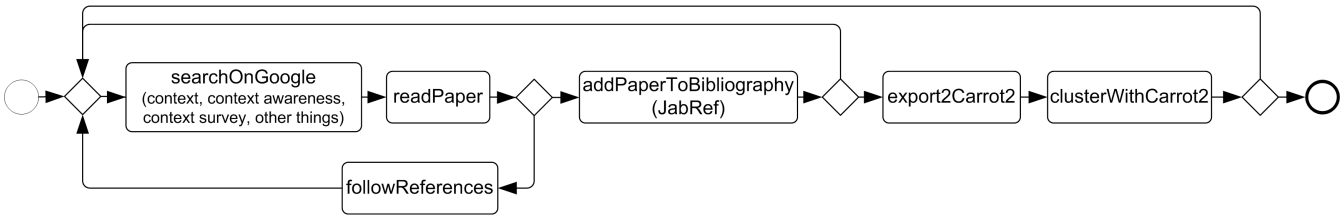


Fig. 1. Research Methodology - Business Process Model (BPMN notation)

TABLE I
EXCERPT OF CONTEXT RELATED BIBLIOGRAPHY

No.	Title	Year	Citations
1	Contextualization as an Independent Abstraction Mechanism for Conceptual Modeling	2007	25
2	A Survey on Context-aware Web Service Systems	2009	83
3	A data-oriented survey of context models	2007	181
4	Towards a better understanding of context and context-awareness	1999	3213

the clustering algorithms. Having differences helps to identify what each algorithm has missed with respect to the other.

The authors of [17] already identified that context is of fundamental importance for cognitive psychology, and computer science. Furthermore it states that in computer science the notion of context has been addressed in several areas such as: artificial intelligence, software development, databases, data integration, machine learning and knowledge representation. Since all these directions have been also identified by our research approach we argue that results are satisfactory in terms of how adequately the mining and clustering algorithms have performed.

In addition based on the information depicted in Figures 2 and 3, context has been used to address many of the future business apps challenges we have enumerated in Section I: adaptation, mobile computing, flexibility, user, modeling, task management, distributed systems, business process models.

IV. CONCLUSION

Context, context-awareness studied for a long time already, it still part of future development trends. Now, because of the changes (i.e. move to the cloud, social computing and so forth), that both IT and economy are witnessing, context is required to tackle the challenges that these changes bring in.

We propose a new research methodology in order to tackle the large number of publications that have been published over time, starting 1991 to present in order to get a more holistic view on the subject of context and context-awareness, by employing text mining and text clustering techniques.

This paper presents results of research that is work in progress. In our future work we will continue to investigate the results and ideas that can be extracted from using the text mining and text clustering based research methodology that we have introduced here. We will investigate semantic wikis, e.g. [18], [19], as systems that can be made context-aware and user-centric [20].

However, our ultimate goal is to provide a holistic design and deployment approach for context-aware user-centric business applications through a (1) unified modeling and methodological approach for context aware applications, and (2) unified execution framework of context aware applications.

ACKNOWLEDGMENT

The paper is supported by the AGH UST Grant 11.11.120.859.

REFERENCES

[1] M. Wang, "Context-aware analytics: from applications to a system framework," http://e-research.csm.vu.edu.au/files/apweb2012/download/APWeb-Keynote-Min.pdf, 2012.

[2] C. McLellan, T. Hammond, L. Dignan, J. Hiner, J. Gilbert, S. Ranger, P. Gray, K. Kwang, and S. Lui, *The Evolution of Enterprise Software*. ZDNet and TechRepublic, 2013. [Online]. Available: http://www.zdnet.com/topic-the-evolution-of-enterprise-software/

[3] A. K. Dey and G. D. Abowd, "Towards a better understanding of context and context-awareness," in *In HUC '99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*. Springer-Verlag, 1999, pp. 304–307.

[4] G. Chen and D. Kotz, "A survey of context-aware mobile computing research," Dartmouth College Hanover, NH, USA, Tech. Rep., 2000.

[5] T. Strang and C. Linnhoff-Popien, "A context modeling survey," in *Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp 2004 - The Sixth International Conference on Ubiquitous Computing*, 2004.

[6] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca, "A data-oriented survey of context models," *ACM SIGMOD Record*, vol. 36, no. 4, pp. 19–26, 2007.

[7] H.-L. Truong and S. Dustdar, "A survey on context-aware web service systems," *International Journal of Web Information Systems*, vol. 5, no. 1, pp. 5–31, 2009.

[8] A. Ligęza and G. J. Nalepa, "A study of methodological issues in design and development of rule-based systems: proposal of a new approach," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 117–137, 2011.

[9] G. J. Nalepa, "Proposal of business process and rules modeling with the XTT method," in *Symbolic and numeric algorithms for scientific computing, 2007. SYNASC Ninth international symposium. September 26–29*, V. Negru and et al., Eds., IEEE Computer Society. Los Alamitos, California ; Washington ; Tokyo: IEEE, CPS Conference Publishing Service, september 2007, pp. 500–506.

[10] G. J. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[11] M. Szpyrka, G. J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. SCI, F. Brazier and et al., Eds. Springer-Verlag, 2011, vol. 382, pp. 249–255.

[12] OMG, "Business process model and notation (bpmn)," OMG, Tech. Rep. formal/2011-01-03, November 2011.

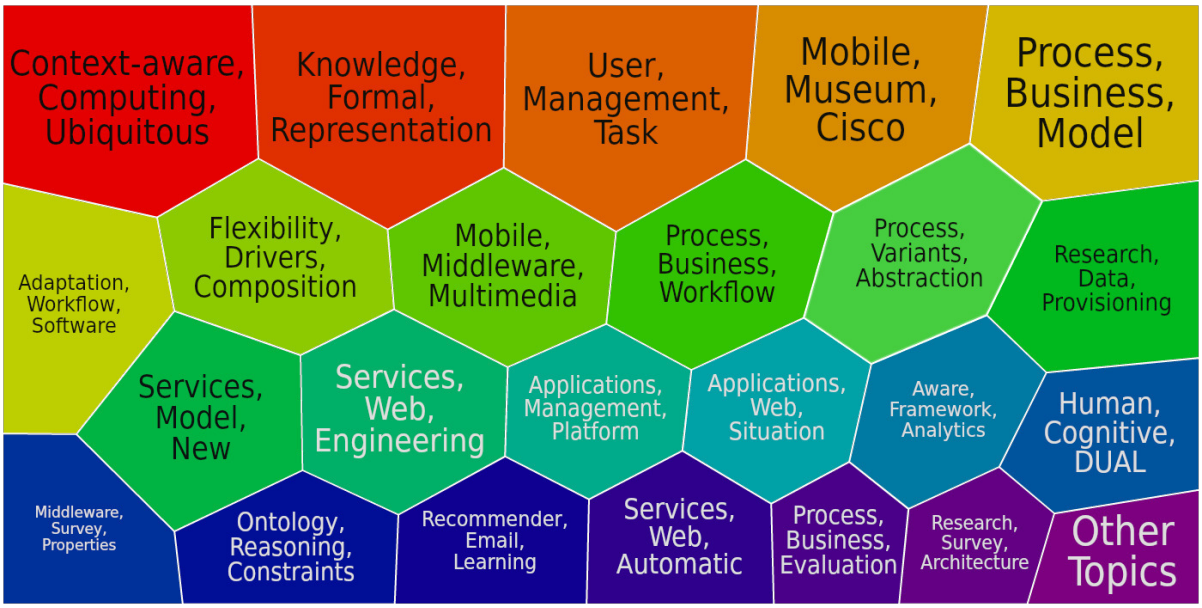


Fig. 2. K-Means – Foam Visual Representation

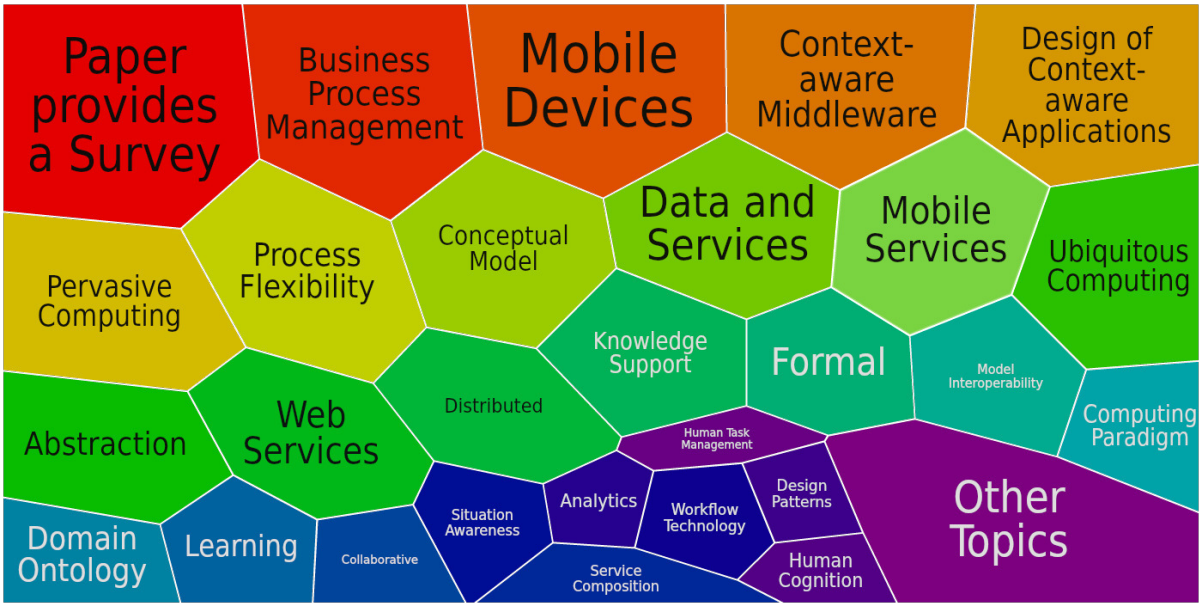


Fig. 3. Lingo – Foam Visual Representation

[13] D. Arthur and S. Vassilvitskii, "How slow is the k-means method?" in *Proceedings of the 2006 Symposium on Computational Geometry (SoCG)*, 2006.

[14] S. P. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–136, 1982.

[15] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[16] S. Osinski, J. Stefanowski, and D. Weiss, "Lingo: Search results clustering algorithm based on singular value decomposition," in *Intelligent Information Systems*, 2004, pp. 359–368.

[17] A. Analyti, M. Theodorakis, N. Spyratos, and P. Constantopoulos, "Contextualization as an independent abstraction mechanism for conceptual modeling," *INFORMATION SYSTEMS*, vol. 32, pp. 24–60, 2007.

[18] G. J. Nalepa, "PlWiki – a generic semantic wiki architecture," in *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, First International Conference, ICCCI 2009, Wroclaw, Poland, October 5-7, 2009. Proceedings*, ser. Lecture Notes in Computer Science, N. T. Nguyen, R. Kowalczyk, and S.-M. Chen, Eds., vol. 5796. Springer, 2009, pp. 345–356.

[19] G. J. Nalepa, "Collective knowledge engineering with semantic wikis," *Journal of Universal Computer Science*, vol. 16, no. 7, pp. 1006–1023, 2010. [Online]. Available: http://www.jucs.org/jucs_16_7/collective_knowledge_engineering_with

[20] S. Bobek, G. J. Nalepa, and W. T. Adrian, "Mobile context-based framework for monitoring threats in urban environment," in *Multimedia Communications, Services and Security: 6th International Conference, MCSS 2013: Kraków, Poland. June 6-7, 2013. Proceedings*, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 963–969

Rapid Application Prototyping for Functional Languages

Martin Podloucký
Department of Software Engineering,
Faculty of Information Technology,
Czech Technical University,
Prague, Czech Republic
Email: martin.podloucky@fit.cvut.cz

***Abstract*—This work addresses the problem of automated graphical user interface generation for functional programs in relation to rapid application prototyping. First an analysis of current state in the field of automated GUI generation is performed. Based on the analysis, the concept of *functionally structured user interface* (FSUI) is introduced. Meta-data system for code annotation is then specified for the Clojure programming language and a transformation from this system to FSUI data model is implemented. Finally, a graphical layer for displaying the actual interface is implemented in Clojure.**

I. Goal

THE focus of this work is to investigate the idea of automated GUI generation for functional programs for application in software prototyping. The main goals are:

1) Analyse the current state in the field of automated GUI generation.
2) Based on the analysis, explore possible approaches to GUI generation for functional languages.
3) Design and implement an automated GUI generator for functional programs.
4) The generator should create the GUI from annotated source code of the program.
5) Both the generator and the GUI should be implemented in functional language.

II. Paper structure

Firstly, the basic concepts, such as application prototyping and functional programming along with the motivation for this work, are explained in section III. Then, the Clojure programming language is introduced in section IV. Analysis of current state in the field of GUI generation is performed in section V. Some possible approaches to a solution of the stated problem, based on results of the analysis, are described in section VI. Finally, the concept of functionally structured user interface is presented in section VII.

III. Introduction

Software engineering industry is evolving and expanding rapidly, constantly searching for better techniques and technologies allowing software to be created faster with lower cost and better resulting quality [1]. Great desire for innovations can be observed in the field of business applications where the cost of the development is a key factor [2]. In this domain, the advantageous fact is that business applications have many common characteristics. They are based on similar principles, work with similar data and they have to deal with alike limitations [2]. This situation creates a good opportunity to use techniques such as prototyping [3] and generative programming [4].

A. Prototyping and generative programming

Application prototyping focuses on creating incomplete versions of the software being developed [3]. The purpose of such prototypes may be to explore possible solutions or to provide a piece of working software to the customer for evaluation in early stages of development [3]. There are several approaches to prototyping [5]: some prototypes are developed only to be discarded after they served their purpose. Such approach is called *throwaway prototyping* or sometimes *rapid prototyping*. Another approach is to incrementally evolve the prototype to fully functional product. This approach is often called *evolutionary prototyping*.

The main concern of either approach is being able to create prototypes as fast as possible [3]. Some techniques that can be used to shorten the prototype development time may be those of generative programming [4] and model driven development [6]. When significant parts of prototypes can be generated from conceptual models or from annotated source code, developers may focus more on implementing application logic rather then, for instance, implementing graphical user interface.

Graphical user interface (GUI) seems like a good candidate for a functionality to be generated out of annotated source code. GUIs used in prototypes do not need to meet too strict requirements on user friendliness since their aim is primarily to be suitable for presentation to the customer or for quick testing of basic functionality of the demanded software [3]. Furthermore, generating the GUI from source code may help to shorten the development time since there is minimum additional and external information needed to generate the GUI. According to [7] the source code already contains enough information to create properly working GUI.

978-83-60810-53-8/$25.00

B. Functional languages

The family of Lisp languages such as Common Lisp [8], Scheme [9], Clojure [10] and others are hereby taken as functional languages. This work is focused on generating GUI from source code written in such a functional language. There are several reasons for choosing functional languages rather than imperative ones as focus of this work.

1) Code in functional languages has simpler structure then a code written imperatively [11], [12]. Since thorough analysis of source code is needed this decision simplifies algorithms used in GUI generation.
2) Functional languages are growing in popularity even for writing business applications [13].
3) There is a number of similar tools for object-oriented languages such as Java or Smalltalk (discussed in section V). It is interesting to investigate which of its functionalities and approaches could be used in the functional world.

IV. The Clojure language

Clojure is a relatively new functional language based on Lisp [10]. It was created by Rich Hickey whose goal was to create mainstream functional language which can compete and complement present mainstream object oriented languages [10]. Basic features and characteristics of Clojure according to Rich Hickey [10] are

Functional programming
: The philosophy behind Clojure is that most parts of most programs should be functional, and that programs that are more functional are more robust compared to programs written imperatively. Clojure provides the common functional tools, however, doesn't force the program to be referentially transparent.

Lisp
: Clojure is a member of the Lisp family of languages, extending the code-as-data system beyond parenthesized lists (s-expressions) to vectors and maps.

Dynamic Development
: Programming in Clojure is interactive. It is not a language abstraction, but an environment, where almost all of the language constructs are reified, and thus can be examined and changed.

Hosted on the JVM
: Clojure is designed to be a hosted language, sharing the Java virtual machine (JVM) type system, garbage collector, threads etc. It compiles all functions to JVM bytecode. Clojure has simple syntax to reference and create Java classes therefore it can easily interoperate with Java and its libraries.

Runtime Polymorphism
: Clojure supports polymorphism at 3 levels. First, almost all of the core infrastructure data structures in the Clojure runtime are defined by Java interfaces. Second, Clojure supports the generation of implementations of Java interfaces. The final and primary language construct for polymorphism is the Clojure multimethod.

Concurrent Programming
: Since the core data structures are immutable, they can be shared readily between threads. However, it is often necessary to have state change in a program. Clojure allows state to change but provides mechanism to ensure that, when it does so, it remains consistent, while alleviating developers from having to avoid conflicts manually using locks etc. This is achieved by using the software transactional memory system (STM) [14].

Clojure is chosen here to represent a Lisp-like functional language. This work therefore focuses on generating GUI from source code written in Clojure. Clojure was chosen as a representative language for several reasons

1) Clojure can easily interoperate with Java and so the GUI can be created using standard Java GUI libraries such as Swing [15].
2) Clojure has very robust mechanism for state manipulation which allows to write the code for the GUI in Clojure itself.
3) Clojure is more suitable for writing business applications than other commonly used Lisps [13]. Therefore, it makes more sense to create rapid prototypes with graphical interfaces especially in Clojure.

V. Current methods for GUI generation

This section analyses current state in the field of automatic GUI generation. The author of this work did not find any current implementation of automated GUI generation for functional languages. Several implementations exist for object oriented languages such as Java, C# or Smalltalk. There is also work [7] which explores automated GUI derivation from programs written in term rewriting systems [16]. Let's first focus on object oriented frameworks.

Most of the object oriented GUI generators share a similar paradigm which is based on annotating domain data model using some kind of meta-data. This annotated model is then used as an input to the generator which then generates not only a graphical interface but often also application logic, database scheme and other components. This is essentially the whole application. In some cases a static generator isn't even necessary and the application's GUI accesses the annotated classes dynamically using reflection. Reflection is an ability of a programming language to change properties and behaviour of objects at runtime.

A. OpenXava

OpenXava [17] was hereby chosen as the main representative of the formerly described approach to object oriented GUI generation and a business application generation. This framework is capable of generating web applications in Java programming language with GUI based on AJAX technology. Input for the generator is a domain data model written in Java, annotated with combination of JPA annotations and

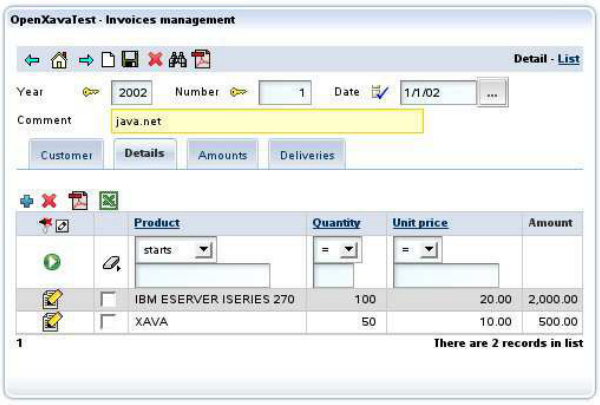


Fig. 1. Example of graphical user interface generated using OpenXava (taken from [17]).

OpenXava's own annotations. The output is a complete web application which uses Hibernate or JPA to store its data. An example of generated GUI is shown in figure 1.

OpenXava is a powerful technology which allows total separation of development of application's business logic and its user interface. Its advantages and disadvantages are more or less common to all hereby mentioned OO frameworks:

1) Independence of particular UI technology. It is possible to implement generators that produce code for different GUI libraries ranging from desktop UI to web or mobile interfaces.
2) Developers are freed from time consuming changes of the graphical interface when the model changes or when migrating to another platform.

However, the formerly described technology has some disadvantages as well:

1) Source code of the data model is often overloaded with meta-data which complicates orientation in the code and potential changes.
2) The user interface should better be derived from user behaviour and adapt to his intuition and habits rather than following the logic of the code underneath the UI [18].

B. Other OO frameworks

The OpenXava framework was used here as a representative of many others that function on similar principles. Some of them are listed along with their brief characteristics and differences.

NakedObjects [19]
: is a framework based on .NET platform. It works with domain model as well, however, the model is now written in languages for CLR – Common Language Runtime. In contrast to the OpenXava, NakedObjects uses reflection instead of code annotations and the resulting UI is created dynamically.

RomaFramework [20]
: works with domain model in Java. Besides code annotations, it makes it possible to use separated XML files to provide meta-data. Both annotation and XML approaches can be used together which can significantly reduce the amount of annotations in source code.

Magritte [21]
: is targeted at dynamic web application written in Smalltalk. Domain classes are annotated with so called description objects which add additional information using naming conventions. Such information is then used to generate GUI, database schemas etc.

There are other OO frameworks based on paradigm described above such as Tynamo [22] (previously named Trails [23]), JMatter [24], Apache Isis [25] and others. The nature of all frameworks mentioned so far is very close to the term of model driven software development (MDSD) [6] where the whole application is generated from some kind of model. As shown above, for decorating the model with additional UI information, a rich annotation language is often required. Even though, according to [7] a source code already contains enough UI information even without using expressive annotations. This idea is demonstrated in [7] by writing a small application computing an average of entered numbers. The application is written in term rewriting language [16]. Generating GUIs using this approach for functional languages is discussed in the next section.

VI. Possible approaches

Several possible approaches for generation of graphical interfaces from functional code were investigated during this work. Here are the main criteria according to which the final solution was assessed.

Practicality
: Useful and practical solution are favourized. For instance, code heavily burdened with vast amount of annotations is not considered practical since it corrupts readability and maintainability. These factors are key in prototyping since prototypes should be developed rapidly [3].

Separation
: The generated GUI should be separated from the application logic so that those two parts can be developed independently.

Generator extendability
: The solution should be extendable enough to allow creation of generators for other platforms such as mobile platforms or web.

A. IO oriented approach

The input/output (IO) oriented approach is based on the idea in [7]. It is hereby named IO oriented since its primary focus is on inputs entered by the user and outputs returned by the program in response. Term rewriting is in its nature close enough to the functional paradigm and thus the technique

described in [7] can be easily adapted for Clojure. The original tree rewriting program is shown in figure 2 and its Clojure implementation is shown by listing 1. This solution, however, has some serious disadvantages.

Listing 1 Program for computing an average rewritten to Clojure.

```
(defn add-number [average size]
  (do
    (label "Add:Result"
           (str "Average of " size " = " average))
    (action
      "Add:Add"
      (add-number
        (/ (+ (* average size) (number 0))
           (inc size))
        (inc size))
      "Add:Finish"
      (alert "Done" "Completed"))))

(defn compute []
  (add-number (number 0)
              (number 1 1 100)))
```

1) The code in Clojure is difficult to read since the GUI directives are heavily intertwined with the code for application logic. This violates our practicality requirement.
2) The GUI is tightly coupled with the application. It is not even possible to run the application without some kind of GUI. Moreover, the GUI directives cannot be easily removed from the program. Even the computational nature of the program has to take the GUI into consideration. This goes against the requirement of separation.
3) Another problem, which is more technical in nature, is interrupting the program. When the program is expecting some input from the user, it is blocking other possible actions such as termination.

Problems described above arise from the very nature of the idea used in [7] thus they cannot be easily overcome.

B. Action oriented approach

All problems described above have essentially the same cause. That is that functions contain directives for graphical interface inside their bodies. From certain point of view we can say that functions are actions that a user can execute. Such an action expects some inputs, produce some output, and as a an implicit side effect, it enables or disables some other actions. This information can be explicitly captured be annotating those functions before or inside their bodies.

The only GUI directive used inside bodies of functions would be a directive for enabling or disabling actions. The GUI generator could, by some kind of source code analysis, identify actions that belong together in the sense of their inputs. Graphical representation such as buttons and input fields belonging to those actions could then be grouped into graphical forms and windows.

According to the solution eventually presented in this article the above approach goes in the right direction. Even though, it is still rejected because serious difficulties are connected to it. The main issue is that the graph representing how actions are enabled and disabled through the flow of the program may heavily depend on the input of the program since we do not impose any restrictions on how the function should determine which actions will be enabled or disabled. Thus, constructing such a graph is in general an undecidable problem (explained in [26]). On the other hand, imposing restrictions that would make this problem decidable would seriously limit expressiveness of the functional language (further details in [26]).

VII. Functionally structured user interface

The above approaches to generating GUI from source code are not very useful still. The IO oriented approach has serious disadvantages and the action oriented approach forces us to overcome undecidability issues. This pushes us to look at the problem from another point of view. An interesting idea is to make the GUI less static and adjust its behaviour according to how functional programs actually work.

The question is how is the functional program essentially different from object oriented program from the user's point of view. By user is hereby meant the end user of the graphical interface of the program. One way to look at this is that when using graphical interface of OO program, the user creates and manipulates objects as primary entities. Functional program, on the other hand, is more about using functions as primary entities. These functions only then create, manipulate and transform objects. Thus, to reflect the different nature of functional programs, the GUI should be somehow function or operation (here we say *action*) centric. This is how we arrive at the term *functionally structured user interface* (FSUI).

A. FSUI specification

The main idea of the FSUI concept is that the UI has the same structure for all Clojure functional programs. Only the description of actions and their inputs and outputs is generated. GUI directives for enabling and disabling actions are not annotations. Instead, they are executable functions which then call the FSUI throw registered callbacks. Thus then, the program may be executed without any generated GUI just from a command line. In such a case, no callbacks are registered for the directives and so those directives do nothing.

Design of the graphical FSUI is shown in figure 3. The graphical FSUI is based on interaction of three basic kinds of components – *actions*, *invokers* and *value holders*.

Actions

An *action* represents a function from the source code of the underlying program. Each action is represented as a button on the left side of the UI. The program starts with some initial actions which can then enable or disable other actions by executing GUI directives inside the body of the function. These are the only directives that are placed inside function

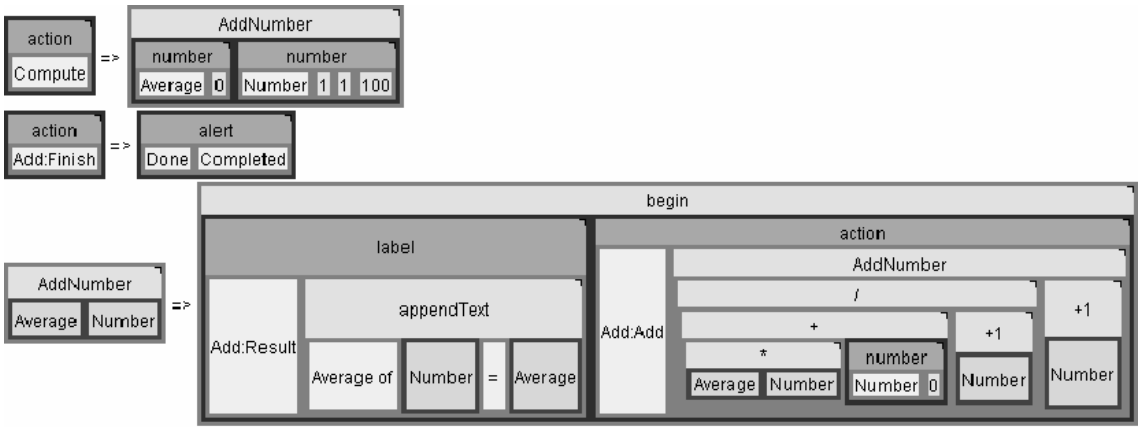


Fig. 2. Program computing an average written in tree rewriting scheme (taken from [7]).

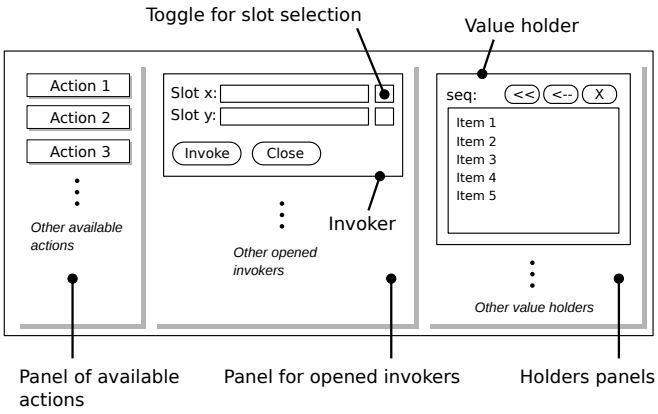


Fig. 3. Design of functionally structured user interface.

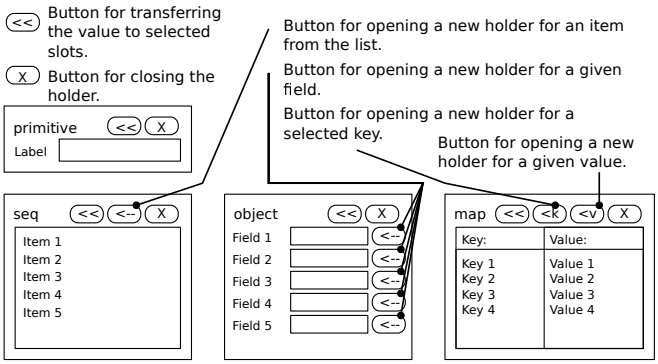


Fig. 4. Four types of holders for corresponding value types.

implementations. The key concept here is that actions do not execute functions right ahead. Instead, they open so called *invokers*.

Invokers

Invoker is a component used to collect inputs for particular action. The user can enter those inputs using *slots* (see figure 3). Slots are represented in the FSUI by text fields. Each slot declares its type so that the user cannot enter invalid values (more about the type system later in this section). The invokers panel in the middle of the UI can contain as many opened invokers as the user wishes. The actual invocation of the represented function is done by clicking on Invoke button. The same function can be executed multiple times since the invoker stays opened until the user closes it. When the function returns a value this value is inserted into so called *value holder*.

Value holders

Value holder displays a graphical component according to the type of value returned by the executed function. There are four types of holders at this time (see figure 4). These types are taken from the type system described later in this section.

Value from a holder can be transferred to a slot with compatible type. The fundamental idea here is that all values in holders are immutable thus they can be changed only by invoking a function on them. This preserves the immutability principle from the functional programming paradigm and makes the UI actually more function oriented than object oriented.

The flow of the program is controlled solely by enabling and disabling actions in the source code. Since all values are immutable, and the user has to execute a function even to create a non-primitive value, the programmer has a great control over what the user can and cannot do at a given moment.

The FSUI is implemented in Clojure in Swing GUI toolkit [15] with help of Seesaw library [27]. This library makes it easier to write GUIs in functional style.

B. The type system

Although the user is limited to create non-primitive values only using functions, he has considerable freedom in what

functions to execute and what values to pass them as arguments. To prevent errors and misunderstandings in the GUI, a function should clearly declare values of what types are expected as input and what is the type of the returned output. As the Clojure language is not statically typed, a type system for annotating functions had to be developed.

These requirements were imposed on the type system

1) The types system should be simple enough so that the user can understand it and be able to relatively easily determine which value can be passed to which function.
2) There is absolutely no need for type inference. All the type information is explicitly written in function annotations.
3) Primitive types as well as composed ones should be implemented. Lists, maps and objects are for now the basic composed types in FSUI. Objects are just heterogeneous key-value structures grouping different kinds of values together.

The type system meeting the above requirements was developed based on [28]. It is used for annotating function bodies in a way that each function is prepended with a type signature. Type signature is a list of type forms specifying value types. The first form in the list specifies the type of the return value and the following forms specify types of the arguments. The grammar for the type signature is shown in listing 2.

Listing 2 Structure of type signature.

```
<signature> ::= (<form> <form-list>)
<form-list> ::= <form> | <form> <form-list>

<form>      ::= <prim> |
                (seq <form>) |
                (map <form> <form>) |
                (object <name>)
<prim>      ::= :bool | :int | :float | :ratio |
                :number | :string | :file
<name>      ::= arbitrary symbol starting
                with a lower case letter
```

C. The generator

As it was previously mentioned, the GUI itself is dynamic which means that the code for the GUI itself doesn't have to be generated. The GUI needs only a description of actions and their respective inputs and outputs. This description is saved in Clojure data structure which is the output of the generator. Input for the generator is annotated source code in Clojure. The generator and the dynamic GUI are both implemented purely in Clojure.

VIII. Discussion

The solution developed in my master thesis [26] and described here contributes a new approach to the field of application prototypes development in functional languages hereby represented by Clojure. The main attributes of this solution, compared to object oriented approaches analysed in section V, are:

1) The UI generation is done using actual source code instead of domain data model.
2) The generator of the FSUI was developed to use more lightweight annotations then the OO solutions usually use.
3) The generated UI is uniform for all generated prototypes.
4) The type system was implemented with simplicity in mind so that the user can understand the GUI quickly.
5) Output of the generation was designed to simplify implementing the FSUI on other platforms.

Possible shortcomings of the whole FSUI and ideas for future development may be:

1) The FSUI is not flexible enough to support specific needs of particular prototypes. More control over the resulting UI layout could be integrated into source code annotations. Specific locations for invokers and value holders may be specified or actions could be nested and produce menu hierarchies.
2) The type system is not flexible enough either since it does not yet support recursive types.
3) This solution is built on top of Lisp-like languages which are dynamically typed. If statically typed languages such as Haskell are used, there will be no need for separate type system and the code annotations could reduce even further.

IX. Conclusion

Focus of this work was on the problem of automated GUI generation for functional languages in relation to application prototyping. Its result is, on the one hand, an analysis of current state in automated GUI generation in object oriented languages, on the other hand, a design and implementation of functionally structured user interface concept using Clojure and Java programming languages.

References

[1] I. Sommerville, *Software engineering*, 9th ed. Pearson, c2011.
[2] J. A. O'Brien and G. M. Marakas, *Management information systems*, 9th ed. McGraw-Hill Irwin, c2009.
[3] M. Smith, *Software prototyping*. McGraw-Hill, c1991.
[4] K. Czarnecki, *Generative programming*. Addison-Wesley, c2000.
[5] J. Nielsen, *Usability engineering*, 1st ed. AP Professional, 1993.
[6] T. Stahl, *Model-driven software development*. John Wiley and Sons, 2006.
[7] J. Jelinek and P. Slavik, "Gui generation from annotated source code," in *Proceedings of the 3rd annual conference on Task models and diagrams*, ser. TAMODIA '04. New York, NY, USA: ACM, 2004, pp. 129–136.
[8] D. S. Touretzky, *Common Lisp: a Gentle Introduction to Symbolic Computation*. Dover Pubns, 2013.
[9] R. K. Dybvig, *The Scheme Programming Language*. The MIT Press, 2009.
[10] R. Hickey, "The clojure programming language," in *Proceedings of the 2008 symposium on Dynamic languages*, ser. DLS '08. New York, NY, USA: ACM, 2008, pp. 1:1–1:1.
[11] B. J. Maclennan, *Functional Programming: Practice and Theory*. Addison-Wesley Professional, 1990.
[12] H. Abelson, G. J. Sussman, and J. Sussman, *Structure and Interpretation of Computer Programs, Second Edition*. McGraw-Hill Science/Engineering/Math, 1996.

[13] L. VanderHart and S. Sierra, *Practical Clojure (Expert's Voice in Open Source)*. Apress, 2010.
[14] N. Shavit and D. Touitou, "Software transactional memory," *Distributed Computing*, vol. 10, no. 2, pp. 99–116, 1997.
[15] J. Elliott, R. Eckstein, M. Loy, D. Wood, and B. Cole, *Java Swing, Second Edition*. O'Reilly Media, 2002.
[16] F. Baader and T. Nipkow, *Term Rewriting and All That*. Cambridge University Press, 1999.
[17] (2013, Feb.) Ajax java framework for rapid application development: Openxava. [Online]. Available: http://www.openxava.org
[18] J. A. Jacko, Ed., *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd ed. CRC Press, 2012.
[19] (2013, Feb.) Naked objects. [Online]. Available: http://nakedobjects.codeplex.com
[20] (2013, Feb.) Roma framework: The new way to conceive web applications. [Online]. Available: http://www.romaframework.org
[21] (2013, Feb.) Google project hosting: Magritte metamodel. [Online]. Available: http://code.google.com/p/magritte-metamodel
[22] (2013, Feb.) Tynamo framework. [Online]. Available: http://tynamo.org
[23] (2013, Feb.) Trails framework. [Online]. Available: http://trails.codehaus.org
[24] (2013, Feb.) Jmatter. [Online]. Available: http://jmatter.org
[25] (2013, Feb.) Apache isis: Domain driven applications, quickly. [Online]. Available: http://isis.apache.org
[26] M. Podloucký, "Automated gui generation for functional data structures," Master thesis, Charles University in Prague, 2012.
[27] (2013, Feb.) Seesaw. [Online]. Available: https://github.com/daveray/seesaw
[28] R. L. Akers, "Strong static type checking for functional common lisp," Doctoral dissertation, University of Texas at Austin, USA, 1995.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 971–976

Assessment of the EPQ probability parameter for scientific articles publishing

Rafał Rumin
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
E-mail: rumin@agh.edu.pl

Piotr Potiopa
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
E-mail: ppotiopa@zarz.agh.edu.pl

***Abstract*—This work presents the analysis of evaluation concerning the articles that are send to publication in academic journals, basing on additional parameters not resulting from essential value of the research work. Currently, majority of article verification algorithms is oriented on the selection of such works that are potentially more strongly influencing the international position of journal. For that purpose, editorial offices, and also reviewers, apply multi-criterion parametric evaluations and accepted parameters have often very subjective character. Presented work makes an attempt to identify used criterion functions i.e. defining evaluation parameters. These parameters were divided onto categories and there was proposed their preliminary verification basing on statistical analysis of already published articles in individual journals. Each parameter has attributed weight function, which allows to defined its impact on the total evaluation of article, and also adaptation of formula to any academic journal. Weight functions will be determined with usage of neural networks or genetic algorithms, aiming to their individual adaptation to particular journal.**

I. Introduction

PREVIOUS investigations over the evaluation of academic journals, cause the continuous improvement of algorithms by which the articles published in these journals are verified. This results from endless aspiration of journals to obtain maximum of points in created rankings (Philadelphia List, Impact Factor, quoting indicators etc.) [2-4].

New appearing methods of the evaluation of journals and modifications of already existing, cause that the essential evaluation of the article can be unsettled in the interest of the parametric evaluation forced by publisher[8-16].

In effect, innovative publications can be inadequately evaluated or not published due to their wrong preparation. Introduced and described below coefficients of scientific articles parametrization are supposed for the task to determine an influence of these subjective parameters on the evaluation of articles in individual journals. Furthermore, there were presented series of factors which, if they will be taken into consideration during writing of scientific articles , have a chance to increase probability of obtaining positive review and in effect the acceptance of publication in renowned journals. In the further process of research works, there is planned realization of automatic information system, which role will be verification of the working version of article, before sending it to the journal and the definition of the probability of obtaining high parametric evaluation. Described parametric evaluation will determine the coefficient EPQ – Estimated Paper Quality. This coefficient will be helpful for scientists who concentrate mainly over essential, and less over the editorial part of their scientific article. The low value of EPQ should induce the author to analyze and supplement his publication before sending article to editorial office of chosen earlier journals.

II. Evaluation of journals

Academic journals are subjected to continuous verification through evaluation of published articles influence on environment of scientists. There exist many parameters evaluating the parametric quality of the journal, here are some of them [20-26]:

- Impact Factor (IF)
- Relative Citation Rates (RCR) / Journal to Field Impact Score (JFIS)
- Article Influence (AI)
- SCImago Journal Rank (SJR)
- Source-Normalized Impact per Paper (SNIP).

Each of them characterizes different factors which influence final evaluation of journals. Different journal evaluation criteria cause the inhomogeneity in resultant rankings. Furthermore, algorithms of evaluation are subjected to continuous changes aiming to the most reliable definition of publications quality. From this reason, the aim of publishing companies, instead of valuing scientific publications having less 'popular' character (though substantially equally good whether even much better), can be wish of achievement of as highest parametric coefficients evaluating other of their publications.

The most popular is Impact Factor (IF) (1). It counts all quoting from particular calendar year, and it divides them by amount of "cited" publications from last two years (C)

978-83-60810-53-8/$25.00

$$IF = \frac{B}{C} \quad (1)$$

Other indicators, although they also reflect the parametric quality rating of journal, are not so popular.

It can be accepted that, as higher evaluation of given journal in the ranking, the article published in it, has a chance to obtain greater range, and consequently receiving greater quantity of quotations. It seems that there exists the conformity of business among the journal and author of the article, however this concerns only wishes of obtaining as maximum quotations quantity at other publishers through the large number of scientists.

Wanting to check our chances for the publication in given journal, often we set incorrect question - *will this journal publish my article?*

To show existing dependences and conflicts of interests between author and editor, one ought to set himself the question:

How my article will help the journal to obtain better position in the ranking (more points in the parametric evaluation of journals)?

The answer is dependent on many factors (the general diagram of dependence among publication publishers and authors is presented on Fig. 1), which can subjectively influence the evaluation of article, aside from its essential value.

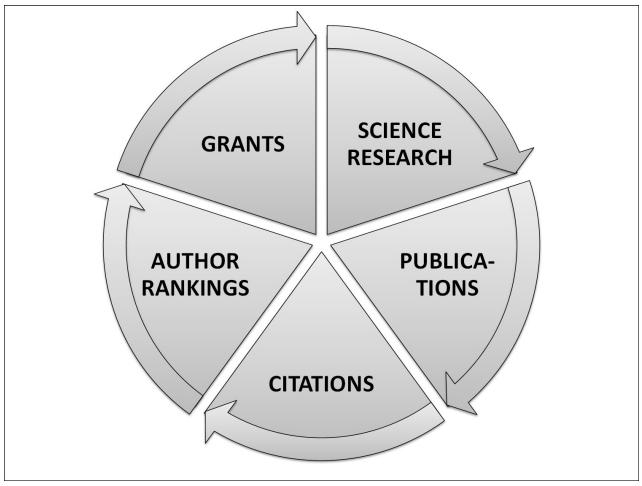


Fig 1. The influence of research work elements on financing and publishing of research

A. The evaluation of Authors

Authors are also subjected to parametric evaluation, targeting the verification of their achievements through the creation of ranking reflecting their contribution to the development of given field. One of main parameters applied in relation to authors of publication is proposed in year 2005 the Hirsch index (h-index) [1]. As easily can be envisaged, such evaluation can be sensitive on manipulation on the side of several cooperating with themselves authors, who mutually will quote their works (aside from their essential contribution into researches). The parametric evaluation of publication issues from the category of scientific research. Scientific researches require financings, and one of the popular sources of learning financing are exploratory grants. To obtain the financing there is expected that the scientist will carry out planned investigations and their effect will have visible influence on given exploratory field. How is measured such influence?

Most readable measure are publications and their quotations. From this reason, scientists who have a suitably high Hirsch index, are treated as trustworthy to commit to them public money on carried researches. Readable dependencies appear between the financing of research, with quantity of publication, and with their quotations which put themselves greater chance for future financial resources.

III. Parametric evaluation of the Article

Every article, except the essential value, can be described by a group of parameters defining its quality from the interest of journal point of view. Here appears mentioned conflict of interests between publishing houses, and authors [17].

In the evaluation of article, the essential value can be estimated by additional parameters defining the range of carried researches, e.g. the article containing theoretical models can classified lower than articles containing, except the theory, also simulation models. As the best will be evaluated articles containing the experimental verification carried researches. Separately, enough high classification can have articles containing rich and complex reviews of the literature from the given field, because this type of articles are quoted often many times. This results from the specific approach of scientists to carried researches and wishes of using elaborated earlier literature review - which often requires a lot of time, and belongs to „little attractive" researches.

Thereby, that at the evaluation of articles value nobody can foresee how often he will be quoted in the future. In the simplification it can be assumed, that in the initial phase of article analysis, each has an evaluation for the essential value on the same level. Since the quantity of elaborated article future quotations cannot be influenced, it can be influenced whom the author quotes in his own publication. This way the quantity of "gained' quotations from the journal's point of view, can be controlled . The issue here is the period of time in which journals are subjected to evaluation in rankings. For the calculation of Impact Factor, there are taken into consideration last 2 years, what means that the auto quotation of other articles which appeared in the same publishing house within a period of last 2 years, have a positive influence on IF indicator increase. Therefore, the publishing house will be willingly promoting art-

icles which already show quotations from their own journal, what is a method to obtaining higher place in the ranking. However, if there exists a group of journals given by the common institution, then cross quotations of other journals belonging to the same publisher are also added value. Here arises a threat regarding the reliability of published articles, because one can apply the mechanism which would permit ranking speculations between journals. Following the paragraphs of this article contain the case study describing such situations.

IV. EPQ - the coefficient of the parametric evaluation of article

To show series of factors participating in the evaluation of given article, there was proposed the EPQ coefficient (Estimated Paper Quality). It can be presented as weighted mean of individual parameters, with suitably assorted weights functions. The value of parameters is standardized so that it contains itself in range from 0 to 1. This type of method descends from Churchman and Ackoff (1954) researches, under the name SAW (Simple Additive Weighting) [18-19]. SAW is one of more popular solutions in MADM type (Multi-Attribute Decision Making) problems of the which undoubtedly is a problem described in work. Elaborated process of EPQ calculation is similar to above methods, however there appear differences in designating of individual parameters. Differences are caused by different way of P_i parameters values determination.

$$EPQ = \frac{1}{n}\sum_{i=1}^{n} P_i * w_i \qquad (2)$$

where P_i is appropriate parameter of evaluation, with following index n appointed , and w_i is weight for given parameter. Below in the table (cf. Table I) there is presented list of parameters together with their asserted values and ranges. All parameters P_i are situated in the same range: $P \in [0,1]$.

V. The example of the EPQ calculation

The definition of the exact value EPQ does not decide about "the success" and the publication of the given magazine article. This will permit however finishing up and improvement of the editorial part which could not take into account mentioned above factors influencing decision of editors and reviewers. Elaborated in such way system using the informatics network , will permit quick definition of the article modification. Outwardly, it will enable based on obtained result EPQ to propose the alternative academic journal which parameters answer to the result. There was calculated the value EPQ for example of publication based on the *Matlab* software.

VI. SEO, Hirsch index and Impact Factor

A. The similarity of the Hirsch Index and Impact Factor to Page Rank, and threats resulting from Black Hat SEO methods?

The growth of the Hirsch index and IF is strictly dependent on the quantity of given author publication quotations. This model can be compared to the published ranking of websites (*PR - Page Rank*) used in Google search engine [5-7]. The similarity refers to the quantity of quotations which correspond to quantities of returnable links indicating given page of data sources.

There are known general methods of influencing the algorithm of search engine in this way, so that the indicated page will be higher in the SERP ranking (Search Engine Results Page). This methods are divided on so called white and black. White Hat SEO - means the positioning of the website in compliance with official guidelines of search engines, what should result in better page adaptation to Webcrawler's and engines of search engines requirements. Good preparation of the website facilitates quick indexing of it in the search engine base of data, however increasing number of valuable references to page (gained naturally and resulting from its popularity and uniqueness) permits its positioning and obtaining of high place in the SERP ranking. As valuable references are acknowledged links from pages about high PR which are often visited by users (e.g. thematic, community websites). There also exists Black Hat SEO which is characterized using all possible gaps in the search engine, for the purpose of raising the ranking of given website. Such effects are achieved through the manipulation with the quantity of returnable links and their "artificial" addition through generating large quantity of pages with links. So many of manipulation methods is the necessity of continuous algorithms change of search and qualitative selection of websites.

From obvious reasons, exact parameters of the algorithm are not revealed for the purpose of their protection before the manipulation. There can be only estimated general dependencies and on their base there can be created algorithms improving the position of website in ranking of searches. Methods of rankings creating e.g. PR and IF, and also H- index, cause the risk of appearing methods taken from SEO, which in the artificial way will manipulate results of mentioned above rankings. Probably there is no possibility of obtaining 100% reliable and objective ranking not burdened with the above risk.

From this reason, the essential evaluation of publication can be shaken, in the interest of the parametric evaluation. This can cause the reverse to intended effect i.e. these rankings will promote less ambitious scientific discoveries, but artificially will overvalue indexes across the elaboration of their manipulation method. Below there is presented case study, which in the mental experiment could result with

TABLE I.
DEFINING PARAMETERS FOR CALCULATION THE EPQ INDICATOR

No.	Parameter P_i	Meaning of value substituted to P_i	Formula on P_i	Range	Initial weight w_i	Range of weight
1	P_1	H – authors Hirsch index	$P_1=(1-\frac{1}{1-H})*(w_1)$	H=[0:inf]	1	[0:1]
2	P_2	I – the quantity of authors indexed publications	$P_2=(1-\frac{1}{1-I})*(w_2)$	I=[0:inf]	1	[0:1]
3	P_3	C – quantity of authors indexed quotations	$P_3=(1-\frac{1}{1+C})*(w_3)$	C=[0:inf]	1	[0:1]
4	P_4	S - degree/ the scientific title of the author (none/engineer/MSc/the doctor/assistant professor/professor)	$P_4=(1-\frac{1}{1+20*S})*(w_4)$	S=[0:5]	1	[0:1]
CONTENT RATING OF ARTICLE						
5	P_5	Gaussian distribution calculated basing on the quantity of all quotations contained by author in the article, where: d - height of the Gaussian curve top, x - quantity of all quotations contained by author in the article, σ - standard deviation of Gaussian distribution, μ - expected value, equal average quantity of quotations devolving on one article in the given journal, a - quotations devolving on one article (k) in the given journal.	$P_5=(d*e^{\frac{-(x-\mu)^2}{2\sigma^2}})*(w_5)$ $\sigma=\sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(x_i-\mu)^2}$ $\mu=\frac{1}{k}\sum_{i=1}^{k}a_i$	d=[1] x=[0:inf] σ=[0:inf] μ=[0:inf] a=[0:inf] k=[0:inf]	1	[0:1]
6	P_6	A - the quantity of quotations coming from archival numbers of the same journal to which the publication is submitted	$P_6=(1-\frac{1}{1-A})*(w_6)$	A=[0:inf]	1	[0:1]
7	P_7	B - the quantity of quotations coming from archival numbers of remaining journals belonging to the same publishing house to which publication is submitted	$P_7=(1-\frac{1}{1+B})*(w_7)$	B=[0:inf]	1	[0:1]
8	P_8	The indicator of the publication originality. O - the quantity of similar articles earlier published by the author. D - the sum of "duplicates", measured by the coefficient of similarity of genuine text and small pictures between previous articles of the author, and with his current publication	$P_8=(\frac{1}{1+O})*(w_8)$ $O=\sum_{i=1}^{n}D_i$	O=[0:inf] D=[0:inf]	1	[0:1]
9	P_9	R_d - the quantity of quoted publications of the current editor of journal to which publication is submitted	$P_9=(1-\frac{1}{1+R_d})*(w_9)$	R_d=[0:inf]	1	[0:1]
10	P_{10}	R_c - the quantity of quoted publications of current reviewer of journal to which publication is submitted	$P_{10}=(1-\frac{1}{1+R_c})*(w_{10})$	R_c=[0:inf]	1	[0:1]
OTHER PARAMETERS						
11	P_{11}	J - the quantity of authors publications quoted by current editor or reviewer of the journal to which publication is submitted	$P_{11}=(1-\frac{1}{1+J})*(w_{11})$	[0:inf]	1	[0:1]
12	P_{12}	K - the quantity of authors common publication articles and current editor or reviewer of journal to which publication is submitted	$P_{12}=(1-\frac{1}{1+K})*(w_{12})$	[0:inf]	1	[0:1]
13	P_{13}	Z - quantity of elements from the range carried researches (the form of survey): review, theory, model, simulation, experiment, lack/other.	$P_{13}=(1-\frac{1}{1+20+Z})*(w_{13})$	[0:5]	1	[0:1]

"artificial" increasing of IF for the journal, or with 'artificial" increasing of the H-index for given scientist.

B. *How to create Journal with IF=100 (Case Study I)?*

In the after-mentioned mental experiment we establish that one publishing house can belong to several academic journals having similar character, or there was undertaken cooperation between publishing houses for the purpose of one common journal strong promotion. The first journal

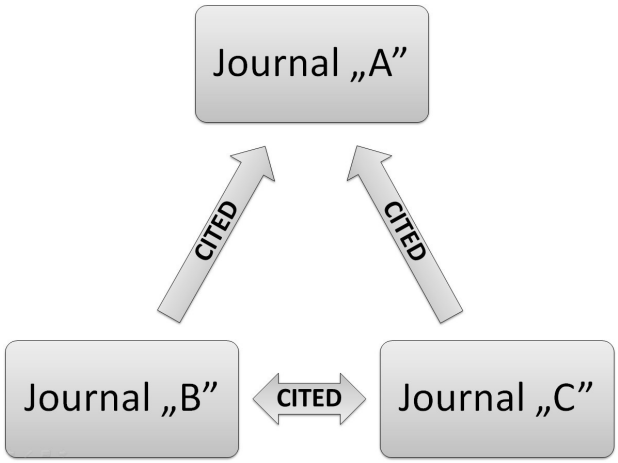


Fig 2. The diagram of quotations hinterland building for promoted journal.

„A" will be promoted, however remaining „B", „C" will constitute the hinterland with place for journal "A" quotations.

In this case arises classical system of hinterland with links (in our case - with quotations), known among computer scientists dealing with the positioning of websites (so called SEO - Search Engine Optimization). The example of building of quotations hinterland is presented on "Fig. 2". On the assumption that the journal "A" will have few articles in one publishing-cycle , then remaining journals can force writing for them authors, to quote several articles from journal "A". So extortionate ranking of promoted journal can have other advantages.

VII. The methodology of designating weights

Particularly essential from the usage of EPQ indicator point of view, is the possibility of weights definition w_i in way compatible to parametric evaluations applied by the given journal. The large number of academic journals causes different approach to the parametric evaluation of accepted to editorial office and the review of article. Basing on the data from previous years, considering all publications printed within the framework of one publishing-title , we are able to determine weights of individual parameters individually for the given journal.

For that purpose we will use neural networks with the feedback which will learn to recognize the influence of given parameter on the positive acceptance of article to the publication. In case of the analysis, already printed publications, we will subordinate the quantity of published articles from the value of individual parameters. The more articles will have e.g. the high parameter P6, the greater influence on the printing of publication has the quantity of archival articles quotations laded from the same journal.

VIII. Application realizing EPQ designating

For the purpose of individual parameters designation, we will use the access to individual databases, among others: SCOPUS, WEB OF KNOWLEDGE, and others. Application in the first instance will collect data: ref. of author, quotations, journal, publication, and then made the evaluation of parametric sent publication. Based on this evaluation, it can propose suggestions ref. introductions of changes in the article, or present proposal of the alternative journal to which the parametric evaluation was in order better. The system architecture may be built based on the client-server methodology what is presented on Fig. 3.

IX. The architectural schema

On the after mentioned Fig. 3 there is presented the general architectural schema of the system.

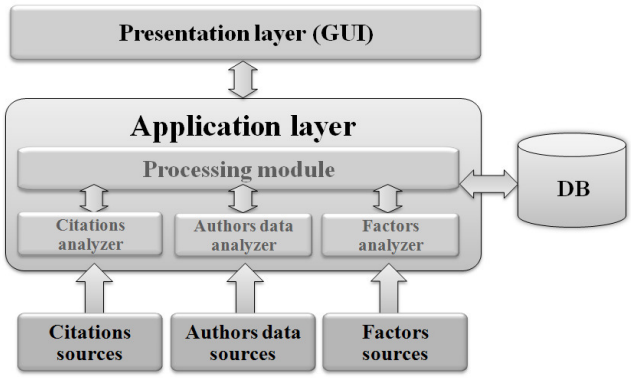


Fig 3. The architecture of proposed information system designating the EPQ coefficient.

In presented architecture system we distinguish:

1. *Presentation layer* - layer of the application responsible for the presentation of results and communication with user, receiving data from user (proposed article, survey for the author)
2. *Application layer* - layer responsible for the resumption of data and processing of results, consists of:
   - *citations analyzer* (module processing the quotation categorizing and counting quotations of authors works.

- *authors data analyzer* (module processing data of authors (also reviewers and editors), checking relations of author with journals across quotations as well as categorizing his achievement)
- *factors analyzer* (module being supposed for the task to process available data sources information of used in the algorithm coefficients for journals and authors)

1. *DB* - layer of database recording source data and results of calculations with application layer, permitting caching of data sources in the situation when data don't need to be refreshed at every operation of weight-coefficients calculation weight- coefficients.
2. *Sources* - layer of gaining data from chosen sources dividing into sources of quotations gaining ("citations sources"), given authors ("authors the date sources") and coefficients used in the algorithm of EPQ count ("factors sources").

The system architecture in case of further development can be calibrated because the module of processing may receive partial results of calculations (weights of component parameters) from individual modules which can be find on separate instances of servers. Every module of gaining data can have the separate database in which will store received results of the data sources indexing In case of presentation layer, the system can communicate with software of the *thin client* type in case of approach users (authors of articles) and with the software of the *fat client* type in case of the administrator who can control work of the processing module (settings control).

III. Conclusion

As this is the elaboration of the preliminary concept of articles parametric evaluation across proposing of the EPQ parameter , only verification on figures will permit the definition of its real effectiveness in the classification of articles to individual academic journals. Methodology is based on foundations that the substantially good article can be worse evaluated, due to remaining factors on which reviewers and editors of journals pay attention. Elaborated system, targets proper verifying and correction of article before delivering to publishing houses. This will permit to carry essential research on equally high level, and to regard of subjective 'expectations' from the side of publishing house in relation to the author. So improved article has greater chances for printing in the renowned journal, what can positively rebound on future publications of many authors.

References

[1] J. E. Hirsch, “An index to quantify an individual’s scientific research output.” *Proc. Nat. Acad, Sci.* (PNAS), 2005, vol. 102, nr 46, s. 16569-16572.

[2] L. Egghe, R. Rousseau, “Introduction to informetrics”. Amsterdam: Elsevier, 1990.

[3] H. F. Moed, M. Vriens, “Possible inaccuracies occurring in citation analysis”. *Journal of Information Science*, 1989, 15, 95-107.

[4] R. Todorov, W. Glanzel, “Journal citation measures: A concise review”. *Journal of Information Science*, 1988, I4. 47-65.

[5] S. Brin, L. Page. “The anatomy of a large-scale hypertextual web search engine”. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, pages 107–117. Elsevier Science Publishers B. V., 1998.

[6] L. Page, S. Brin, R. Motwani, T. Winograd, “The PageRank citation ranking: Bringing order to the web”. Technical Report, Stanford Digital Library Technologies Project, 1998.

[7] S. Maslov, S. Redner, “Promise and Pitfalls of Extending Google’s PageRank Algorithm to Citation Networks”, *Journal of Neuroscience* 2008, 28, 11103.

[8] D. Zhou, S. A. Orshanskiy, H. Zha, C. L. Giles, “Co-Ranking Authors and Documents in a Heterogeneous Network”, *INTERNATIONAL CONFERENCE ON DATA MINING*, 2007.

[9] P. Chen, H. Xie, S. Maslov, S. Redner. “Finding Scientific Gems with Google”. *J.INFORMET.*, 1:8, 2007.

[10] E. Garfield. “Citation analysis as a tool in journal evaluation”. *Science*, 178(60):471–479, November 1972.

[11] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel. “Coauthorship networks in the digital library research community”. *arXiv.org:cs/0502056*, 2005.

[12] M. Bianchini, M. Gori, and F. Scarselli. “Inside pagerank”. *ACM Trans. Inter. Tech.*, 5(1):92–128, 2005.

[13] De Moya, F. “The SJR indicator: A new indicator of journals’ scientific prestige”, 2009, Arvix. (arxiv.org/abs/0912.4141)

[14] E. Garfield, “The history and meaning of the journal impact factor”. JAMA 2006;295:90-3.

[15] M. Amin, M. Mabe, “Impact factors: Use and Abuse”. *Perspectives in Publishing, 2000 - October, 1,* 1-6

[16] Mu-Hsuan Huang, Wen-Yau Cathy Lin, "The influence of journal self-citations on journal impact factor and immediacy index", *Online Information Review*, 2012, Vol. 36 Iss: 5, pp.639 – 654

[17] C. Wenneras, A. Wold ”Nepotism and sexism in peer-review”. *Nature*. 1997 May 22;387(6631):341-3.

[18] C.W. Churchman, R.L. Ackoff, “An approximate measure of value”. *Journal of the Operational*, *Research Society of America*, 1954, **2**(2), 172–187.

[19] D. Widayanti , S. Oka, S. Arya, “Analysis and Implementation Fuzzy Multi-Attribute Decision Making SAW Method for Selection of High Achieving Students in Faculty Level” *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 1, No 2, January 2013, ISSN 1694-0784

[20] SNIP and SJR at Journal Metrics (www.journalmetrics.com)

[21] SCImago Journal Rank (SJR) (http://www.scimagojr.com/)

[22] Source-Normalized Impact per Paper (SNIP) (www.journalindicators.com)

[23] Impact Factor (IF) (http://thomsonreuters.com/products_services/science/free/essays/impact_factor)

[24] h-index (http://help.scopus.com/robo/projects/schelp/h_hirschgraph.htm)

[25] Article Influence (AI) (www.eigenfactor.org)

[26] Relative Citation Rates (RCR)/Journal to Field Impact Score (JFIS)

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 977–980

Fuzzy Multi-attribute Evaluation of Investments

Bogdan Rębiasz, Bartłomiej Gaweł and Iwona Skalna
AGH University of
Science and Technology
Krakow, Poland
Email: {brebiasz, bgawel, iskalna}@zarz.agh.edu.pl

***Abstract*—Most companies have a large number of projects that they would like to do for various reasons. However, usually there is never enough time and money available to complete all of them. Selecting a portfolio from available project proposals is crucial for the success of each company. This paper proposes a practical framework for modelling projects portfolio selection problem with fuzzy parameters resulting from uncertainty associated with decision makers' judgment. A fuzzy multi-attribute decision-making approach is adopted. A two-step evaluation model that combines fuzzy AHP (*Analytic Hierarchy Process*) and fuzzy TOPSIS (*Technique for Order Preference by Similarity Ideal Solution*) methods is used to rank potential projects. The proposed approach is illustrated by an empirical study of a real case from steel industry involving five criteria and ten projects.**

I. Introduction

DECISIONS on investment projects have a direct impact on a company's success. They are, however, particularly difficult, because of the ubiquitous uncertainty associated with any business activity. This causes that the project portfolios selection (PPS) becomes an increasingly complex decision task, which in turn motivates managers to utilise modern techniques and tools to optimise capital allocation.

At present, there are a lot of methods that can be applied to solve PPS problems, including Economic Analysis, Decision Theory, Optimisation and Multi-criteria methodologies. In order to deal with both financial and non-financial project attributes, the multi-criteria decision making (MCDM) analysis is a preferred approach. The goal of the multi-criteria decision making analysis is to "provide a set of attributes aggregation methodologies that enable the development of models considering the decision makers' (DMs') preferential system and judgement policy" [7]. In general, MCDM methods may be divided into two groups: multi-objective decision making (MODM) and multi-attribute decision making (MADM). The latter have been used to solve problems with discrete decision choices and a predetermined or limited number of alternative choices. A comparative study on various MCDM methods is presented, e.g., in [1] and [8].

In this paper, an MADM approach to project portfolio selection is applied. The classical approach is expanded to deal with uncertainty expressed in the form of fuzzy numbers. There is a range of scientific publications which develop very sophisticated methods for describing uncertainty. Meanwhile, according to the survey of Hubbard [9], modern enterprises still asses and mitigate risk using old fashioned methods which have not evolved much for several decades. This paper attempts to fill out the gap between the theory and practise. A practical framework to deal with the PPS problem is developed. The reminder of this paper is organised as follows. Section II briefly describes problems with modelling uncertainty in PPS. Section III presents methodology used to solve the PPS problem. The proposed framework is described in Section IV. Numerical example is shown in Section V. The paper ends with concluding remarks.

II. Risk and uncertainty in project portfolio selection

There is no universally accepted definition of business risk and uncertainty, but in the PPS context they may be understood as potential problems with availability and certainty of information, and also imprecise choices. To deal with uncertainty in PPS, it must be first noted that PPS usually consists of two stages. In the first stage, projects are selected on the basis of the threshold criteria which are determined by decision-makers. In order that a project can pass to the next stage, it must strictly fulfil these criteria. The selected projects are input for an MADM method, which usually first calculates the weights of criteria, and then determines the ranking of potential projects. Each part of an MADM method is associated with different type of uncertainty. The main source of uncertainty in determination of criteria weights is imprecision of expert judgements. Due to cognitive biases, decisions may be deviated from a standard of rationality or good judgement. To take into account these systematic errors, fuzzy numbers are used instead of crisp numbers.

In this paper, fuzzy criteria weights are obtained using a fuzzy AHP method. Some researchers believe that classical Saaty's AHP method has some weaknesses which are connected with uncertainty. In [18] authors points out that mapping experts judgement to crisp numbers and cognitive biases generates uncertainty which is not taken into account by the classical AHP method and may have huge impact on the results. To deal with this problem, some researches fuzzified AHP (e.g., [4] has considered trapezoidal fuzzy intervals for comparison ratios in AHP and [5] has proposed approach for triangular case).

Some criticise the fuzzy AHP and argue that it does not give much different results then the crisp version. However, it is important to note that the main criticism is based on assumption that comparison ratios are based on expert consensus. In practise, comparison ratios are usually averages of expert

978-83-60810-53-8/$25.00

ratios. As long as there are no agreement between experts, everyone of them interprets linguistic variables in different ways. In this situation, expert verbal possibilities should better be translated into fuzzy than crisp numbers.

The second aspect of representing model uncertainty by fuzzy numbers concerns the type of fuzzy numbers that should be used. Generally, there are two approaches to fuzzification of comparison ratios – fuzzy numbers [19], [21] and fuzzy intervals [13]. The empirical study shows [3] that membership functions of numerical equivalents of linguistic terms are similar to fuzzy numbers, which are not distributed equidistantly along the possibility scale and which vary considerably in symmetry and vagueness.

Usually, after the first stage, the calculated criteria weights are defuzzified. In the proposed approach, fuzzy weights are passed to the second stage. This guaranties that uncertain judgements of decision-makers are taken into account also during the second phase of PPS.

The second phase of PPS determines the ranking of potential projects based on the weights obtained in the first stage. In this phase, uncertainty concerns attributes of alternatives. The attributes are divided into two groups: objective (numerical) and subjective (linguistic). A majority of authors argue that only subjective criteria should be described in terms of fuzzy numbers. In the proposed approach it is assumed that quantification of financial attributes of investment projects should be modelled as mixture of possibility and probability distribution.

III. Methodology

The proposed methodology of selecting an efficient portfolio of investment projects consists of the following steps. First, multiple criteria that are considered in the decision-making process are identified. Then, criteria weights are calculated according to the fuzzy AHP methodology. After constructing the relationship of a criteria decision matrix, the fuzzy TOPSIS approach is used to achieve the final ranking results.

The AHP method ([16]) is a flexible MCDM tool for complex problems where both qualitative and quantitative aspects need to be considered ([2]). The AHP integrates different measures into a single overall score for ranking alternatives. By reducing complex decisions to a series of simple, pairwise comparison judgements, then synthesising the results, the AHP not only helps the analysts to arrive at the best decision, but also provides a clear rationale for the choices made [5]. The fuzzy AHP [5] is the fuzzy extension of AHP to deal with the fuzziness of the data involved in the decision making process. Fuzzy AHP enables decision makers to specify preferences in the form of natural language expressions about the importance of each performance attribute.

TOPSIS [10] is another popular approach to MCDM. The main idea is that the best alternative should have the shortest distance from the (positive) ideal solution and the farthest distance from the negative ideal solution. The TOPSIS method has also been extended in different ways to deal with fuzzy numbers. The simplest one is to change fuzzy MCDM into a crisp one by using defuzzification. This approach, however, can lead to the loss of information. Another approach is to define a crisp Euclidean distance between fuzzy numbers. An approach based on α-cuts can also be found in the literature [20].

IV. Proposed framework for project portfolio selection

Based on the methodology described in Section III, a new approach to project portfolio selection problem is proposed. It consists of four stages described in the following subsections.

A. Identification of available investment projects and criteria

First, a committee of decision-makers who come from different managerial levels is formed. They identify m potential investment projects $A_1,\ldots,A_m$ and n criteria $C_1, \ldots, C_n$ that each project must fulfil. To properly assess a project, many factors should be considered [14], [15]. McKown and Mohamed [12] presents multi-criteria project selection where uncertainty of profitability parameters is described by fuzzy numbers. They point out that the selection of investment projects should consist of two kinds of parameters – financial (e.g., net present value) and non-financial (e.g., social, environmental, strategic an organisational). The method proposed in this paper allows to aggregate financial and non-financial indicators. First, the criteria are divided into two groups – objective and subjective. Objective criteria are described by fuzzy numbers which usually result from fuzzy modelling or aggregation of historical data. Subjective criteria are qualitative criteria with values that are specified by decision makers in the form of linguistic variables. Here, linguistic variables are transformed into fuzzy numbers (triangular or trapezoidal). This simplifies further ranking of projects.

B. Calculation of synthetic importance weights

Obviously, the problem of calculating the importance weights of the criteria is a typical multi-variable and multi-objective optimisation problem. To calculate importance weights of the criteria the fuzzy AHP is used. To make a pairwise comparison, a linguistic scale is developed. Table I provides summary of translation developed based on [3]. The final scores of criteria are also represented by fuzzy numbers.

C. Development of performance ratings for projects

The performance ratings of objective and subjective parameters are calculated. At the end of this stage the threshold selection is made. Only those projects which have passed the threshold selection are taken into account in the next stage of the PPS.

D. Calculating hierarchy of projects using fuzzy TOPSIS

The hierarchy of projects is established. Then, the overall ranking of projects is calculated. The ranking allows a decision-maker to select the most appropriate investment option.

Projects	Capital investment	C1			C2					C3		C4				C5
	000's PLN	C1.1	C1.2	C1.3	C2.1	C2.2	C2.3			C3.4	C3.5	C4.1	C4.2	C4.3	C4.4	C5.1
							C2.3.1	C.2.3.2	C.2.3.3							
P1	(135,150,165)	(-66 083.9, 14 376.1, 127 533.0,225 122.1)	(-5.1,4.5, 9.3,12.1)	2.1	middle	stability	average	average	external	widely used	neutral	without growth	no impact	positive	avaliable	one
P2	(216,240,264)	(-1 334 440.0, 628 247, 493 130.0, 1 757 191.0)	(-4.3,-1.0, 3.3,7.4)	4.5	big	stability	average	average	well-developed	widely used	increase	moderate growth	no impact	nautral	need to train	one
P3	(1 270,1 430,1 590)	(-1 00 0694.0, -24 717, 659 279.1, 1779 179.1)	(-2.3,-0.9 , 2.3,5.9)	5.2	big	stability	average	average	well-developed	widely used	increase	moderate growth	no impact	positive	need to train	more than two
P4	(1 600,1 780 ,1 995)	(-258 273.0, -111 259.0, 202 888.0, 5 023 315.0)	(-1.1,-0.8 , 1.3,6.2)	4.6	big	stability	average	average	well-developed	widely used	large increase	moderate growth	no impact	positive	need to train	more than two
P5	(375, 410, 450)	(-356 417.0, -140 463.0, 291 199.4, 785 944.0)	(-3.3,-0.8 , 2.4, 7.9)	3.2	big	growth	best	below average	well-developed	widely used	large increase	moderate growth	improve condition	neutral	need to train	more than two
P6	(465, 515, 565)	(-368 432.0, -102371.0, 560 595.0, 1 123 142.0)	(-3.0,-0.5 , 2.9, 7.8)	3.1	big	growth	average	average	well-developed	widely used	increase	moderate growth	degradation	very positive	need to train	more than two
P7	(125, 138 , 150)	(-102 722, -50 513.0, 245 279.2, 391 245.1)	(-1.7 ,-0.8, 5.9, 9.1)	2.1	big	growth	best	average	well-developed	highest	increase	moderate growth	degradation	very positive	need to train	more than two
P8	(170, 190, 210)	(-144 482, -70 168.0, 235 800.5, 448 462.5)	(-2.8, -0.6 , 4.9, 7.9)	2.3	big	growth	best	average	well-developed	highest	increase	moderate growth	degradation	very positive	need to train	more than two
P9	(250, 320, 350)	(-137 252.0, 23 797.3, 129 599.0, 301 074.2)	(-3.8, 1.4 , 5.9, 6.9)	3.3	small	stability	best	below average	no network	highest	large increase	moderate growth	degradation	positive	need to train	more than two
P10	(20, 22, 24)	(-12 344.0, 4 322.1, 16 123.0, 28 144.0))	(-4.1, 2.2, 5.9, 7.6)	4.1	big	small	average	lower	no network	widely used	neutral	moderate growth	no impact	neutral	difficulties in recruiting	one

Fig. 1. Description of the projects alternative

TABLE I
COMPARISON OF RELATIVE IMPORTANCE OF CRITERIA FOR FUZZY AHP

Linguistic terms	**Crisp intensity of importance**	**Fuzzy intensity of importance**
Equally important	1	$(1, 1, 1)$
Moderately more important	3	$(1, 3, 5)$
Strongly more important	5	$(2, 5, 6)$
Very strongly more important	7	$(6, 7, 8)$
Extremely more important	9	$(8, 9, 9)$

TABLE II
PAIRWISE COMPARISON MATRICES

	Criteria				
	C1	C2	C3	C4	C5
C1	(1,1,1)	(2,5,6)	(8,9,9)	(8,9,9)	(8,9,9)
C2	(0.17,0.2,0.5)	(1,1,1)	(1,3,5)	(1,3,5)	(1,3,5)
C3	(0.11,0.11,0.13)	(0.2,0.34,1)	(1,1,1)	(1,1,1)	(1,1,1)
C4	(0.11,0.11,0.13)	(0.2,0.34,1)	(0.2,1,1)	(1,1,1)	(1,1,1)
C5	(0.11,0.11,0.13)	(0.2,0.33,1)	(0.2,1,1)	(0.2,1,1)	(1,1,1)

TABLE III
PAIRWISE COMPARISON MATRICES - SUBCRITERIA

	Sub-criteria			
	C1.1	C1.2	C1.3	
C1.1	(1,1,1)	(1,1,1)	(6,7,8)	
C1.2	(1,1,1)	(1,1,1)	(6,7,8)	
C1.3	(0.13,0.14,0.17)	(0.12,0.14,0.17)	(1,1,1)	
	C2.1	C2.2	C2.3	
C2.1	(1,1,1)	(0.12,0.14,0.17)	(1,3,5)	
C2.2	(6,7,8)	(1,1,1)	(0.16,0.2,0.5)	
C2.3	(0.2,0.33,1)	(2,5,6)	(1,1,1)	
	C3.1	C3.2		
C3.1	(1,1,1)	(6,7,8)		
C3.2	(0.13,0.14,0.17)	(1,1,1)		
	C4.1	C4.2	C4.3	C4.4
C4.1	(1,1,1)	(0.17,0.2,0.5)	(0.2,0.33,1)	(1,3,5)
C4.2	(2,5,6)	(1,1,1)	(1,3,5)	(8,9,9)
C4.3	(1,3,5)	(0.2,0.333,1)	(1,1,1)	(8,9,9)
C4.4	(0.2,0.33,1)	(0.11,0.11,0.13)	(0.11,0.11,0.13)	(1,1,1)

TABLE IV
PAIRWISE COMPARISON MATRICES - ATTRIBUTES

	Attributes		
	C2.3.1	C2.3.2	C2.3.3
C2.3.1	(1,1,1)	(1,3,5)	(1,1,1)
C2.3.2	(0.2,0.333,1)	(1,1,1)	(1,3,5)
C2.3.3	(1,1,1)	(0.2,0.333,1)	(1,1,1)

V. NUMERICAL EXAMPLE

The proposed approach was applied for PPS in steel industry. There are five criteria $C1, \ldots, C5$ – financial, market, technology and environment, staff and compliance with the company's strategic objective. Each of them is divided into subcriteria. The objective ones are NPV, IRR, Pay-back period, the rest is subjective. There is also the third level of subcriteria for the $C2$ criterion. They are called attributes.

To calculate weights of criteria, a team of decision makers make pairwise comparison. The results of this comparison are presented in Tables II, III and IV. Then, using the fuzzy AHP global priorities are obtained (Table V). The priorities are presented in terms of fuzzy numbers. It can be noticed that the higher hierarchy of the criteria is, the wider fuzzy number are. For example, fuzzy weight $C2.3.1$ range between 0 to nearly 0.3. This illustrates the well-known phenomenon of accumulation of uncertainty. That is why in next step the consistency degree should be used (e.g., fuzzy preference

TABLE V
IMPORTANCE WEIGHTS OF INDIVIDUAL REQUIREMENTS

	Weight		Weight
C1	(0.429,0.633,0.917)	C3.1	(0.039,0.056,0.124)
C2	(0.073,0.175,0.372)	C3.2	(0.006,0.008,0.018)
C3	(0.051,0.064,0.122)	C4.1	(0.003,0.007,0.039)
C4	(0.051,0.064,0.122)	C4.2	(0.013,0.036,0.145)
C5	(0.051,0.064,0.122)	C4.3	(0.008,0.019,0.085)
C1.1	(0.181,0.292,0.471)	C4.4	(0.001,0.003,0.013)
C1.2	(0.181,0.292,0.471)	C5.1	(0.051,0.064,0.122)
C1.3	(0.025,0.042,0.072)	C2.3.1	(0.002,0.027,0.313)
C2.1	(0.006,0.05,0.307)	C2.3.2	(0.001,0.02,0.264)
C2.2	(0.018,0.061,0.204)	C2.3.3	(0.002,0.015,0.127)
C2.3	(0.011,0.063,0.31)		

TABLE VI
FINAL RANKING OF PROJECTS

Project	Rank	Project	Rank	Project	Rank
P9	0.7154	P3	0.6817	P5	0.6770
P10	0.7101	P7	0.6789	P6	0.6714
P1	0.7095	P8	0.6782	P2	0.6615
P4	0.7011				

programming).

In the next step, evaluation matrix is created. Matrix consists of 15 criteria and 10 projects ($P1, \ldots, P10$). The objective criteria are characterised by fuzzy intervals. The level of subjective criteria are specified by experts. The subjective criteria are translated into triangular fuzzy numbers.

In the presented example, there are two kinds of subjective attributes – some of them describe patterns, and some of them judgements. Market size criterion $C2.1$ and prospects for market growth criterion $C2.2$ belong to first group. They describe the belief of decision maker that market for project i will behave in accordance with some pattern. For example, pattern *stable* means the dynamic of the market growth which may be described by the fuzzy number $(-1.02, 0, 1.02)$.

The second group that is subjective criteria represents judgements of experts. Therefore, they are treated as ordinal fuzzy variables. Since all of subjective criteria are ordinal (variable with order), thus fuzzy ordinal rank transformation is used. After translation of linguistic variables – the fuzzy TOPSIS is applied. The obtained final ranking of projects is presented in Table VI.

VI. CONCLUSION

The evaluation and selection of industrial projects is one of the most important aspects of PPS. This paper proposed a combined fuzzy MADM approach based on fuzzy AHP and fuzzy TOPSIS techniques. A real world case study from steel industry was presented to explain approach. The paper introduced fuzzy decision making concept, when some data is burden with uncertainty. It is argued that if a fuzzy MADM problem is defuzzified into crisp one to early, then the advantage of modeling uncertainty becomes negligible. The rational approach is to defuzzify imprecise values at the very end of methods. Based on this argument, we perform deffuzzification at the very end of MADM method during calculate weight of criteria.

More research is needed to examine projects interaction and dependency. Further research is also required with respect to the subjective criteria of project selection. The problem of quantifying the qualitative factors remains a difficult and sometimes controversial tasks.

REFERENCES

[1] N.P. Archer and F. Ghasemzadeh, *An Integrated framework for project portfolio selection*, International Journal of Project Management, 7(4), 207–216, 1999.

[2] M. Bevilacqua and M. Braglia, *The analytic hierarchy process applied to maintenance strategy selection*, Reliability Engineering & System Safety, 70(1), 71–83, 2000.

[3] F. Bocklisch and S.F. Bocklisch and J.F. Krems (n.d.), *How to Translate Words into Numbers? Fuzzy Approach for the Numerical Translation of Verbal Probabilities*, Lecture Notes in Computer Science Volume 6178, 614–623, 2010.

[4] J.J. Buckley, *Fuzzy Hierarchical Analysis*, Fuzzy sets and systems, 17(3), 233–247, 1985.

[5] D.-Y. Chang, *Applications of the extent analysis method on fuzzy AHP*, European Journal of Operational Research, 95, 649–655, 1996.

[6] C.T. Chen, *Extension of the TOPSIS for group decision-making under fuzzy environment*, Fuzzy Sets and Systems, 114, 1–9, 2000.

[7] M. Doumpos and C. Zopounidis, *Multiattributes decision aid classification methods*, Boston, Kluwer Academic Publishers, 2002.

[8] J. Figueira and S. Greco and M. Ehrgott, *Multiple attributes decision analysis: State of the art surveys*, New York: Springer, 2005.

[9] D.W. Hubbard, *The Failure of Risk Management: Why It's Broken and How to Fix It*, John Wiley and Sons, 31–35, 2009.

[10] C.L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, New York, Springer-Verlag, 1981.

[11] M. Kordi, *Comparison of fuzzy and crisp analytic hierarchy process (AHP) methods for spatial multicriteria decision analysis in GIS*, Diss. University of Gävle, 2008.

[12] S. Mahomed, A.K McKown, *Modelling project investment decisions under uncertainty using possibility theory*, International Journal of Project Management 19, 231–241, 2001.

[13] P. Liu and Y. Su, *The Extended TOPSIS Based on Trapezoid Fuzzy Linguistic Variables*, Journal of Convergence Information Technology, 5(4), 38–53, 2010.

[14] B. Rebiasz, *Fuzziness and randomness in investment project risk appraisal*, Computers & Operations Research, 34(1), 199–210, 2007.

[15] B. Rebiasz, *Selection of efficient portfolios-probabilistic and fuzzy approach, comparative study*, Computers & Industrial Engineering, Perganon, 2013 (in print).

[16] T.L. Saaty, *The analytic hierarchy process*, New York, McGraw-Hill, 1980.

[17] T.L. Wand, *Fuzzy discounted cash flow analysis*, In: Evans GW, Karwowski W, Wilhelm MR (Eds.) Applications of Fuzzy Sets Methodologies in Industrial Engineering, Elsevier, 91–102, 1989.

[18] Ch.Ch. Yang and B.Sh. Chen, *Key Quality Performance Evaluation Using Fuzzy AHP*, Journal of the Chinese Institute of Industrial Engineers, 21(6), 543–550, 2004.

[19] S. Mahmoodzadeh and J. Shahrabi, *Project selection by using fuzzy AHP and TOPSIS technique*, International Journal of Humanities and Social Sciences, 1(3), 135–140, 2007

[20] Y.-M. Wang and T.M.S. Elhag, *Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment*, Expert Systems with Applications, 31, 309–319, 2006.

[21] J. Wang and K. Fan and W. Wang, *Integration of fuzzy AHP and FPP with TOPSIS methodology for aeroengine health assessment*, Expert Systems with Applications, 37(12), 8516–8526, 2010.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 981–984

Increase in the Competitiveness of SMEs using Business Intelligence in the Czech-Polish border areas.

Milena Tvrdíková
VŠB-Technical University of
Ostrava, 17.listopadu 15/2372, 708
00 Ostrava, Czech Republic
Email: milena.tvrdikova@vsb.cz

***Abstract*—The paper addresses tools that support knowledge-based management. These tools are referred as Business Intelligence. Competitive Intelligence is also discussed. They transport data into comprehensive information about company processes and real-world impact on their progress. Basic principles of these applications are also described. The paper presents the results of the questionnaire survey conducted among SMEs in the Czech-Polish border area. The questions are focused on the use of these tools and intensions in cloud exploitation. The results and the lack of show a considerable interest in the use of Business Intelligence applications awareness of business managers with regard to Cloud computing possibilities.**

I. Introduction

GLOBAL economic models are undergoing profound changes which include e.g. global competition, global transfers and changes in the way of work in many industries or the pressure on the qualification of the workforce. In developed countries, this causes a gradual shift to an economy of intangible assets and relationships [1]. These changes are underpinned by the key role of information and communication technologies (ICT) providing modern infrastructure, which allows implementing most of the changes. Simultaneously, ICT provides tools to increase performance, competitiveness and innovation in virtually all areas of the economy.

Managers currently have access to a considerable amount of data available from a variety of sources. To make decisions, they need an efficient tool that would help them process data to obtain necessary information. Such tools are Business Intelligence (BI) applications that specifically aim to provide decision support to managers. These tools allow them to analyses the situation in their own company, provide access to information from external company environment and based on the processed data, provide the managers with early warning of existing threats and highlight new business opportunities [2].

At the same time, the range of Competitive Intelligence (CI) tools and methods is expanding. The CI tools and methods introduce into management sophisticated methods of work for contextual search in external information sources through the Internet [3].

New opportunities for managers are also brought about by the new Cloud Computing (CI) technologies provided by ICT companies as a service. While part of company managers is aware of the existence of CI, the majority are still distrustful of it or have no idea what the term means.

II. Increasing the Entrepreneurial Potential of Using ICT in the Czech-Polish Border Area

Since 2012, the Department of Applied Informatics, VSB-TU Ostrava has been conducting a research "Application software for decision support in small enterprises in the Czech-Polish border area". This research is conducted jointly by the Department of Economic Informatics, University of Economics in Katowice.

The survey results also confirmed that the adoption of ICT present an adaptive challenge, not a technical problem. It provides the SMEs with several advantages, especially at the tactical and operational level of management.

- At the level of operational management: an increase in the quality of data management, communication, decision making, data exchange, improving cash-flow control, gradual transition to work with digital documents, shortening the response time to queries (improving customer relations, company profile).
- At the tactical level: rapid response to changes, promotion of teamwork, more flexible scheduling, flexible processing of offers, better integration of business processes, and overall improvement in efficiency and effectiveness.

Although information is today recognized as an important means of creating added value, there is a lack consistency between ICT and business. The basis to solve these problems can be found in the balance between the global and information strategies of the company. These strategies must be prepared on the same basis and with the same weight (attention). The truth is that most SMEs typically do not pay attention to the information strategy – creating IS and ICT development plan in the company in the long term. They are focused on current issues in the struggle to keep the company on the market.

A company information strategy mainly aims to strengthen the link between IS and ICT development in the company and its global strategy in order to subsequently increase its competitiveness and support the development of new forms of business.

When preparing the IS strategy in the longer term, it is necessary to consider four aspects of the IS: achieving the intended objective (effectiveness), efficiency, reliability and continuity of its development.

978-83-60810-53-8/$25.00

III. Methodology, Data, Results and Discussion

Part of the survey focuses on segmenting entities by type of the applications used and by their relationship to the type of ICT service provision. A sub-aim is to familiarize the respondents with a range of options currently offered to purchase, operate and maintain ICT and to determine the level of awareness of companies about the possibility to use cloud computing. Previous results in these areas confirm very similar approaches in the border area concerned.

The survey also includes a questionnaire survey carried out in electronic form on both Czech and Polish sides of the region. The outcome of the project was 160 replies from both Czech (100 responses) and Polish (60 responses) companies. The return rate of the questionnaire on the Czech side was 18.4% (105 responses of 572 questionnaire views). The return rate of the questionnaire on the Polish side was 16.5% (75 responses of 455 questionnaire views). Average time to fill out a questionnaire was about 11 minutes.

The questionnaire is divided into eight interconnected parts: **A, B** (present and future use of ICT tools), **D** (data sources to support decision-making processes), **M** (modules used within IS), **E** (what are your preferences in software procurement), **F** (software functions important to select ICT), **G** (way of ICT maintenance) and **I** (company identification).

The part concerning the applications used is divided into four groups:

A1, B1 – Application for personal IT – office software; antivirus; compression software.

A2, B2 – IT support for key processes in your organization (production, storage, logistics, billing, human resources, inventory, accounting, purchase and sale, etc.).

A3, B3 – Information technology at the tactical management level (MIS – an economic software, using data provided by systems at the operational management level).

A4, B4 - Comprehensive BI solution.

For these four types of applications, intensity of use in the present is examined (variables **A1-A4)**, as well as the assumption of utilizing these applications in the future (variables **B1-B4).**

The first part of the analysis is focused on distributing the relative frequencies of using ICT tools in SMEs in the Czech-Polish border area in the second half of 2012 in %. The results are shown in Table 1.

In the Czech border area, the first level is used the most (L1) – Simple Office ICT tools with 97%, followed by the second level supporting the management of the sub-processes (L2) in 93% of the SMEs. The third level - Comprehensive ICT tools of MIS System (L3) are used only in 62% of companies and BI ICT (L4) are supported in 42% of SMEs. In the Polish border area, ICT tools of L1/L2/L3/L4 levels are used in 97% / 93% / 56% / 28% of SMEs. The first two levels are comparable between regions, but the use of integrated MIS and BI are in the Czech border area significantly higher by about 6% and 14%.

In terms of future development in ICT tools used in both border areas, the use of Simple Office ICT tools assumes the same level, but in the case of ICT tools for the sub-process analysis in SMEs there is stagnation in the Polish region and a slightly increase of 3% is expected in Czech SMEs. On the other hand, it is beneficial planned to increase of ICT using for L3 (a comprehensive MIS) by 10.2% and 5.6% in the Czech and Polish SMEs, respectively. This trend is also detected for of BI tools (by 18.4% in the Czech SMEs and only by 8.3% in the Polish SMEs).

The above results show a considerable interest in and expansion of the use of BI applications.

Another issue of interest includes part E of the questionnaire "What are your preferences in software procurement?". This section offers the respondents four answers, as shown in Figure 2. Out of 96% of valid responses on the Czech side and 56% of valid responses on the Polish side, the evaluation was carried out for the entire region, as well as separately for the Czech and Polish sides.

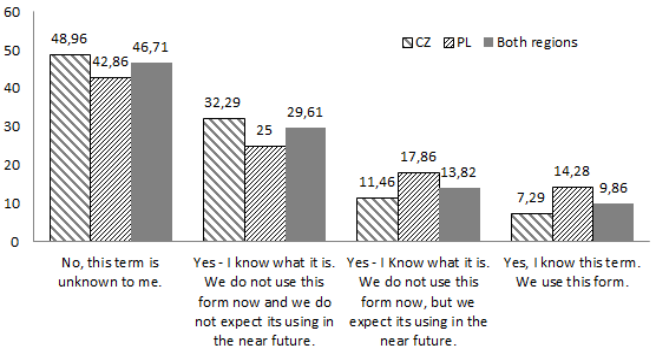


Fig. 2 Empirical distribution of the knowledge of the cloud technologies

The questionnaires about confirmed the expected lack of awareness of managers of what Cloud Computing (CC) is to offer. The correlation between the company size and the

TABLE. I:
EMPIRICAL DISTRIBUTION OF THE USE OF ICT TOOLS IN THE SURVEYED REGIONS

		Valid percentage frequency of using contemporary and future ICT tools							
region	**ICT usage**	**A1**	**B1**	**A2**	**B2**	**A3**	**B3**	**A4**	**B4**
CZ	rarely	3.0	3.1	7.0	4.1	38.0	27.8	57.6	39.2
	often or continuously	97.0	96.9	93.0	95.9	62.0	72.2	42.4	60.8
PL	rarely	3.3	2.3	6.7	6.8	43.9	38.3	71.9	63.6
	often or continuously	96.7	97.7	93.3	93.2	56.1	61.7	28.1	36.4

knowledge of the term CC is a weak, blat statistically significant.

IV. Business Intelligence Tools And Competitive Intelligence, Cloud Computing

BI is system of tools, design solutions and organizational measures to enable the organizational management based on knowledge. These tools have been increasingly applied in companies worldwide. They are specifically aimed at supporting the needs of managers. They form a part of the overall company IS that works with selected or modified data, and that uses these modifications to become the bearer of comprehensive information characterizing the relevant processes in the company. Primarily, they are used to identify and locate specific phenomena in the company, and subsequently to perform their in-depth analysis. An area of BI consists of a number of separate components, with its own architecture and methodology [4].

Many case studies confirm that BI may be utilized in an organization for:

- Increasing the effectiveness of strategic, tactic and operational planning including: modelling different variants in the development of an organization; informing about the realization of enterprise's strategy, mission, goals and tasks; trends, results of introduced changes and realization of plans; identifying problems to be tackled; providing analyses of products, employees, regions; providing analyses of deviations from the realization of plans for particular organizational units or individuals;
- Creating or improving relations with customers, mainly: adequate knowledge about customers for sales representatives so that they could promptly meet their customers' needs and identifying market trends;
- Analyzing and improving business processes and operational efficiency of an organization.
- Providing knowledge and experience which emerged while developing and launching new products onto the market; providing knowledge on particular business processes. [5].

The BI applications use the "OLAP" (On-Line Analytical Processing) technology and Data Mining for advanced analyses. Data Mining allows searching of correlations in large volumes of data, which were not known in advance. The ultimate goal is to provide readable, well-organized, analyzable and readily available information from the maximum number of corporate databases and external sources, which can be utilized in the management [6].

CI represents a set of following activities: definition, collection, analysis and distribution of information and knowledge about clients, competitors and other aspects of the external environment surrounding the organization. These activities are carried out in order to reduce the risk of threats from external business environment, mapping potential opportunities and reaching competitive position.

If we ask ourselves the question what difference there is between CI and BI, the correct answer is that the difference is not very significant. The basis for this assertion is the understanding of the meaning of "intelligence" as the ability to use knowledge assets in action. In our case, these actions relate to the strategic management of an organization.

Thus, the CI only represents another development step, introducing into management sophisticated working methods in connection with the development of advanced technologies for contextual search in external information sources throughout the Internet.

CC is sharing HW and SW means via networks, changing the traditional IT processes and business models. It allows a more efficient use of computing and other data center sources and service providers and brings users to meet their requirements for the speed of the deployment of services, their quality and availability at a transparent price [4]. As already stated above, CC is based on virtualization. Storage, servers, applications and desktops are separated from the actual physical information infrastructure of the business. Virtualization enables greater efficiency and flexibility of IT while reducing IT costs. CC also brings benefits and risks. Benefits of CC:

- The applications or services are provided from centralized data centers through a network, thereby eliminating the software management on each PC.
- Users are not required to know the technology, nor manage its operation on their own. Access to applications and data located on a server is facilitated through a web browser (SaaS - Software as a Service).
- HW can also be provided as a service (IaaS - Infrastructure as a Service). The same applies to the computing platform (PaaS - Platform as a Service).
- Dynamically scalable resources and elasticity.
- Reducing financial costs; the provider supplies the product to multiple users (multitenancy).
- CC changes ICT to a service. Computing power becomes a commodity that we buy and scale as needed.

Limiting factors of CC:

- A possible risk of failing to maintain permanent operation of IT via the Internet – reliability.
- Increased costs for the transfer of large volumes of data.
- Concerns about the security of sensitive data and data in general.
- Lack of control over one's own data, valuable data located off the company.
- Problems with managing permissions and roles with growing company portfolio of CC applications.

A very important contribution of CC for the customer is the transfer of risk and responsibility to the service provider. The supplier is responsible for implementation, audit, security, monitoring, necessary capacity plan, maintenance and support, as well as for availability management.

By using the CC, one can solve the problem of SMEs - lack of availability of many ICT due to their prices and the required infrastructure. In the current economic situation,

CC offers SMEs a viable solution to secure access to the necessary technologies [7].

Currently, CC services offer on-line provision of virtually all software products that can be virtualized into the cloud.

Users no longer have to worry about the management of applications, servers and computer networks and can focus on selecting the range and quality of services purchased from the provider, measuring their use and the prices they pay.

V. Conclusion

A company's global strategy is currently an active strategy. Its quality is dependent on the support of the existing IS. If the IS/IT are well designed and used, they may affect the company's competitiveness. The goal is an interconnection between the development of the information system and the global strategies of a company or an institution [8].

The ICT has become a necessary tool for enterprises, public administration and citizens in most countries. Good strategic management of companies and institutions is conditional upon continuous IS management and maintaining its integrity.

In EU-25, there are approximately 23 million SMEs, representing 99% of all EU companies and providing jobs to approximately 75 million people [9].

The significant role of SMEs is certainly beyond any doubt. It is the ability to respond quickly to changing conditions or absorb free labor that is seen as irreplaceable. SMEs struggle to gain access to capital. This problem persists despite the fact that commercial banks in recent years show a growing interest in these otherwise risky clients that mostly want relatively small loans. The above suggests that capital-intensive investment in ICT is often not a priority for the management of these companies [10].

Changes in company strategy and processes require changes in hardware, software, data storage sites and telecommunication equipment. The quality of business processes is often dependent on the capabilities and features provided by company IS.

Sufficient amount of information nowadays becomes the key factor for success in all fields of human activity. However, obtaining information as such is not enough; they must meet certain parameters. Valuable, meaningful piece of information is one which is provided at the right time and the right place, which it is relevant, correct, complete, and meeting many other requirements.

Projections suggest that in the next decade, we should expect a very rapid development in conventional as well as unconventional approaches to IT and information processing. Possible results are not attractive only in terms of technology, but they can significantly contribute to cost savings and increased productivity in all areas of business and management [11].

Acknowledgment

The results in the paper are based on findings of a research Fund of microprojects Euroregion Silesia, project CZ.3.22/3.3.04/12.02994 (BENEFIT 7) and the European Social Fund within the project CZ.1.07/2.3.00/20.0296.

References

1. J. Voříšek and O. Novotný, "Digital Path to Prosperity" (Synthesis of the collection of studies), VŠE-CSSI, Prague, September 2009.
2. J. Ministr and, M. Števko, “Human Resources Requirements for Professional Management of ITSCM Process”. In IDIMT-2010: Information Technology: Human Values, Innovation and Economy. pp. 57-64, Trauner Verlag, Linz, 2010.
3. Z. Molnár and J. Střelka, „Competitive Intelligence for Small and Middle Enterprises”. In IT for Practice 2011, pp. 89-108, TUO EkF, Ostrava, Czech Republic, 2011.
4. M. Tvrdíková and O. Koubek, „The Use of Cloud Computing (SaaS) for Small and Medium Enterprises to Improve the Quality of their Information Systems“. In IDIMT-2011: Interdisciplinarity in Complex Systems, pp. 389-391, Jindřichův Hradec: Trauner Verlag, Linz, 2011.
5. C. M. Olszak, „The Business intelligence-based Organization- new chances and Possibilities”. In the International Conference on Management, Leadership and Governance. pp. 241-249, Bangkok, Thailand, 2013.
6. R. Němec and F. Zapletal, „The Perception of User Satisfaction in Context of Business Intelligence Systems’ Success Assessment“. In IDIMT-2012: ICT Support for Complex Systems. pp. 203-211, Trauner Verlag, Linz, 2012.
7. P. Rozehnal, „Trends in Management of Companies Caused by the Impact of ICT“. In IDIMT- Interdisciplinary Information Management Talks, pp. 135-142, Linz: Trauner, 2012.
8. J. Hanclova, M.Lukacik and K. Szomolanyi, „A VEC Model of the Czech and Slovak Economies”. In Proceedings of 28th International Conference on Mathematical Methods in Economics. pp. 208-213, České Budějovice: Czech Republic, 2010.
9. (http://ec.europa.eu/enterprise/policies/sme/files/sme_definition/sme_user_guide_cs.pdf)
10. http://www.komora.cz/hk-cr/hlavni-zpravy/art_23005/pocty-malych-a-strednich-firem-rostou-hlavne-v-trznich-sluzbach.aspx)
11. J. Redolia, J., R. Mompo, J. Garcia-Diezc, M. Lopez-Coronadoc, “A model for the assessment and development of Internet-based information and communication services in small and medium enterprises”. Techninovation, vol. 28, pp. 424–435, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 985–989

Implementation of the Big Data concept in organizations – possibilities, impediments and challenges

Janusz Wielki
Opole University of Technology
ul. Waryńskiego 4, 45-047 Opole,
Poland
Email: Janusz@Wielki.pl

***Abstract*—This paper is devoted to the analysis of the Big Data phenomenon. It is composed of seven parts. In the first, the growing role of data and information and their rapid increase in the new socio-economical reality, are discussed. Next, the notion of Big Data is defined and the main sources of growth of data are characterized. In the following part of the paper the most significant possibilities linked with Big Data are presented and discussed. The next part is devoted to the characterization of tools, techniques and the most useful data in the context of Big Data initiatives. In the following part of the paper the success factors of Big Data initiatives are analyzed, followed by an analysis of the most important problems and challenges connected with Big Data. In the final part of the paper, the most significant conclusions and suggestions are offered.**

I. Introduction

INCREASING amounts of data are streaming into contemporary organizations as a result of the rapidly growing quantity of data being generated not only by the organizations themselves but also in the organizations' business environments by both their stakeholders and other entities operating there. Thus, it is in this context that such expressions as "a data-centric world" have become more and more common [1].

The above mentioned processes are significant elements of the socio-economical changes taking place worldwide, where the extremely dynamic development of increasingly powerful and pervasive information technology has an important role to play. Advancements in this field have been a significant catalyst for the transformation of the contemporary economy and the emergence of an "interconnected economy". This new type of economy, in the resource dimension, is a knowledge-based economy, where the most meaningful form of capital is intellectual capital [2]. Under these conditions, from an organization's point of view, the ability to collect the right data and information and to transform it effectively into useful knowledge becomes an increasingly important issue.

Recently, the field of information technology has begun to enter into a new era, as a result of the intensification of the progress being made there. It is an era where processing power and data storage have become virtually free, while networks and cloud-based solutions provide users with global access and pervasive services. As a result of these processes, Big Data sets are being generated which have grown exponentially in size [3]. In 2012, about 2.5 exabytes of data were created every day, with this amount doubling about every forty months [4]. Generally, 90% of the global data in existence has been created over the last two years [5].

As a result, the amount of data and information available for organizations for analysis is exploding [4]. This provides organizations with completely new operating possibilities, while simultaneously generating numerous new challenges. In this context the term "Big Data" has emerged and is being used more and more commonly in the business world.

II. Big Data and the Most Important Sources of the Growth of Data

The term Big Data is not universally understood and applied, leading to various approaches to analyzing it. According to the Leadership Council for Information Advantage, this term is not precise "(…) it's a characterization of the never-ending accumulation of all kinds of data, most of it unstructured. It describes data sets that are growing exponentially and that are too large, too raw or too unstructured for analysis using relational database techniques" [6]. On the other hand, NewVantage Partners describes Big Data as "a term used to describe data sets so large, so complex or that require such rapid processing (sometimes called the Volume/Variety/Velocity problem), that they become difficult or impossible to work with using standard database management or analytical tools" [8]. It is important to underline that Big Data not only relates to the storage and consumption of original content but also to the data connected with this consumption [9].

Generally, there have been some significant trends that have caused a considerable increase in data generation [7]. The first trend, the growth in traditional transactional databases is chiefly connected with the fact that organizations are collecting data with greater granularity and frequency. This is due to various reasons such as the increasing level of competition, increasing turbulence in the business environment and the growing expectations of customers. All of these factors require organizations to react rapidly and with maximum flexibility to the changes taking place and then adjust to them. In order to be able to do this, they are forced to conduct more and more detailed analysis concerning marketplaces, competition and the behavior of consumers [7].

The second trend, the increase of multimedia content, is connected with the rapid increase in the use of multimedia in the various industries of the contemporary economy, such as the health care sector where over 95 % of the clinical data

978-83-60810-53-8/$25.00

generated is in video format. Generally, multimedia data already accounts for over half of Internet backbone traffic and it is predicted that this share will grow to 70% by the end of 2013 [7].

The next trend which has caused a growth in the amount of data being generated is the development of the phenomenon called "The Internet of Things", where the number of physical objects or devices that communicate with each other without any human interference is increasing at a fast pace. They link with each other in a wired or wireless manner, often using IP protocols. As they are equipped with various sensors or actuators they collect and send huge amounts of data [10]. By 2015 the amount of data generated from the 'Internet of Things' will grow exponentially as the number of connected nodes deployed in the world is expected to grow at a rate of over 30% per year [7].

Social media is the next extremely significant source of the increase of data. Facebook users alone generate huge amounts of data. In 2011 the 600 million active users of this social platform spent over 9.3 billion hours a month on the site, with the average Facebook user generating 90 pieces of content (photos, notes, blog posts, links, or news stories) [7]. A year later the number of Facebook users reached 1 billion. Research conducted at the beginning of 2012 showed that if only messages are considered, users receive an average of nearly twelve messages a month, and send nine [11]. In the case of YouTube, every minute 24 hours of video is uploaded, while over the same timeframe Twitter users send 98000 tweets [7], [12]. In addition, smart phones are playing an increasingly important role in social networks. Although the penetration of social networks is increasing for both PCs and smartphones, it is significantly higher for smartphones. If frequent users are considered in the case of PC's it is 11% p.a. while in the case of smartphones it is 28% p.a [7]. This has caused a rapid increase in mobile data traffic which doubled between the third quarter of 2011 and the third quarter of 2012. It is predicted that mobile data traffic will grow twelve fold by 2018 [13].

III. Opportunities and Benefits Connected With Big Data Utilization

The development of the Big Data phenomenon and its associated tools and techniques, is not something which has been separated from the wider processes which have been taking place in organizations over recent years. In fact, it is becoming more and more common in organizations concerned with the field of analytics and has significantly expanded the possibilities available within the scope of business intelligence (BI) tools.

Given their role in providing organizations with numerous possibilities and opportunities in the sphere of analytics, business intelligence systems are well suited for aggregating and analyzing structured data [14]. But there are, however, some types of analyses that BI can not handle. These mainly relate to situations where data sets become increasingly diverse, more granular, real-time and iterative.

Such types of unstructured, high volume, fast-changing data, pose problems when trying to apply traditional approaches based on relational database models. As a result, it has become apparent that there is growing demand for a new class of technologies and analytical methods [6].

There are many diverse benefits arising from the utilization of Big Data, depending on the sector of the economy, as has been confirmed by the results of research conducted by the McKinsey Global Institute. These results show the transformational potential of Big Data in such diverse domains as health care, public sector administration, retail, manufacturing and personal location data [7]. According to the results of a survey conducted in summer 2012 by NewVantage Partners, among C-level executives and function heads from many of America's leading companies, there are seven basic groups of benefits connected with Big Data initiatives. Better, fact-based decision making (22%) and an improved customer experience (22%) are the most important of these benefits, coupled with the overall message that the expectation is to make better decisions faster. The other groups of benefits include: increased sales (15%), new product innovations (11%), reduced risk (11%), more efficient operations (10%) and higher quality products and services (10%) [15].

Organizations use Big Data platforms to give them answers to important questions in seconds rather than months. Thus, the key value of Big Data is to accelerate the time-to-answer period, allowing an increase in the pace of decision-making at both the operational and tactical levels [14], [15]. An extremely important new element, in the context of decision-making, connected with the Big Data phenomenon, is the possibility for constant business experimentation to guide decisions and test new products, business models, and customer-oriented innovations. Such an approach even allows, in some cases, for decision making in real-time. There are many examples of companies using this in practice. For example multifunctional teams in Capital One perform over 65,000 tests each year. They experiment with combinations of market segments and new products. The online grocer FreshDirect adjusts, on a daily basis or even more frequently, prices and promotions based on online data feeds. Tesco is another example. This company gathers transaction data on its millions of customers through a loyalty card program and uses it to analyze new business opportunities. For example, it looks at how to create the most effective promotions for specific customer segments and how to inform them about decisions concerning pricing, promotions, and shelf allocation [17]. Walmart is another example. This company created the Big Data platform (The Online Marketing Platform) which is used, among other things, to run many parallel experiments to test new data models [16]. Also such dot-com giants as Amazon, eBay and Google have been using testing in order to drive their performance [17].

According to the McKinsey Global Institute, five key ways in which Big Data creates value for organizations can be distinguished [7]:

- creating transparency by integrating data and making it more easily accessible to all relevant stakeholders,
- enabling experimentation to discover needs, expose variability, and improve performance,

- segmenting populations in order to customize actions,
- replacing or supporting human decision making with automated algorithms,
- innovating new business models, products and services.

Generally, the results of the research conducted in February 2012 among 607 executives from around the world by the Economist Intelligence Unit confirm the value of Big Data utilization by companies. The surveyed executives claim that Big Data initiatives have improved the performance of their organizations over the past three years by around 26%. Simultaneously they expect that such initiatives will improve performance by an average of 41% over the next three years [18]. In addition, it is worth noticing that according to the results of the research of Brynjolfsson et al., firms where decision making is based on data and business analytics have 5-6% higher output and productivity. Decision making based on data and business analytics also impacts on other performance measures such as asset utilization, equity return and market value [18].

As in the case of BI initiatives Big Data systems have been used for two purposes - human decision support and decision automation. According to the results of the above mentioned research conducted by the Economist Intelligence Unit, Big Data is used, on average, for decision support 58% of the time and for decision automation around 29% of the time, based on the level of risk connected with the decision [14].

IV. Tools, Techniques and the Most Useful Data in the Context of Big Data Initiatives

The effective implementation of Big Data initiatives requires an undertaking of appropriate organizational actions, including ensuring organizations are provided with all the necessary resources to enable analysis of the ever-growing data sets to which they have access. In this context, the application of proper techniques and technologies is one of the key issues. In practice, organizations use many various techniques and technologies to aggregate, manipulate, analyze, and visualize Big Data. They come from various fields such as statistics, computer science, applied mathematics, and economics. Some of them have been developed intentionally and some of them have been adapted for this purpose. Examples of techniques utilized for the analysis of Big Data are: A/B testing, data fusion and data integration, data mining, machine learning, predictive modeling, sentiment analysis, spatial analysis, simulation or time series analysis. Examples of technologies used to aggregate, manipulate, manage, and analyze of Big Data are: Big Table, Cassandra, Google File System, Hadoop, Hbase, MapReduce, stream processing, visualization (tag cloud, clustergram, history flow, spatial information flow) [7].

Increasingly, there are a number of new analytical toolkits for the analysis of Big Data. Examples of such solutions are [19]:

- Alterian, TweetReach (network intelligence tools for real-time analysis of the reactions and responses to changes of industry players),
- NM Incite, Social Mention, SocMetrics, Traackr, Tweepi (sentiment analysis tools for estimating the buzz around a product or service, influencer intelligence tools for identifying key influencers and targeting for marketing or insights),
- Attensity, Autonomy (live testing tools for getting direct feedback from users on new products or ideas, data mining tools for text-analytics to estimate market size).

In addition, a very important element of Big Data initiatives is properly trained people. In this context, a specific type of worker is indicated, known as data scientists, who are properly trained to work with Big Data. In practice, it means that they should be people who know how to discover the answers to an organization's key questions from huge collections of unstructured data. These people should be a hybrid of analyst, data hacker, communicator and trusted advisor [20]. In addition to analytical abilities and substantial and creative IT skills, they should be close to the products and processes inside the organization [21]. As the acquisition of in-depth domain knowledge from data scientists typically takes years [14], most organizations build platforms to close the gap between the people who make decisions and data scientists, such as that created by Walmart - the Social Genome Platform. It facilitates cooperation among buyers, merchandisers, product managers and other people who have worked in retail for years and data scientists [22].

In addition to proper techniques, tools and people, the basic resource required for Big Data initiatives is appropriate data. As was mentioned earlier, a lot of data from various sources is currently flowing into contemporary organizations but not all Big Data sets are equally valuable,. Business activity data such as sales, purchases, costs etc. is definitely the most important source of data. Office documentation is the second key source of data, closely followed by social media. In certain sectors such as healthcare, pharmaceutical, and biotechnology, data sets from social media are more important that those from office documentation. The other important types of data sets include: point-of-sale data, Website clickstream data, RFID/logistics data, geospatial data, telecommunications data, telemetry data [18].

V. Success Factors of Big Data Initiatives

Through an analysis of implemented Big Data initiatives, various success factors can be determined, each with their own set of recommendations.Marchand and Peppard have identified five important guidelines for the success of a Big Data project. They include [24]:

1. Placing people at the heart of the Big Data initiative.
2. Emphasizing information utilization as the way to unlock value from information technology.

3. Equipping IT project teams with cognitive and behavioral scientists.
4. Focusing on learning.
5. Worrying more about solving business problems than about deploying technology.

Based on their experiences gained from cooperation with companies from data rich industries, Barton and Court, on the other hand, came to the conclusion that full exploitation of data and analytics requires three capabilities [25]:

1. Choosing the right data. In this context two aspects are important: creative sourcing of internal and external data and upgrading IT architecture and infrastructure for easy data merging.
2. Emphasizing information utilization as the way to unlock value from information technology. In this context two aspects are important: focusing on the biggest drivers of performance and building models that balance complexity with ease of use.
3. Equipping IT project teams with cognitive and behavioral scientists. In this context two aspects are important: creating simple, understandable tools for people on the front lines and updating business processes and developing capabilities to enable tools utilization.

According to Barth et al., organizations that benefit from Big Data base their activities on three fundamental issues [21]:

1. Paying attention to data flow as opposed to stocks.
2. Relying on data scientists and product and process developers rather than data analysts.
3. Moving analytics away from the IT function, into core business, with operational and production functions.

VI. The most significant obstacles and challenges connected with Big Data eferences

As with other IT-related initiatives, Big Data also has its own set of problems and challenges. The Economic Intelligence Unit research mentioned earlier indicates some of the impediments to the effective utilization of Big Data for decision-making [14]. "Organizational silos" were the most significant barrier (55,7%), which result from the fact that data connected with particular organizational functions (i.e. sales, distribution etc.) are collected in "function silos" rather than pooled for the benefit of the entire company. The second (50,6%), although no less important, issue is the lack of appropriately skilled people (data scientists) prepared to analyze data. The third aspect (43,7%) is the excessively long time it takes organizations to analyze huge data sets. As was mentioned earlier, organizations expect to be able to analyze and act on data in real time. The fourth barrier (41,7%) is the difficulties concerned with the analysis of ever increasing amounts of unstructured data. Finally, the inability of senior management to view Big Data in a sufficiently strategic way (34,9%) is the fifth key impediment [14].

McAfee and Brynjolfsson indicate five management challenges which prevent organizations from reaping the full benefits of Big Data utilization. They are: leadership, talent management, technology, decision making, company culture.

When considering leadership, having more or better data does not guarantee success. The leaders still have to have a vision of the organization's development, set clear goals, understand the market, etc. Big Data changes the way organizations make many of their decisions. Talent management is connected with the necessity of providing the organization with the right people (such as data scientists) who are prepared to work with huge sets of data. The next challenge relates to the problem of assuring the data scientists have the proper tools to handle the Big Data. Although the technology alone is not enough to succeed in Big Data initiatives, it is a necessary part of it. The next challenge is connected with the problem of ensuring there are mechanisms in place to guarantee that the information and the relevant decision-makers are in the same location. It is important to make sure that the people who understand the problems are able to use the right data and to work with people who have the necessary problem solving skills.

The final challenge is connected with changes related to organizational culture. The key issue in this context is to make decisions as data-driven as possible, instead of basing them on hunches and instinct [4]. It is worth mentioning that the significance of such cultural transformation is also mentioned in other research e.g. that concerning sectoral Big Data projects [23].

In addition, the numerous challenges connected with data and legal rights should be noted. They relate to such issues as copyright, database rights, confidentiality, trade marks, contract law, competition law [1]. There is another important challenge also connected with legal aspects. It relates to the transparency in data collection practices. A further important risk is around the utilization of Big Data to increase the automation of decision-making. There is one more important danger which is underlined in the context of Big Data. It is connected with the fact that Big Data might not be providing the whole picture for a particular situation. There are several reasons for this i.e. biases in data collection, exclusions or gaps in data signals or the constant need for context in conclusions [26].

At the same time, existing pre-Big Data challenges and threats are still developing, such as the problem of securing collected data and information. These issues chiefly relate to the problem of how to protect competitively sensitive data and data that should be kept private by organizations (e.g. various types of consumer data) [7]. As a result, the problems connected with the broadly defined security of the IT infrastructure of organizations and protection against various attacks becomes an even more important issue than previously [27]. The increasing dependence, as a result of the Big Data phenomenon, of organizations on the efficient and reliable functioning of their IT infrastructure, means that securing it has become even more important.

VII. Conclusions

The rapidly growing amount of data which organizations have at their disposal and the opportunities connected with its practical utilization are increasingly changing the processes relating to making decisions at various organizational

levels. Thus, Big Data offers huge potential to positively impact on the functioning of organizations generally and gives them a competitive advantage. Companies are now trying to utilize to an even greater degree the opportunities and chances that are emerging.

But if initiatives aimed at the practical usage of Big Data sets are to be successful at giving an organization a competitive advantage and be of value, it is not enough to just collect and own the appropriate data sets. In fact, this is only the starting point of every Big Data initiative. Further essential elements are suitable analytical models, tools, skilled people, and organizational capabilities. Lack of all of these necessary components can lead to a situation whereby instead of expected benefits there is only disappointment and a belief that Big Data initiatives are only the next wave in a long line of management fads.

Generally, although Big Data solutions have a huge potential for both commercial organizations and governments, there is uncertainty concerning the speed with which they can be utilized in a secure and useful way [3].

REFERENCES

[1] Kemp Little LLP, "Big Data – Legal Rights and Obligations", http://www.kemplittle.com/Publications/WhitePapers/Big%20Data%20-%20Legal%20Rights%20and%20Obligations%202013.pdf, January 2013.

[2] D. Tapscot, A. Williams, *Radical Openness*. New York: TED Books (Kindle Edition), 2013.

[3] National Intelligence Council, "Global Trends 2030", http://globaltrends2030.files.wordpress.com/2012/11/global-trends-2030-november2012.pdf, December 2012.

[4] E. Brynjolfsson, A. McAfee, (2012), "Big data: The management revolution", *Harvard Business Review*, pp. 60-68, October 2012.

[5] A. Lampitt, "Hadoop: Analysis at massive scale", in *InfoWorld*, http://resources.idgenterprise.com/original/AST-0084522_IW_Big_Data_ rerun_1_all_sm.pdf, pp. 8-12, Winter 2013.

[6] LCIA, "Big Data: Big Opportunities to Create Business Value", http://poland.emc.com/microsites/cio/articles/big-data-big-opportunities/ LCIA-BigData-Opportunities-Value.pdf, 2011.

[7] McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity", http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf, May 2011.

[8] NewVantage Partners, "Big Data Executive Survey: Themes & Trends", http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Themes-Trends.pdf, 2012.

[9] J. Gantz, D. Reinsel, "Extracting Value from Chaos", http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf, June 2011.

[10] M. Chui, M. Löffler, R. Roberts, "The Internet of Things", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=4&L3= 116&ar=2538, March 2010.

[11] K. Hampton, L. Goulet, C. Marlow, L. Rainie, "Why most Facebook users get more than they give", *Pew Internet & American Life Project*, http://pewinternet.org/~/media//Files/Reports/2012/PIP_ Facebook%20users_2.3.12.pdf, February 3, 2012.

[12] HP, "Big security for big data", http://www.hpenterprisesecurity.com/collateral/whitepaper/BigSecurityforBigData0213.pdf, December 2012.

[13] Ericsson, "Ericsson Mobility Report", http://hugin.info/1061/R/1659597/537300.pdf, November 2012.

[14] Capgemini, "The Deciding Factor: Big Data & Decision Making", http://www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/?d=6C800B16-E3AB-BC55-00F4-5411F5DC6A8C, February 2012.

[15] NewVantage Partners, "Big Data Executive Survey: Creating a Big Data Environment to Accelerate Business Value", http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Accelerate-Business-Value.pdf, 2012.

[16] Walmartlabs, "Big Data Platform and Demand Generation", http://www.walmartlabs.com/platform/, 2013.

[17] J. Bughin, M. Chui, J. Manyika, "Clouds, big data, and smart assets", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=20&L3= 75&ar=2647, August 2010.

[18] E. Brynjolfsson L. Hitt, H. Kim, "Strength in Numbers", http://ssrn.com/abstract=1819486, December 12, 2011.

[19] M. Harrysson, E. Metayer, H. Sarrazin, "How 'social intelligence' can guide decisions", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=21&L3=37&ar=3031, November 2012.

[20] T. Davenport T., D. Patil, "Data scientist", *Harvard Business Review*, pp. 70-76, October 2012.

[21] P. Barth, R. Bean, T. Davenport, "How Big Data is Different", *Sloan Management Review*, Fall 2012.

[22] R. Ferguson, "It's All About the Platform: What Walmart and Google Have in Common", *Sloan Management Review*, http://sloanreview.mit.edu/article/its-all-about-the-platform-what-walmart-and-google-have-in-common/, December 5,.2012.

[23] P. Groves, B. Kayyali, D. Knott, S. van Kuiken, "The_big_data_revolution_in_healthcare", *McKinsey Quarterly*, http://www.mckinsey.com/ insights/health_systems/~/media/7764A72F70184C8EA88D805092D72D58.ashx, January 2013.

[24] D. Marchand, J. Peppard, "Why IT fumbles analytics", *Harvard Business Review*, pp. 104-113, January-February 2013.

[25] D. Barton, D. Court, "Making advanced analytics work for you", *Harvard Business Review*, pp. 78-83, October 2012.

[26] R. Ferguson, "Competitive Advantage with Data? Maybe … Maybe Not", *Sloan Management Review*, http://sloanreview.mit.edu/article/competitive-advantage-with-data-maybe-maybe-not/?utm_source=facebook&utm_medium=social&utm_campaig, March 26, 2013.

[27] J. Wielki, *Modele wpływu przestrzeni elektronicznej na organizacje gospodarcze*, Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego, 2012.

Agent Day 2013

DURING the last decade agent systems have been constantly developed and many approaches were successful in such areas as transport, decision support, distributed monitoring and control systems, computation. Such immanent features of the agents like autonomy or task-orientation assured the result. These (and others) features of the agent systems may be further extended to produce new paradigms for constructing scalable computing systems with focus on grid computing, cloud computing, resilient control and monitoring systems, SCADA and many more.

It is also worth noticing, that hardware available several years ago, such as computing clusters, are nowadays outperformed by simple desktop sets or even portable computers. Even environment appliances such as gaming consoles or GPUs may be used to perform complex computations in parallel, much faster than even multi-core CPUs. Such a plethora of possibilities calls for promotion and dissemination of results and ideas connected or inspired by agent-based systems or their features.

Topics

For the "Agent Day" Session we would like to give guidance in the subject, as the following topics, that do not exhaust all the possibilities:

- multi-agent management, scheduling, load-balancing
- multi-agent computation and simulation
- stochastic and structural modeling of complex systems
- distributed and grid computing
- software reuse in complex systems
- GPU and gaming consoles for computing
- nature-inspired, evolutionary and memetic computing
- scalability, extendability, resilience in complex systems
- mobile robotics
- data-intensive computing
- various application of multi-agent systems

Steering Commitee

Cetnarowicz, Krzysztof, AGH University of Science and Technology, Poland

Dobrowolski, Grzegorz, AGH University of Science and Technology, Poland

Nawarecki, Edward, AGH University of Science and Technology, Poland

Schaefer, Robert, AGH University of Science and Technology, Poland

Organizing Committee

Byrski, Aleksander, AGH University of Science and Technology, Poland

Dobrowolski, Grzegorz, AGH University of Science and Technology, Poland

Kolodziej, Joanna, Cracow University of Technology, Poland

Nalepa, Grzegorz J., AGH University of Science and Technology, Poland

Program Committee

Ambroszkiewicz, Stanislaw, Institute of Computer Science, Polish Academy of Sciences, Poland

Atzmueller, Martin, Kassel University, Germany

Burczynski, Tadeusz, Cracow University of Technology, Poland

Burguillo Rial, Juan Carlos, University of Vigo, Spain

Byrski, Aleksander, AGH University of Science and Technology, Poland

Canadas, Joaquin, University of Almeria, Spain

Carvalho, Marco, Florida Institute of Technology, United States

Cervenka, Radovan, Whitestein Technologies AG

Cetnarowicz, Krzysztof, AGH University of Science and Technology, Poland

Cotta, Carlos, University of Malaga, Spain

Dajda, Jacek, AGH University of Science and Technology, Poland

Danoy, Gregoire, University of Luxembourg, Luxembourg

Dobre, Ciprian, Politehnica of Bucharest, Romania

Dobrowolski, Grzegorz, AGH University of Science and Technology, Poland

Eleftherakis, George, The University of Sheffield International Faculty, CITY College, Greece

Florea, Adina Magda, University Politehnica of Bucharest, Romania

Fong, Simon , University of Macau, Macao S.A.R., China

Grzech, Adam, Wroclaw University of Technology, Poland

Kaleta, Mariusz, Warsaw University of Technology, Poland

Kisiel-Dorohinicki, Marek, AGH University of Science and Technology, Poland

Kliazovich, Dzmitry, University of Luxembourg, Luxembourg

Kolodziej, Joanna, Cracow University of Technology, Poland

Koukam, Abder, IRTES-SeT Université de Technologie de Belfort Montbéliard, France

Letia, Ioan Alfred, Technical University of Cluj-Napoca, Romania

Ligęza, Antoni, AGH University of Science and Technology, Poland

Michalewicz, Zbigniew, University of Adelaide, Australia

Nalepa, Grzegorz J., AGH University of Science and Technology, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 993–998

Learning sensors usage patterns in mobile context-aware systems

Szymon Bobek*, Krzysztof Porzycki†, Grzegorz J. Nalepa‡
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059, Krakow, Poland
*szymon.bobek@agh.edu.pl, †kporzyck@student.agh.edu.pl, ‡gjn@agh.edu.pl

***Abstract*—Context-aware mobile systems have gained a remarkable popularity in recent years. Mobile devices are equipped with a variety of sensors and become computationally powerful, which allows for real-time fusion and processing of data gathered by them. However, most of existing frameworks for context-aware systems, are usually dedicated to static, centralized architectures, and those that were designed for mobile devices, focus mainly on limited resources in terms of CPU and memory, which in nowadays world is no longer a big issue. Mobile platforms require from the context modelling language and inference engine to be simple and lightweight, but on the other hand – to be powerful enough to allow not only solving simple context identification tasks but also more complex reasoning. These, with combination of a large number of sensors and CPU power available on mobile devices result in high energy consumption of a system. The original contribution of this paper is a proposal of an intelligent middleware for mobile context-aware frameworks, that is able to learn sensor usage habits, and minimize energy consumption of the system.**

I. Introduction

RESEARCH in the area of pervasive computing and ambient intelligence aims to make use of context information to allow devices or applications behave in a context-aware, thus "intelligent" way. Dey [1] defines context as *"any information that can be used to characterize the situation of an entity*. The *information* in Dey's definition may be: (1) location of the user (spatial context), (2) presence or absence of other devices and users nearby, or collaboration with other users (social context), (3) time (temporal context), (4) user *behavior* or *activity*, and possibly (5) any other environmental data gathered by microphones, light sensors, etc.

The variety of sensors available on nowadays mobile devices allow for complex contex-based reasoning, but at the same time requires a lot of resources and energy.

Although there are many frameworks and middlewares developed for context-aware systems [2], [3], [4], they do not provide full support for all of the challenges that we believe are crucial for mobile computing (e.g. smartphones or tablets), with respect to the context modelling and context-based reasoning. Those are:

Energy efficiency – most of the sensors, when turned on all the time, decrease the mobile device battery level very fast. This reflects on usability of the system and ecological aspects regarding energy saving.

Data privacy – most of the users do not want to send information about their location, activities, and other private data to external servers. Hence, the context reasoning should be performed by the mobile device itself.

Resource limitations – although mobile phones and tablets are becoming computationally powerful, the context aware system has to consume as low CPU and memory resources as possible in order to be transparent to the user and other applications.

System responsiveness – in mobile environment context changes very fast. Hence, no delays are admissible in processing contextual data.

Context data distribution – in mobile pervasive environments many devices produces huge amount of contextual information, hence the quality measures should be developed and distribution methods designed to fit characteristics of such unstable and dynamic network [5].

All of these require from the modelling language and inference engine to be simple and lightweight. On the other hand, the model should be powerful enough to allow not only solving simple context identification tasks but also more advanced context processing and reasoning.

This gives motivation for developing a solution that will allow for using advanced reasoning and modelling techniques, with as low energy cost as possible. The original contribution of the paper is a proposal of an intelligent middleware for mobile context aware frameworks, that is able to learn sensor usage habits, and minimize energy consumption of the system.

The rest of the article is organized as follows: In Section II an existing context aware systems and frameworks are presented, and the motivation of the paper is given. The architecture that can be used in combination with our approach is presented in Section III. The Section IV discusses the learning algorithm used for intelligent middleware and Section V presents an evaluation of the algorithm. Finally, summary and directions for future work are given in Section VI.

II. State of the art and motivation

In recent years, a lot of development was devoted to build applications that use mobile devices to monitor and analyse various user contexts. The availability of application distribution platforms for common mobile operating systems, e.g. Google Play for Android stimulated the popularity and adoption of such solutions. However, most of them focus only on a very narrow application area of context awareness. Most of them are end user applications, and not more generic

978-83-60810-53-8/$25.00

frameworks. Some selected representative cases are briefly described below.

A. Context aware systems

The SocialCircuits platform [6] uses mobile phones to measure social ties between individuals, and uses long- and short-term surveys to measure the shifts in individual habits, opinions, health, and friendships influenced by these ties.

Jung [7] focused on discovering social relationships (e.g., family, friends, colleagues and so on) between people. He proposed an interactive approach to build meaningful social networks by interacting with human experts, and applied the proposed system to discover the social networks between mobile users by collecting a dataset from about two millions of users. Given a certain social relation (e.g., isFatherOf), the system can evaluate a set of conditions (which are represented as propositional axioms) asserted from the human experts, and show them a social network resulted from data mining tools.

Sociometric badge [8] has been designed to identify human activity patterns, analyse conversational prosody features and wirelessly communicate with radio base-stations and mobile phones. Sensor data from the badges has been used in various organizational contexts to automatically predict employee's self-assessment of job satisfaction and quality of interactions.

Eagle and Pentland [9] used mobile phone Bluetooth transceivers, phone communication logs and cellular tower identifiers to identify the social network structure, recognize social patterns in daily user activity, infer relationships, identify socially significant locations and model organizational rhythms.

Beside research projects, there exist also a variety of application that are used for gathering information about context from mobile devices, like SDCF [10], AWARE [1], JCAF [11], SCOUT [12], ContextDriod [13], Gimbal [2]. These are mostly concerned with low-level context data acquisition from sensors, suitable for further context identification. On the other hand, they do not provide support nor methodology for creating complex and fully customizable context-aware systems and do not provide any mechanisms for limiting energy consumption of the system.

What is more, all of the approaches described above use their own dedicated methods for gathering and maintaining context. These methods are mostly not applicable for reuse, or their functionality is limited to simple context matching. Some of the systems do not provide any support for context modelling nor context reasoning, limiting their functionality only to identifying and collecting contextual information.

B. Context aware frameworks

To solve the issue of reusability of the system, a lot of frameworks were designed. These frameworks are based on many different architecture paradigms which pros and cons in terms of energy efficiency, responsiveness, and privacy were presented in this Section.

[1]http://www.awareframework.com
[2]https://www.gimbal.com/

The system described in [14] uses *direct sensor access* architecture which is usually not very energy efficient, however it preserver privacy issues, since no communication with external servers is usually needed, and the interpretation of the sensor data as well as reasoning is performed directly on the host device

The CoBrA system [15] was build on *centralized context server* architecture. This approach is especially useful when a context-aware system is composed of many mobile devices with limited resources. The server relieves mobile agents from performing reasoning tasks. On the other hand, one has to consider privacy issues connected with sending private contextual data to remote server, quality of service issues, etc. This approach is also characterized with rather low responsiveness that stems from a possible lack of network connection or communication delays.

Service oriented architecture with combination of *distributed architecture* was used in SOCAM [16] system. In context-aware applications this architecture is used mainly in pervasive environment, where variety of context information from many different sources has to be processed. This architecture usually does not preserves privacy nor energy efficiency issues since usually it assumes communication over the web between each of its elements. Although SOCAM provides architecture for distributed mobile systems, it mostly solves problems of a low memory and CPU power of mobile agents, which nowadays is no longer a big issue for most of the mobile devices like smart-phones or tablets. On the other hand, energy efficiency issue is still a big problem, which was not addressed by none of the solutions described in this Section.

This gives motivation for developing an architecture that will allow for advanced context-based reasoning and modelling, but at the same time allow for minimizing energy usage costs of sensors that are needed in such reasoning. An overview of the proposed system is presented in following Section.

III. Intelligent middleware approach

The proposed solution incorporates an idea of a mobile device as an autonomous context-aware entity, equipped with intelligent middleware layer and context-based inference layer. The intelligent middleware act as a proxy between context sources and inference layer. It is able to learn sensor usage patterns and thus adjusting sampling rates to significantly improve energy consumption of the system (See Section IV).

The architecture of a system that may use intelligent middleware approach should consist of three main elements:

1) sensors layer – responsible for gathering data from sensors and performing initial preprocessing of them,
2) inference layer – responsible for context based reasoning and knowledge management, and
3) intelligent middleware layer – acting as an intelligent proxy between sensors layer and the inference layer.

The *Sensor Layer* gathers data directly from mobile device sensors. Due to the different possible sensor types (GPS, Accelerometer, Bluetooth), different methods for interpreting

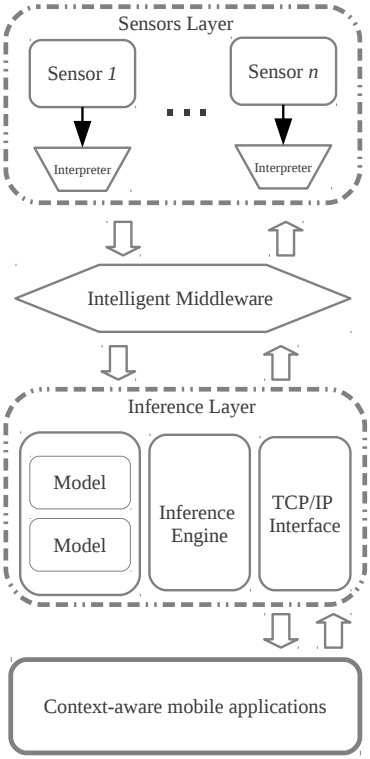


Figure 1. Architecture of the mobile context aware framework

these data are required. Hence, each sensor has its own interpreter module that is responsible for initial preprocessing of the raw data. Data preprocessing is triggered by the intelligent middleware.

The *Inference Layer* is responsible for performing reasoning, based on the model (knowledge base) and the working memory elements (facts). The inference engine may be a rule engine, first-order logic reasoner, probabilistic inference module, or any other custom approach. However, we argue, that to allow more complex reasoning tasks than just simple context classification, the best choice is lightweight rule engine [17].

The *Intelligent Middleware* is responsible for exchanging information between sensors layer and inference layer. The working memory is shared between all models stored within the inference layer, acting as a *knowledge cache*. Therefore, it minimizes the number of required requests to the sensors layer, improving power efficiency of the entire system.

The idea of separating Intelligent Middleware from inference layer is that it is able to learn sensors usage habits, and in consequence adapt itself to the individual device characteristics. It automatically generates a model of usage habits from historical data and based on that model data it adjusts the sampling rates for the sensors appropriately. It improves power efficiency of the system, since sampling rates are not fixed, but leaned from the usage patterns. On the other hand it may help in increasing responsiveness of the system, since the learned model allows predicting not only future sensor activity, but also context-aware application needs. Hence, it is possible to get the desired context in advance, before the application actually requests it. It can be especially useful in cases when context cannot be obtained by the middleware directly from the sensor layer, but has to be for example downloaded over the internet. However in this paper we focus only on the power efficiency advantage of the usage of the intelligent middleware approach.

The following sections describes in details the learning algorithm used, and provide an evaluation on a simple use case scenario.

IV. Learning algorithm

Input data. The algorithm takes as an input a vector of m percepts. Each percept is described by a pair (X_i, Y_i) interpreted respectively as time of percept and sensor activity state. Such a notation results in two vectors X, Y of size m such that $\forall_{i<m} 0 \leq X_i \leq 24 \wedge (Y_i = -1 \vee Y_i = 1)$. Time equals numbers of hours passed since last midnight and percept is:

$$Y_i = \{^{-1\ for\ inactive\ state}_{1\ for\ active\ state}$$

Learning objective. Sensor activity depends largely on its stochastic and inaccessible environment. Being so, it is impossible to predict it with absolute certainty, however, often some part of its variance can be explained by time of a day. The algorithm proposed aims to exploit that possibility by finding a function determining probability of sensor usage given time of a day $F(t) = P(X = 1|t)$. Problems of learning conditional probability are often addressed in Machine Learning by using logistic regression. The following paragraphs define necessary concepts and present the problem in terms of logistic regression with accordingly chosen parameters.

Hypotheses set. Finding objective function $F(t)$ is achieved by searching a hypotheses set H. Each function h in hypotheses set has to have the following properties:

1) be continuous,
2) be defined in range $< 0; 24 >$,
3) $h(0) = h(24)$,
4) return values in range $< 0; 1 >$ (probability).

To perform search it is necessary to represent each function in H in a general form parametrized by some vector w of length $2n+1$, such that every combination of parameters in w will yield in a proper hypothesis $h = H_w$. Such representation allows to transition from searching a set of functions to searching a linear space $\mathbb{R}^{2n+1}$.

A representation that has the first three of required properties is given below:

$$S(\omega, t) = \omega_0 + \sum_{i=0}^{n-1} \left(\omega_{2i+1} \cos\left(\frac{i * t * 12}{\Pi} \right) \right) + \sum_{i=0}^{n-1} \left(\omega_{2i+2} \sin\left(\frac{i * t * 12}{\Pi} \right) \right)$$

It may be understood as a sum of some first terms of Trigonometric Fourier Series parametrized by vector ω. Using only low frequency components is desirable because they are

most likely to describe habits of usage that usually occurs in long sequences of same actions. The only requirement left, that is – unbound return values, can be addressed by composing function $S(\omega, t)$ with sigmoid function:

$$\theta(x) = \frac{1}{1 + e^{-x}}$$

The resulting and correct hypotheses set parametrized by vector w is given by formula:

$$H_\omega(t) = \theta(S(\omega, t))$$

Interpreting hypothesis as probability

Assumed interpretation that $P(y = 1|x) = h(x)$ implies that $P(y = -1|x) = 1 - h(x)$. Because $h(x) = \theta(s(\omega, t))$, and the properties of θ: $\theta(-s) = 1 - \theta(s)$, the resulting probability formula is drawn:

$$P(y|x) = \theta(y * S(\omega, t))$$

This formula is based on the assumption made earlier, that y takes 1 for active and -1 for inactive state.

Learning input data Out of all possible functions in the hypotheses set one has to be chosen in terms of its lowest cost. In order to perform such a selection a cost measure has to be defined. The suggested measure is a combined probability of all the observations in a learning set. The higher the combined probability the better a hypothesis describes the user habit that gave rise to such sensor readings. Derivation of final formula to be optimized: $\max_\omega \prod_{i=0}^{m-1} P(Y_i|X_i)$ Maximizing an expression is equivalent to maximizing its logarithm:

$$\max_\omega \ln \prod_{i=0}^{m-1} \theta(Y_i * S(\omega, X_i))$$

$$\max_\omega \frac{1}{m} \ln \prod_{i=0}^{m-1} \theta(Y_i * S(\omega, X_i))$$

$$\min_\omega -\frac{1}{m} \ln \prod_{i=0}^{m-1} \theta(Y_i * S(\omega, X_i))$$

$$\min_\omega -\frac{1}{m} \sum_{i=0}^{m-1} ln(\frac{1}{\theta(Y_i * S(\omega, X_i))})$$

Formula in such a form is then subject to optimization. The optimization algorithm used in this case was gradient descend with initial ω coefficients set all to 0. Fast convergence to unique value is always achieved thanks to minimized formula being always convex.

V. Evaluation

We implemented a prototype of an intelligent middleware, that learns user habits based on the usage of device sensors (in this case a GPS sensor). We assumed that the GPS sensor is active if the speed of the device exceeds some fixed threshold, otherwise the sensor was assumed to be inactive. This reflects to the cases where someone was moving or not.

The learning and evaluation process is presented in Figure 2. We first performed an acquisition of the GPS sensor data and save it to SQLite database. We collected samples from 5 consecutive days, which later were preprocessed offline to be ready for the learning algorithm. The main aim of the preprocessing phase was to decide weather the GPS sensor was active or not, depending on the speed threshold. After the learning process was finished, we moved learned model back again to the mobile device and based on that, we were adjusting sampling rates of the sensors.

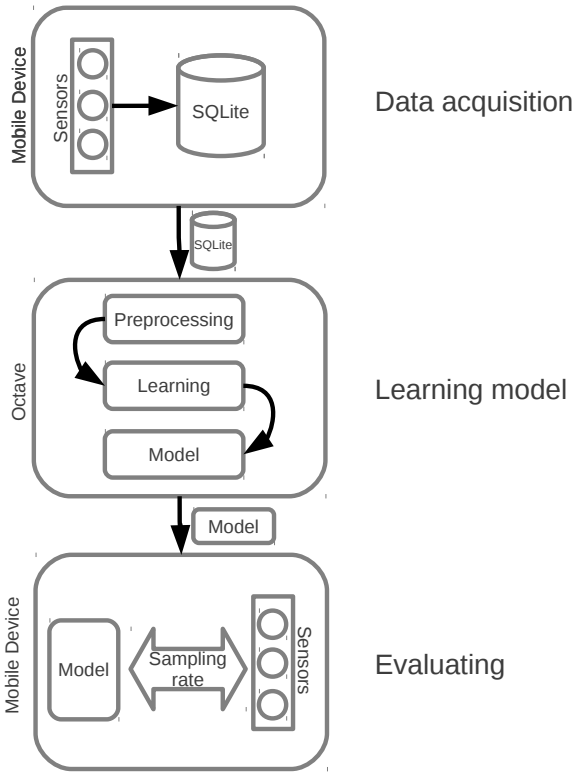


Figure 2. Learning and evaluation process of the intelligent middleware approach.

In the following Sections more details about implementation and evaluation results is presented.

A. Implementation

The prototype of the learning algorithm was written in Octave, and the evaluation of the learned model was performed on a Samsung Galaxy S II smartphone with Android 4.2 Jelly Bean installed.

The Octave learning phase was performed according to the learning algorithm described in Section IV. Fragment of a gradient descent source code responsible for learning parameters of the model is presented below.

```
for i=1:max_iterations,
  derivatives = zeros(nparams,1);
  for j=1:nsamples,
    product = -Y(j)*(X(j,:)*weights);
    error(i) += log(1+exp(product))/nsamples;
    sigm = sigmoid(product);
    for k=1:nparams,
       derivatives(k)=derivatives(k) +
```

```
            sigm*(Y(j)*X(j,k));
    end
  end
  derivatives = -derivatives / nsamples;
  weights -= learning_rate*derivatives;
end
```

During the evaluation phase, we used an Android device with a model of the sensor usage habits generated by the Octave algorithm. The algorithm that was adjusting sampling rates based on the learned model, performed following steps:

- Sample GPS sensor with a rate predicted by the intelligent middleware algorithm.
- When any movement is discovered, start sampling with the highest possible rate called `baseFreq` (we fixed this to be 1 second).
- After some fixed period of time called `continuityThreshold` (10 seconds in our approach), if no sensor activity was discovered, return to sampling rate predicted by the intelligent middleware algorithm.

The source code fragment responsible for this is presented below:

```
if(timeFromLastActivity < continuityThreshold){
   newFreq = baseFreq;
} else{
  float probability =
    middleware.getProbability(clockTime);
  int multiplier =
    (1.0f - probability) * scaleFactor + 1.0f;
  newFreq = baseFreq * multiplier;
}
if(newFreq != refreshFrequency)
  rescheduleUpdatesFromProviders(newFreq);
```

B. Results

We made experiments on two identical devices carried by the same person. One device was equipped with and intelligent middleware algorithm implemented and the other does not. Both devices were fully charged at the beginning of the experiment and was not recharged during it. We decided to use speed threshold equal to 5 km/h. With lower thresholds, the difference between intelligent middleware approach and the other one was hardly visible, because of the errors in GPS sensor readings which results in fake "active" states. As depicted in Figure 3, the intelligent middleware approach allowed for 50% battery saving than in case of the device without the algorithm implemented.

Figure 3 presents a proportion of the time that both devices worked on the battery. The right plot shows the time that the device without the intelligent middleware implemented worked, and the left plot presents a work time of the device with the intelligent middleware implemented.

The distance error which we define as a difference between the GPS samples generated in our approach and referenced samples generated by the approach without learning algorithm, is presented in Figure 4. The average distance error of the presented data equals 0.053 km. The high error in several

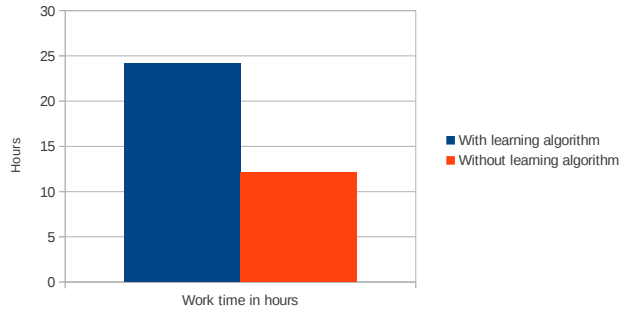


Figure 3. Difference in power consumption for device with and without learning algorithm implemented, with speed threshold set to 5km/h.

places on the plot is a result of noisy readings of GPS sensor rather than an algorithm fault.

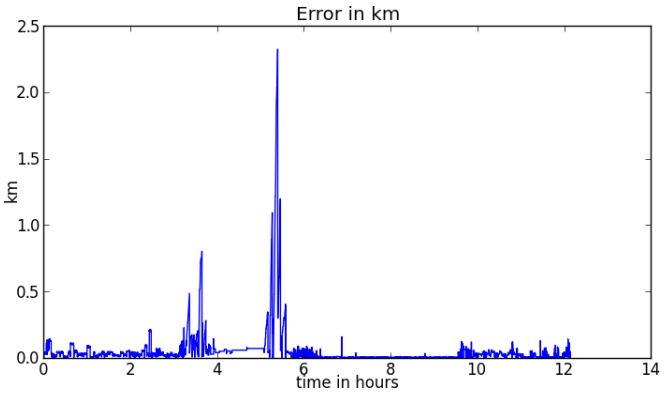


Figure 4. Error between position designated by the intelligent middleware approach and the reference data.

VI. Summary and Future Work

In this paper we presented a prototype of an intelligent middleware approach that is able to learn sensor usage habits and adjust sensor sampling rates to minimize energy consumption of the context-aware system. The middleware was presented as a part of a context reasoning platform tailored to the needs of such intelligent distributed mobile computing devices. We argue that most of the existing solutions are not fully applicable to mobile architectures, and does not fulfil energy efficiency needs of context-aware distributed systems. The presented approach was designed to solve that issue, however we believe that it is suitable for predicting not only future sensor activity, but also context-aware application needs. Hence, it is possible to get the desired context in advance, before the application actually requests it. It can be especially useful in cases when context cannot be obtained by the middleware directly from the sensor layer, but has to be for example downloaded over the internet.

We used a logistic regression algorithm to learn sensor usage model form historical data. This allowed for adjusting sampling rates of the sensors according to usage probability.

Evaluation on a real device showed that we can gain up to 50% of energy saving using this algorithm.

As a future work the implementation of the algorithm for an Android device is planned to allow real-time online learning and full evaluation of the intelligent middleware approach, not only for a GPS, but also other sensor like accelerometer, gyroscope, etc.

It is also planned to implement the learning algorithm that uses Markov chains, and compare it to existing implementation. We plan to design and develop an architecture dedicated for mobile context aware applications equipped with an intelligent middleware layer and rule based inference layer provided by the HeaRT [18] inference engine, which is a lightweight rule-based engine that uses XTT2 [19] notation for knowledge representation. This will allow for lightweight reasoning [20] and also verification of context models [21]. We plan to incorporate and evalueate the middleware in the context-aware system for monitoring threats in urban environment proposed in [22]. We also believe that it would be valuable to compare challenges and problems in mobile context-aware computing with an area connected with research about wireless sensor networks [23]. This two fields of science can possibly benefit from exchanging solutions and ideas especially regarding energy efficiency and resource limitations issues.

ACKNOWLEDGMENT

The paper is supported by the AGH University of Science and Technology Grant 11.11.120.859.

REFERENCES

[1] A. K. Dey, "Providing architectural support for building context-aware applications," Ph.D. dissertation, Atlanta, GA, USA, 2000, aAI9994400.

[2] F. Sahafipour and R. Javidan, "A comparative study of context modeling approches and applying in an infrustructure," *Canadian Journal on Data Information and Knowledge Engineering*, vol. 3, no. 1, 2012.

[3] T. Strang and C. Linnhoff-Popien, "A Context Modeling Survey," in *In: Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp 2004 - The Sixth International Conference on Ubiquitous Computing, Nottingham/England*, 2004. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.2060

[4] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive Mob. Comput.*, vol. 6, no. 2, pp. 161–180, Apr. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.pmcj.2009.06.002

[5] P. Bellavista, A. Corradi, M. Fanelli, and L. Foschini, "A survey of context data distribution for mobile ubiquitous systems," *ACM Comput. Surv.*, vol. 44, no. 4, pp. 24:1–24:45, Sep. 2012. [Online]. Available: http://doi.acm.org/10.1145/2333112.2333119

[6] I. Chronis, A. Madan, and A. S. Pentland, "Socialcircuits: the art of using mobile phones for modeling personal interactions," in *Proceedings of the ICMI-MLMI '09 Workshop on Multimodal Sensor-Based Systems and Mobile Phones for Social Computing*, ser. ICMI-MLMI '09. New York, NY, USA: ACM, 2009, pp. 1:1–1:4.

[7] J. J. Jung, "Contextualized mobile recommendation service based on interactive social network discovered from mobile users," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11 950–11 956, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2009.03.067

[8] D. Olguin, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART B: CYBERNETICS*, pp. 43–55, 2009.

[9] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, Mar. 2006.

[10] M. Atzmueller and K. Hilgenberg, "Towards capturing social interactions with sdcf: An extensible framework for mobile sensing and ubiquitous data collection." in *Proc. 4th International Workshop on Modeling Social Media*. ACM Press, 2013.

[11] J. E. Bardram, "The java context awareness framework (jcaf) – a service infrastructure and programming framework for context-aware applications," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, H.-W. Gellersen, R. Want, and A. Schmidt, Eds. Springer Berlin Heidelberg, 2005, vol. 3468, pp. 98–115. [Online]. Available: http://dx.doi.org/10.1007/11428572_7

[12] W. V. Woensel, S. Casteleyn, and O. D. Troyer, *A Framework for Decentralized, Context-Aware Mobile Applications Using Semantic Web Technology*, 2009.

[13] B. van Wissen, N. Palmer, R. Kemp, T. Kielmann, and H. Bal, "ContextDroid: an expression-based context framework for Android," in *Proceedings of PhoneSense 2010*, Nov. 2010. [Online]. Available: http://sensorlab.cs.dartmouth.edu/phonesense/papers/Wissen-ContextDroid.pdf

[14] E. Elnahrawy and B. Nath, "Context-aware sensors," in *EWSN*, ser. Lecture Notes in Computer Science, H. Karl, A. Willig, and A. Wolisz, Eds., vol. 2920. Springer, 2004, pp. 77–93.

[15] H. Chen, T. W. Finin, and A. Joshi, "Semantic web in the context broker architecture." in *PerCom*. IEEE Computer Society, 2004, pp. 277–286.

[16] T. Gu, H. K. Pung, D. Q. Zhang, and X. H. Wang, "A middleware for building context-aware mobile services," in *In Proceedings of IEEE Vehicular Technology Conference (VTC*, 2004.

[17] A. Ligęza and G. J. Nalepa, "A study of methodological issues in design and development of rule-based systems: proposal of a new approach," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 117–137, 2011.

[18] G. J. Nalepa, "Architecture of the HeaRT hybrid rule engine," in *Artificial Intelligence and Soft Computing: 10th International Conference, ICAISC 2010: Zakopane, Poland, June 13–17, 2010, Pt. II*, ser. Lecture Notes in Artificial Intelligence, L. Rutkowski and [et al.], Eds., vol. 6114. Springer, 2010, pp. 598–605.

[19] G. J. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[20] G. Nalepa, S. Bobek, A. Ligęza, and K. Kaczor, "Algorithms for rule inference in modularized rule bases," in *Rule-Based Reasoning, Programming, and Applications*, ser. Lecture Notes in Computer Science, N. Bassiliades, G. Governatori, and A. Paschke, Eds., vol. 6826. Springer Berlin / Heidelberg, 2011, pp. 305–312.

[21] G. Nalepa, S. Bobek, A. Ligęza, and K. Kaczor, "HalVA – rule analysis framework for XTT2 rules," in *Rule-Based Reasoning, Programming, and Applications*, ser. Lecture Notes in Computer Science, N. Bassiliades, G. Governatori, and A. Paschke, Eds., vol. 6826. Springer Berlin / Heidelberg, 2011, pp. 337–344. [Online]. Available: http://www.springerlink.com/content/c276374nh9682jm6/

[22] S. Bobek, G. J. Nalepa, and W. T. Adrian, "Mobile context-based framework for monitoring threats in urban environment," in *Multimedia Communications, Services and Security: 6th International Conference, MCSS 2013: Kraków, Poland. June 6-7, 2013. Proceedings*, 2013.

[23] C. S. Raghavendra, K. M. Sivalingam, and T. Znati, *Wireless sensor networks*. Springer, 2006.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 999–1005

System Design and Implementation Decisions for ParaMoise Organisational Model

Mateusz Guzek
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg
6, rue R. Coudenhove-Kalergi
Luxembourg, Luxembourg
Email: mateusz.guzek@uni.lu

Grégoire Danoy
Computer Science and Communications Research Unit
University of Luxembourg
6, rue R. Coudenhove-Kalergi
Luxembourg, Luxembourg
Email: gregoire.danoy@uni.lu

Pascal Bouvry
Computer Science and Communications Research Unit
University of Luxembourg
6, rue R. Coudenhove-Kalergi
Luxembourg, Luxembourg
Email: pascal.bouvry@uni.lu

***Abstract*—ParaMoise is a novel organisational model that permits to specify parallel and concurrent systems' organisation and reorganisation. Workflows, locks and multiple organisation managers are the entities that differentiate this model from it antecedent, the $\mathcal{M}$oise$^+$ framework. All these entities must be efficiently designed and implemented to ensure the practical usage of the theoretically formulated model. The main challenge here is the distributed synchronisation of workflows and locks, that will maximise the performance of the system. This paper presents and analyses different workflows and locks management approaches that can be used to achieve this goal: from basic centralised or middleware based solutions, towards truly decentralised coordination mechanisms.**

I. Introduction

ORGANISATIONAL models are used in Multi-Agent Systems (MAS) to facilitate the teamwork between agents. They define the coordination and cooperation mechanisms between agents, resulting in a model that can be reused in various systems and environments, without need to create a custom solution for each of them. Additionally, a model of organisation enables to explicitly represent the social aspects of a MAS, which can be useful for both agents and external observers of the MAS.

Multiple organisational models were introduced [1]–[3], each of them with its own properties and assumptions about MAS architecture. The cited frameworks are notable, as they are general purpose and can fit into multiple domains that benefit from the MAS paradigm.

In this study, we further extend the MOISE [1] organisational model, which relies on a three dimensional description of the Organisation Specification (OS). The OS consists of a Structural Specification (SS) that describes roles together with their hierarchy and possible interactions, groups, and links between roles and groups, a Functional Specification (FS) that describes the schemas, further divided into goals, which are in turn grouped into missions, and a Deontic Specification (DS) that binds the SS and FS by a set of deontic modalities, which enforce or allow agents playing specific roles to commit to missions.

Moreover, MOISE clearly divides the general description of the organisation (OS) from the instantiation of that description, called Organisational Entity (OE). An OE consists of an OS, a set of agents, and the elements that create a valid instance of the OS using this set of agents, e.g. functions that determine the current assignment of agents to roles, groups, and missions, the set of currently existing groups, the set of applied deontic modalities.

The next step in the development of organisational models is considering the benefit and the cost of running an explicit organisation infrastructure in a system. The rationale for running an organisation is to facilitate reaching the desired states by the system. In case of a dynamic environment, it is likely that the organisation may decrease its efficiency due to changes in its environment. As a result, the *reorganisation* may be necessary to adapt the system [4]. The urge for the reorganisation can be especially important for large scale distributed systems, that may trigger reorganisation not only because of external environmental changes, but also due to internal events in MAS. Additionally, the concept of *artifact*, a general and abstract representation of object that can be perceived and used by agents, may be applied to represent organisation [5].

The state-of-the-art development in the field of modelling is ParaMoise [6], that enhances its predecessors by introducing novel concepts coming from the distributed and parallel computing field: workflows, locks, alternative or redundant execution paths, transactions, and failure handling mechanisms, as well as multiple managers of organisation. In effect, the resulting model offers more possibilities to execute parallel and concurrently, without removing or diminishing any of the properties of the antecedent models. The final goal of this development is improving the distribution properties of a MAS, which shall result in an increased performance and reliability, which are essential for dynamic, large scale systems.

However, ParaMoise is a theoretical approach, which application and performance will depend on a proper design and implementation of the proposed mechanisms. In this work, we aim to address this issue by discussing possible alternative designs. In this context, our contributions are the proposals of various design and implementation possibilities for the ParaMoise model divided into two groups:

978-83-60810-53-8/$25.00

a) Centralised: classical tools such as databases that supports transactions, which can be possibly seen as artifacts by MAS.

b) Decentralised: artifacts distributed among agents. In this context we consider that the responsible Organisation Manager (OrgManager) could host the organisational artifacts, which can be also delegated to a new auxiliary Organisation Carrier (OrgCarrier) role that sole purpose is to host the organisational artifacts. We also present the possible solutions for preventing deadlocks that can occur in case of multiple locks, followed by some further refinements of organisation to minimise the resulting synchronisation overheads. Finally, we highlight the impact of using distributed algorithms, as they give agents the possibility to choose the organisational artifact type according to their needs.

The rest of the article is organised as follows. Section II provides a state-of-the-art on ParaMoise, artifacts and distributed synchronisation. Section III presents a centralised solution approach that fulfils the basic requirements for the implementation, while section IV discusses the organisation distributions possibilities. Finally, section V describes the advantages of distributed artifacts and section VI concludes the paper.

II. State of the Art

This section is divided into three parts. Section II-A describes in more details the ParaMoise model, then section II-B presents the artifact-based approaches that are important in the discussed design concepts, and finally section II-C describes the basic algorithms that can be used for the distributed concurrency control.

A. ParaMoise

This section describes the main concepts introduced in the ParaMoise [6] organisational model. ParaMoise is a novel organisational model based on the MOISE [1] and $\mathcal{M}$oise$^+$ [7] models. One of the assumptions of Moise models is the full *autonomy* of agents, i.e. the agents decide by themselves what to do and when, given their current deontic situation, which in turn defines possible rewards or penalties for some performed actions. As a result, the system does not need any central scheduler that will assign tasks to agents, contrary to other state-of-the-art solutions such as GPGP/STÆM [3].

The ParaMoise model is based on the state-of-the-art definitions of organisational models [7], [8], and defines an OE as a tuple [6]: $\langle OS, \mathcal{A}, \mathcal{GI}, \mathcal{SI}, \mathcal{O}, sg, ar, am\rangle$, where OS is the organisational specification; $\mathcal{A}$ is the set of agents; $\mathcal{GI}$ is the set of group instances; $\mathcal{SI}$ is the set of social schemes; $\mathcal{O}$ is the set of current deontic modalities; $sg : \mathcal{GI} \rightarrow \mathbb{P}(\mathcal{GI})$ maps each group to its subgroups; $ar : \mathcal{A} \nrightarrow \mathbb{P}(\mathcal{R} \times \mathcal{GI})$ maps agents to the roles they are playing in the groups; $am : \mathcal{A} \nrightarrow \mathbb{P}(\mathcal{M} \times \mathcal{SI})$ maps agents to the missions they are committed to in the social schemes.

The first major contribution of the ParaMoise model is the *Workflow Specification* (WFS), which is a way to present goals and dependencies between them as a workflow. WFS is defined as [6]: $\langle \mathcal{G}, \mathcal{E}, \mathcal{M}, mo, nm, alt, fh\rangle$, where $\mathcal{G}$ is the set of global goals; $\mathcal{E}$ is the set of precedence relations; $\mathcal{M}$ is the set of mission labels; $mo : \mathcal{M} \rightarrow \mathbb{P}(\mathcal{G})$ is the function that specifies the mission set of goals ; $nm : \mathcal{M} \nrightarrow \mathbb{N} \times \mathbb{N}$ specifies the boundaries (min,max) of number of agents committed to the mission in well formed WFS; $alt : \mathcal{E} \rightarrow \mathbb{P}(\mathcal{E})$ is the function specifies the precedence relations alternatives; $fh : \mathcal{G}_p \nrightarrow \mathbb{N}$ specifies the failure handling mechanism for a primitive goal in terms of maximum number of allowed repetitions. An example of WFS is presented in Figure 1.

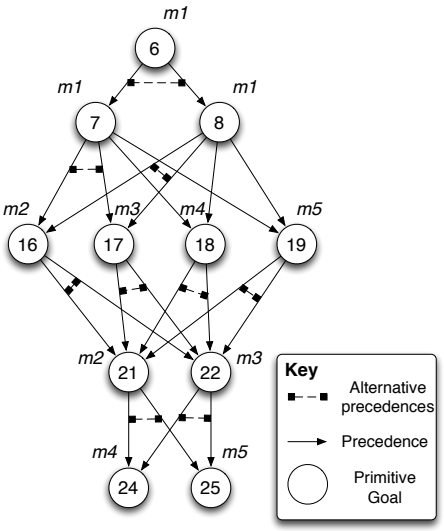


Fig. 1. An example of ParaMoise WFS.

An instantiation of a WFS by some agents is referred to as a *Workflow* (WF). The latter is defined as a tuple [6] $\langle WFS, es, gs, exe, gf\rangle$, where WFS is the workflow specification, $es : \mathcal{E} \rightarrow \{active, inactive, discarded\}$ is the function that maps edges to their activity status label; $gs : \mathcal{G}_p \rightarrow \{waiting, possible, executing, suspended, achieved, discarded\}$ is the function that specifies statuses of primitive goals; $exe : \mathcal{G}_p \nrightarrow \mathbb{P}(\mathcal{A})$ is the function that specifies the set of agents executing a goal; $gf : \mathcal{G}_p \rightarrow \mathbb{N}$ specifies numbers of repetitions of primitive goals. The status of the goals in the system changes according to the state transition diagram presented in Figure 2. The final example of a workflow usage is presented in Figure 3, which presents the capability of tracking the execution status.

The WFS and WF enable more parallelism, since they permit to represent an arbitrary structure of dependencies between goals. They are combined with *locks* to ensure mutual exclusion during reorganisation. The locks are defined as $\langle ROE, type\rangle$, where ROE is the reduced organisational entity (the elements of the OE on which the lock applies) and $type \in \{read, write\}$ specifies the type of the lock. Before a reorganisation, a lock must be created for all modified (write lock) or accessed (read lock) elements of the organisation. To minimise the scope of locks, they can be applied to a subset of

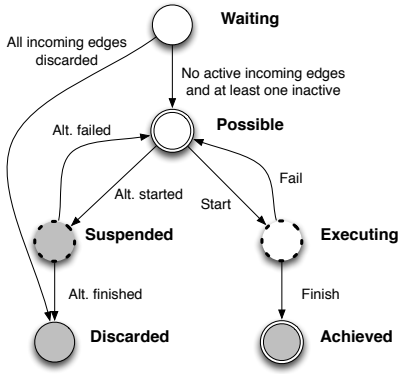


Fig. 2. State transition diagram of goals status in ParaMoise model [6].

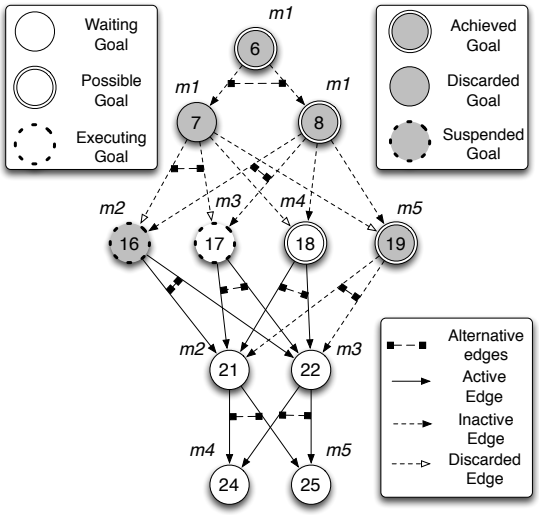


Fig. 3. An example of WF during execution

a set, or for a subdomain of a function. ParaMoise also introduces multiple managers of the organisation (OrgManagers), which fulfil the requirement for an effectively concurrent system.

An efficient access to the workflows and locks is crucial to achieve a high performance of concurrent and parallel execution and reorganisation in a system that applies ParaMoise. Therefore, it is important to find a design appropriate for a considered scenario. In this article we present two approaches: a basic centralised one, applicable for small scale systems, which is interesting as it underlines the basic requirements of the implementation of the model, and the decentralised approach with various possible design choices and their anticipated consequences.

The essential requirements for any implementation of the organisational model are:

1) the existence of all elements of the defined organisation,
2) the accessibility of all the existing elements by the agents with appropriate permissions,
3) the execution of workflow that ensures their correct state transitions,
4) the effective lock and reorganisation mechanisms.

The concept of artifacts can directly meet these requirements.

B. Artifact-Based Frameworks

ORA4MAS (Organisational Artifacts for Multi-Agent Systems) [5] is an approach that describes organisational entities as *artifacts*, based on the Agents and Artifacts (A&A) framework [9]. Artifacts are abstraction of interactive objects that can be perceived and used by agents. An artifact is defined by observable *properties* that represent its state, *operations* that determine its functionality, *links* that describe its relations with other artifacts, *events* that can be emitted in certain conditions, and corresponding *manual* that instructs the agents how to use artifacts. In this context we can see the artifacts as tools that can be used to achieve the goals of agents. ORA4MAS describes the theoretical foundations for using artifacts as the basis of reorganisation, but neglects some system designs aspects. It is a centralised solution based on the paradigms of the A&A framework.

Another drawback of ORA4MAS is the absence of reorganisation. It is partially covered by the JaCaMo [8] framework, which uses a central artifact to perform reorganisation by a single OrgManager that requires halting the whole organisation.

In this context, ParaMoise offers parallel and concurrent reorganisation at runtime, performed by multiple OrgManagers. In the same way ParaMoise enhances standard execution mechanisms, enabling arbitrary precedences between goals, novel possibilities of alternative goals, and a failure handling mechanism. However, the ParaMoise model does not propose any exact design and implementation, but only mentions the usage of artifacts as a perspective. This work proposes to answer to this need. In the remainder of the paper we discuss alternative scenarios, arguing that using well-known and established solutions results in an abundance of choices in which artifacts are one of the basic concepts, being an interface between agents and organisation support systems.

C. Decentralised Synchronisation

Decentralised synchronisation problems are crucial for distributed computing and distributed systems. In contrast to easier case of centralised systems synchronisation, they must be solved taking into account such properties as lack of the global knowledge or communication delays. As a result, a number of solutions to the problem were proposed to solve some main issues for ParaMoise: mutual exclusion (locks) and transactions [10].

1) Mutual Exclusion: Exemplary mutual exclusion algorithms are the Ricart and Agrawala algorithm [11] and token-based algorithms. We focus here on the basic algorithms, despite further refinements were proposed (e.g. Maekawa [12] or Sigma [13]) as they underline the common characteristics of this type of algorithms.

The Ricart and Agrawala algorithm requires communications between all agents, that have the possibility to access the critical section. As a result, the cost of synchronisation is $2(n-1)$, where n is the number of agents. Additionally, in the basic forms, the failure of any of the agents disturbs the proper work of the algorithm. These limitations prohibit usage of such algorithms in a large groups of agents. On the other hand, the algorithm is fair and in optimistic case leads to a fast resolution of the problem.

The token-based algorithms [10] ensure mutual exclusion by using a unique token for a critical section. These can be applicable for a set of distinct critical sections or resources. The token can be passed according to various strategies, e.g. continuously moving in the ring or using more sophisticated hierarchical structures.

2) Transactions: Another aspect of concurrency control is proper transaction handling, which is required for an effective WF management. Three main approaches can be distinguished [10].

Two phase locking (2PL) [14] is based on the standard lock mechanism. The locks are created in the two phases: in the first phase the locks are consecutively acquired according to the needs of a transaction and then in the following phase they are released.

The optimistic concurrency control [15] assumes that violations of mutual exclusion are rare, and it is possible to repair the potential damages done by such violation. As a consequence, there must exist an efficient repair mechanism and the collisions cannot be destructive. This is most effective in the case of relatively rare occurrence of conflicting write operations in a system [16].

The pessimistic timestamp ordering [10] is the last approach presented here. It associates the demands to access elements being part of a critical section with additional read and write timestamps. During the evaluation of the incoming transactions, the timestamps are used to check if the incoming transactions conflict with the ongoing ones. The approach is safer than the optimistic concurrency control, as it avoids the potential problems instead of resolving them.

III. Centralised Design

The centralised solution for synchronisation can be achieved with classical tools used for mutual exclusion and transactions. A central entity is responsible for keeping all the information about the state of the organisation. In case any agent needs to acquire knowledge about any part of the organisation, e.g. the agent's roles, obligations or known agents, it can query this entity.

The concurrent access control is performed centrally and in result does not pose a major challenge. A system of role-based access can effectively enforce that only the entitled agents can access specific elements of the organisation. The transaction mechanism can be straightforwardly applied to the execution of workflows, ensuring the correct state transitions of goals. Finally, lock creation and checking is done by a single entity that can prohibit any forbidden overlaps.

As a result, we can see this entity as a database which stores all elements of the organisation and grants access to them only to the roles that have the required permission. The workflows are stored inside the database and their status can be changed using the mechanism of transactions. The database routines ensure that the created lock does not overlap with other locks.

For an agent in the system, the database is seen as an artifact that stores the information about the organisation with well-defined interfaces to perform organisational actions. It can return information about the organisation or be exploited as a synchronisation tool used for efficient teamwork, as it holds the workflow state. Finally, the artifact has interfaces that enable OrgManagers to change the shape of the organisation by modifying the current state of the organisation in a safe way.

IV. Decentralised Design

This section presents the variety of possible design choices for the ParaMoise model and discusses their properties. Firstly, it describes the basic decentralisation capabilities and concepts in section IV-A, which introduces the decentralised artifacts described in section IV-B. Artifacts management problems and the corresponding organisational challenges are presented in section IV-C, while the solving of possible deadlock problems is discussed in section IV-D.

A. Decentralised Middleware

The most straightforward way to decentralise the system is to use an existing solution to distribute the centralised middleware, e.g. a database. This involves correct replication schema together with synchronisation of replicas. From the MAS design perspective these problems of distributed computing are out of the scope of this paper, as logically there is still one entity that is distributed, possibly with multiple equivalent interfaces. Therefore, the following paragraphs describes the applicability of decentralised synchronisation algorithms for the ParaMoise model. Following the structure of Section II-C, we describe two main issues: Mutual Exclusion and Transactions.

1) Mutual Exclusion: Solving the mutual exclusion problem is an essential design decision for ParaMoise, as it effectively determines the locks mechanism, its performance and properties. The Ricart and Agrawala algorithm is applicable for the ParaMoise model. The need for broadcast communication may rise scalability issues, however proper division schemas may result in more applicable solutions, which are discussed further in Section IV-D. The other discussed solution, token-based algorithms, seems to have limited applicability in the ParaMoise model. The lock in ParaMoise could have an arbitrary form, which makes the token impractical. There could be either a large number of tokens that could create significant overhead, or in the opposite case a small amount of token responsible for major organisational elements would decrease the possible number of concurrent reorganisations. Additionally, gathering multiple

tokens to perform reorganisation may lead to increased waiting time and deadlocks.

2) Transactions: The transactions mechanism is necessary to properly modify the status of goals during progress of WF, which is the core issue for the efficiency of the parallel and concurrent execution in a ParaMoise system. A WF is also used to express all the schemas in the organisation, including reorganisation processes. 2PL conceptually fits the ParaMoise locks and could be directly applied. In the case of using optimistic concurrency control for reorganisation, there is a need to ensure that the occurrence of a conflict will not result in breaking the consistency, by applying an effective rollback mechanism. The property of allowing the conflicts to happen is not acceptable in MAS organisations, as achieving some of the goals by agents can be impossible to undo, making the rollback impossible. Additionally, the changes of reorganisation shall be directly mapped into the behaviour of agents, which could lead to an inconsistent state in an organisation. Pessimistic timestamp ordering is applicable thanks to its more cautious nature. As it ensures that the organisation will be unique at any moment in time, it is applicable for usage in ParaMoise. The specifics of the solution require a mechanism for comparing the timestamps, which could be directly performed by an artifact created for each WF. We discuss more generally the usage of decentralised artifacts in the following section.

B. Decentralised Artifacts

The next step toward decentralised system is a logical distribution of artifacts among the distributed system. In this way, the agents are no longer using the monolithic organisational artifact, but they access a set of distinct artifacts. An example of such a division may be a system that stores the definition of each role as a separate artifact.

The logical division can lead to practical consequences: as the elements of the organisation are separated among artifacts, locks are distributed among the artifacts, therefore decreasing the complexity of checking for possible conflicts. Additionally, this approach eliminates performance bottleneck, and decreases the risk of single point of failure. On the other hand, larger reorganisations can require interactions with several artifacts, increasing therefore the complexity of the operation, leading to the possibility of deadlocks, and in case of lack of redundancy failure of any of the artifacts can lead to breaking down the whole organisation.

C. Role-Based Synchronisation

The concept of decentralised artifacts can be further advanced by merging it with the concept of roles. Agents of specific roles would be responsible for maintaing organisational artifacts. The responsibility may hold for the whole system, or for a specific group. A natural choice for the role responsible for artifacts is the OrgManager, however this solution would add another functionality to this role. As OrgManagers are already responsible for coordinating organisation and executing reorganisation, we propose a new role in the structural specification: Organisation Carrier (OrgCarrier).

The OrgCarrier responsibility is to maintain the organisational artifacts. To keep the control over parts of the organisation, OrgManagers have authority over OrgCarriers, i.e. the latter should follow the orders of the former. We present the novel structural specification in Figure 4. The authority link of the OrgManager pointing to the root role *soc* is transitively propagated to the OrgCarrier role. However, this is the only connection of OrgCarrier, since none of *Org* roles, except OrgManager, has knowledge about the OrgCarrier. This structure can fulfil its goals, being transparent for the rest of the system.

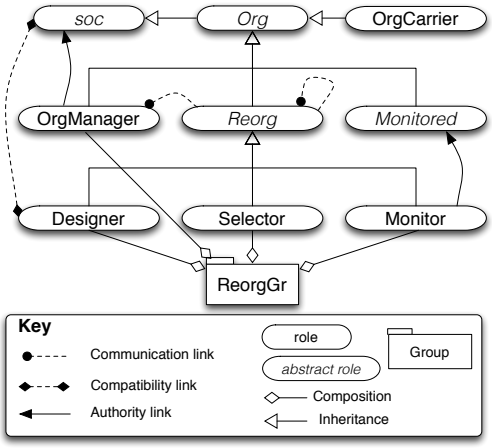


Fig. 4. Organization group structure with OrgCarrier role.

D. Resolving Deadlocks

As previously mentioned, introducing multiple locks distributed among artifacts leads to the possibility of deadlocks. An example leading to a deadlock is an attempt to create locks on two organisational elements by two OrgManagers. If they actions are synchronised, but they create locks in reverse order, a deadlock occurs. Such behaviour may be simply overcome by adding a timeout for each of the locks, however this may lead to poor system performance (waiting for timeout) or aborting lock for a valid operation that lasts longer than expected [16].

Another approach to solve the problem of deadlock could be the coordination between OrgManagers, using an algorithm such as Ricart and Agrawala. Each OrgManager broadcasts the description of the lock it wants to create together with a timestamp. Other OrgManagers must reply that they allow to create the lock. In case of conflict, the OrgManager that detects it takes a decision based on the timestamps. The lock with earlier timestamp has priority. Tie braking may be implemented, e.g. favouring the lowest agent ID number. The inherent drawback of this solution is its scalability: each agent must receive and send a message coming from the initiator of the procedure. Additionally, unreliable channels or agents that occur in dynamic systems may break this schema,

with the probability of failure growing with the number of OrgManagers.

The organisational model properties can be used to mitigate such negative behaviours. The responsibility of a subset of OrgManagers can be restricted to specific elements of the organisation. As an example, OrgManagers could form groups associated with a role. Additionally, to ensure that major reorganisation spanning across multiple roles can be executed, there must exist a subset of OrgManagers responsible for the whole organisation. As a result, each lock concerning an element of the organisation must be checked by the subset of agents (i.e. in a group) that contain the OrgManagers responsible for this element as well as the globally responsible OrgManagers. The effectivity of such solution is based on the assumption that the major reorganisations are relatively rare. Additionally, the division of elements of the organisation must be done with a granularity that ensures correct system behaviours.

V. Discussion

A. Advantages of Artifact Driven Organisation

Representing organisations with artifacts has an additional added value: agents are controlling the artifacts, thus they have the power to alter even the organisation implementation. For example, agents may initially choose to use the centralised monolithic artifact while the system is of small scale. However, together with the growth of the system agents may decide that this solution has reached its limits and shall be changed to better scale with the new situation. Then, by correct mapping of the existing organisation to new artifacts agents can replicate the existing OE and start to use the new type of artifacts.

Agents can learn how to use different organisational artifacts, what are their strong and weak points in terms of performance, reliability, recoverability, etc. Moreover, this solution can be used to perform updates, maintenance or archive the organisational artifacts. In case one type of artifacts starts to present erroneous behaviour, agents have possibility to choose another one.

VI. Conclusion

The ParaMoise model can be used using various designs and implementations. From the discussed alternatives, we see the Ricart and Agrawala family of algorithms, 2PL, and pessimistic timestamp ordering as the most fitting low-level primitives. We consider that artifacts could play a major role as interfaces between agents and systems that agents use. Artifacts are easy to distribute, can embed specific access control and synchronisation mechanisms, and they enhance the autonomy of agents. The paper also introduces the OrgCarrier role that can facilitate the management of the organisation. The properties of decentralised algorithms may require additional structure of the OrgManagers, for example by adding managers responsibility zones.

The future work includes experimental testing of the discussed solutions as well as implementing ParaMoise as a general purpose framework. We intend to use ParaMoise in a system optimising and managing Cloud Computing infrastructures, which could validate the approach. In this context, the optimisation of an organisation model to achieve the system objectives is a prospective research direction.

Acknowledgment

M. Guzek acknowledges the support of the National Research Fund of Luxembourg (FNR) and Tri-ICT, with the AFR contract no. 1315254. This work was completed with the support of the FNR INTER-CNRS-11-03 Green@cloud.

References

[1] J. Hübner, J. Sichman, and O. Boissier, "A model for the structural, functional, and deontic specification of organizations in multiagent systems," in *Advances in Artificial Intelligence*, ser. Lecture Notes in Computer Science, G. Bittencourt and G. Ramalho, Eds. Springer Berlin / Heidelberg, 2002, vol. 2507, pp. 439–448.

[2] V. Dignum, "A model for organizational interaction: based on agents, founded in logic," Ph.D. dissertation, Proefschrift Universiteit Utrecht, 2003.

[3] V. Lesser, K. Decker, T. Wagner, N. Carver, A. Garvey, B. Horling, D. Neiman, R. Podorozhny, M. N. Prasad, A. Raja, R. Vincent, P. Xuan, and X. Q. Zhang, "Evolution of the gpgp/tæms domain-independent coordination framework," *Autonomous Agents and Multi-Agent Systems*, vol. 9, pp. 87–143, 2004.

[4] J. Hübner, J. Sichman, and O. Boissier, "Using the $\mathcal{M}$oise$^+$ model for a cooperative framework of mas reorganisation," in *Advances in Artificial Intelligence, SBIA 2004*, ser. Lecture Notes in Computer Science, A. Bazzan and S. Labidi, Eds. Springer Berlin / Heidelberg, 2004, vol. 3171, pp. 481–517.

[5] J. Hübner, O. Boissier, R. Kitio, and A. Ricci, "Instrumenting multi-agent organisations with organisational artifacts and agents," *Autonomous Agents and Multi-Agent Systems*, vol. 20, pp. 369–400, 2010.

[6] M. Guzek, G. Danoy, and P. Bouvry, "Paramoise: Increasing capabilities of parallel execution and reorganization in an organizational model," in *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems, AAMAS'13*. IFAAMAS, May 2013, pp. 1029–1036.

[7] J. F. Hübner, J. S. Sichman, and O. Boissier, "Developing organised multi-agent systems using the $\mathcal{M}$oise$^+$ model: Programming issues at the system and agent levels," *International Journal of Agent-Oriented Software Engineering*, vol. 1, no. 3/4, pp. 370–395, 2007.

[8] A. Sorici, G. Picard, O. Boissier, A. Santi, and J. F. Hübner, "Multi-Agent Oriented Reorganisation within the JaCaMo infrastructure," in *Proceedings of The Third International Workshop on Iinfraestructures and tools for multiagent systems: ITMAS 2012*, Valencia, Espagne, 2012, pp. 135–148.

[9] A. Ricci, M. Viroli, and A. Omicini, "The A&A programming model and technology for developing agent environments in MAS," in *Programming Multi-Agent Systems*, ser. LNCS, M. Dastani, A. El Fallah Seghrouchni, A. Ricci, and M. Winikoff, Eds. Springer, Apr. 2008, vol. 4908, pp. 89–106, 5th InternationalWorkshop (ProMAS 2007), Honolulu, HI, USA, 15 May 2007. Revised and Invited Papers. [Online]. Available: http://www.springerlink.com/content/92370q174328841j/

[10] A. S. Tanenbaum and M. V. Steen, *Distributed Systems: Principles and Paradigms*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[11] G. Ricart and A. K. Agrawala, "An optimal algorithm for mutual exclusion in computer networks," *Commun. ACM*, vol. 24, no. 1, pp. 9–17, Jan. 1981. [Online]. Available: http://doi.acm.org/10.1145/358527.358537

[12] M. Maekawa, "A n algorithm for mutual exclusion in decentralized systems," *ACM Trans. Comput. Syst.*, vol. 3, no. 2, pp. 145–159, May 1985. [Online]. Available: http://doi.acm.org.proxy.bnl.lu/10.1145/214438.214445

[13] W. Chen, S. Lin, Q. Lian, and Z. Zhang, "Sigma: A fault-tolerant mutual exclusion algorithm in dynamic distributed systems subject to process crashes and memory losses," in *Proceedings of the 11th Pacific Rim International Symposium on Dependable Computing*, ser. PRDC '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 7–14. [Online]. Available: http://dx.doi.org/10.1109/PRDC.2005.57

[14] P. A. Bernstein, V. Hadzilacos, and N. Goodman, *Concurrency Control and Recovery in Database Systems*. Addison-Wesley, 1987.

[15] H. T. Kung and J. T. Robinson, "On optimistic methods for concurrency control," *ACM Trans. Database Syst.*, vol. 6, no. 2, pp. 213–226, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/319566.319567

[16] Coulouris, J. Dollimore, and T. Kindberg, *Distributed Systems: Concepts and Design (4th Edition) (International Computer Science)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1007–1012

Using the Evaluation Nets Modeling Tool Concept as an Enhancement of the Petri Net Tool

Michał Niedźwiecki, Krzysztof Cetnarowicz
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow, Poland
Email: {nkg, cetnar}@agh.edu.pl

Krzysztof Rzecki
Cracow University of Technology
ul. Warszawska 24, 31-155 Krakow, Poland
Email: {krz}@iti.pk.edu.pl

***Abstract*—Petri net modeling has well-known algorithms, so it is easy to develop computer tools to build, edit, and analyse these networks. These tools are designed to be able to add extensions giving additional functionality, such as an extension for evaluation networks.**

Evaluation networks are not as popular as Petri Net modeling, but they turns out that, in modeling of some of the problems, evaluation networks makes analysis of them very clear and intuitive. Unfortunately there are no mathematical tools and computer programmes for evaluation networks use. Fortunately, under certain assumptions, an evaluation network can be converted into a Petri net. This article presents an idea of how to convert an existing Petri net computer programme to draw evaluation nets and convert them into Petri nets in order to use existing tools for Petri net analysis.

Evaluation nets are well suited for modeling negotiation protocols between two parties represented by servers or software agents. This article provides an example of such a protocol presented in three versions: a sequence diagram UML, Petri net and Evaluation nets.

I. Introduction

PETRI net is one of the few languages for modeling distributed systems. It was invented in the 1960s by the German mathematician Carl Adam Petri, and described in his doctoral dissertation [1]. Petri nets are used to model concurrent systems [2], the discrete [3] synchronisation process [4], etc. They are used in computer science, and in other fields such as biology, medicine and chemistry [5].

Modeling Petri nets may be performed by a number of computer programmes, such as WoPeD [6], YASPER [7], CPN Tools [8] or PIPE2 [9]. They allow for graphical editing, interactive visualisation, and analysis of automatic Petri net.

A new areas of Petri net application arose, so various extensions were created. As a result, there are coloured Petri nets [10], timed Petri nets [11], stochastic Petri nets [12], etc.

Among these extensions are evaluation nets (*E*-Nets), the first use of which was associated with an analysis of the operation of information systems. *E*-Nets were invented by Garry J. Nutt and Jerre D. Noe, and described in Nutt's dissertation [13]. *E*-Nets are very poorly distributed. Teaching materials for the *E*-Nets are not widely available, and nor are computer programms for it. Unfortunately, *E*-Nets were not developed particularly dynamically for many years.

E-Nets are well-suited for modeling communication between systems [14]. However, the lack of programmes that work with *E*-Nets makes use of them very difficult, and the implementation of such programmes from scratch increases the cost of the study.

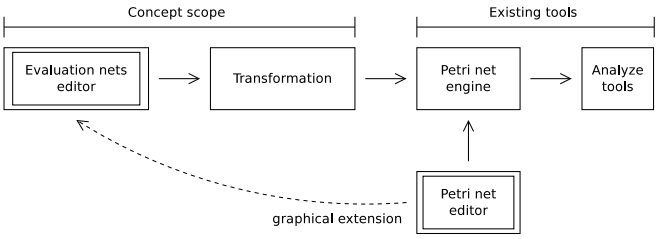


Fig. 1. Petri net editor is extended to support elements of evaluation nets. Then, when you call the analyser, evaluation nets will have been transformed to Petri net and as such referred to the process analyser.

The main motivations for writing an article on this theme are:

1) Evaluation networks are useful for designing algorithms for negotiations, such as between agents, servers, etc., yet their universal notation is very readable for humans.
2) The project of an algorithm may be verified by including the performance simulation of the evaluation nets. But there is no such network simulator.
3) Some elements of *E*-Nets can easily be presented in Petri net, so existing Petri net simulators can be used.

Our concept uses an opportunity to convert *E*-Nets (some restrictions were imposed) to Petri nets, and uses existing Petri net analysers for analysis by a visual extension of an existing Petri net editor, which draws evaluation nets on the screen and sends the ready Petri net to the analyser (Fig. 1).

The terms *E*-Nets and evaluation nets usually are used interchangeably, but for the record, in the rest of this article the names will be applied differently. The term "*E*-Nets" will be used for the original evaluation nets [15] and the variation created for this concept will be referred to as "evaluation nets."

The formalism of Petri Net were well described in [16] i [17]. That concept is based on those descriptions of Petri Net and it's extensions (Table I).

II. Evaluation nets

An *E*-net is a connected set of locations over the set of allowable transition schema. The formalism of original *E*-nets was described in [18] and [15].

978-83-60810-53-8/$25.00

In oryginal *E*-nets location may be connected with only one input and one output transition. Location in *E*-Nets can have no token (empty) or one token (full). For example in `F-TRANSITION` (Fig. 6) firing is not possible when the c location is full.

Since their creation, *E*-Nets have evolved. Some extensions were presented in [19]. These extensions were used in [14].

A lot of *E*-Nets structures can be transformed to Petri nets with coloured tokens [10] and an inhibitor arc [20] extension.

Token can have attributes. Attributes are in *Resolution procedure* and *transition procedure*. If attributes determine *E*-Nets, then conversion to Petri nets may be impossible. Therefore, for the simulation, we have to use external procedures to process the data in the token's attributes and to decide about the firing transition.

Unfortunately the usage of many external procedures in the cases of resolution locations and transition procedures seems to make the proposed model complex. The Petri net tools for analysis are not able to solve or make easiest the usage and the efficiency of proposed model. However, in some cases, it's enough to replace those procedures by simple conflict structures.

If we want to create an automatic converter from *E*-Nets to Petri nets, then we must impose additional restrictions on *E*-Nets. In this article we describe our variation of *E*-Nets. Some extensions are graphical only and these improve readability, but some extensions change the *E*-Net flow:

1) multi-connections between locations and transitions were Enabled, but this may produce conflict structures (similarly to with Petri nets),
2) if a conflict structure contains resolution locations then the order of firing is determined by the execution of these *resolution procedures*,
3) coloured token — a token with enumerated value (for example: colour) as an attribute,
4) transition schema — may use token colour instead of 0 and 1 values (something like switch-cache transition),
5) resolution location — can return a coloured token, return value cannot be determined from other tokens but can be randomised,
6) outer request location — a special kind of resolution location, represents request coming from outside the evaluation net (indicated by a rectangle with a double line),
7) transition with sender — transitions which are able to send tokens (represented by a horizontal arrow labeled with the corresponding message name), always after firing.

All these differences between evaluation nets and Petri nets, our extensions and transitions are described in Table I.

III. Transforming evaluation nets into Petri nets

In [18], there were described 5 primitives. Primitives are basic (trivial) constructions of evaluation nets (Fig. 6). All more complex evaluation nets are built by joining and extending of those primitives. *E*-Nets primitives may be replaced by their equivalents from Petri net (Fig. 7). Thus, we can acquire an evaluation net converted into a Petri net.

All original *E*-Nets primitives from Fig. 6 can be transformed with inhibitor arcs, colour tokens and conflict structures determined by e.g. controlled transitions (Fig. 7). `T-`, `F-` and `J-TRANSITION` may be transformed without any extensions but additional places, transitions and tokens are required.

Transformation of `X-` or `Y-TRANSITION` must be made by a lot of additional elements because we must emulate complex *transition schema* and all possible states of *resolution location*. With coloured Petri nets the number of additional elements may be reduced to 3 and 7 additional transitions for `X-` and `Y-TRANSITION`. Also added a lot of new connections. (Table II).

Transformation of a triangular location is trivial (use standard connection). The transformation of transition with sender is more difficult (Fig. 5).

Other transformation issues are described in Table I.

Following such transformation, Evaluation Nets can be transformed to Petri net analysis (Fig. 1).

A. Example

It turns out that evaluation networks are well-suited for modeling protocol negotiations between the two parties, such as that shown in Fig. 2. This protocol is based on the CNP [26] and is part of Complex Negotiations [27]. The specified protocol defines two participants: initiator and participant. If the initiator wants to start negotiations, it sends a message to the participant CFP (call for proposal). If the participant is interested in negotiations, the initiator receives the offer. Otherwise, the initiator sends the message *refuse* to end the negotiations as a failure. When Initiator is offered, decide whether a) accept the offer (accept), thus ending the negotiations successfully, b) reject the offer and end the negotiations a failure or c) rejects the offer but made a counteroffer and waits for a response from the participant.

Thanks to evaluation networks it is possible to design a computer system using the previously described negotiation protocol. Fig. 4 shows an evaluation net divided into two subnets, one belonging to the initiator, the second to the participant. Subnets do not have direct connections with each other (using arcs), but they exchange messages.

Initially, both participants are in a state of inactivity (`Idle`). Then the initiator is ordered to enter into negotiations, and as a result a token appears in l_1. This causes firing of a_1, which in addition to placing a token in l_4 sends a message called `CFP` (call for proposal). Initiator goes to l_4 (`Wait`) and waits for a response $\mathtt{S}_i$ from the participant. If the answer does not appear within the required time, the wait can be interrupted by a request l_3 (`Timer`). This is support for an emergency situation, which ends the Initiator net evaluation of the state of l_8 (`fail`).

In the meantime, the message `CFP` is received by l_{10} (located in the participant) which results in the appearance of the token and firing of a_5. Then the participant's evaluation

TABLE I
THE DIFFERENCES AND TRANSFORMATION METHOD BETWEEN EVALUATION NETS AND PETRI NET

Evaluation nets item	**Petri net equivalent**	**Original *E*-Nets limitations**	**Limitations of our evaluation nets variation**	**Original Petri net limitations**	**Useful Petri net variations**	**Transformation**
token	token	simple token or attributed token	first attribute is reserved for colour	simple token	colored token [10], object token [21]	When we use only coloured attribute value then we can use coloured token. Otherwise we must use object token.
location connections	place connections	connected to at least one transition	similarly to Petri net	no limitations	-	No conflicts.
token in location	token in place	no tokens (empty) or one single token (full) or one attributed token (full)	no tokens (empty) or one coloured token (full) with other attributes or not	no limitations of token count	inhibitor arcs [20], capacities of places [22]	Use inhibitor arcs of places.
transition schema	enabling rules, firing rules	schema values in tuple: 0 or 1 or "*e*" (only in right hand side of a schema), enabled transition cannot be disabled, it must fired (no conflicts)	may be use colour value instead to 0 or 1, conflict resolving determined by execution finish of resolution procedure or randomise function, transition can be disabled	input places must contains one or more token, output places not constrained enabling rules	inhibitor arcs [20], reset arcs [23], capacities of places [22]	Can be emulated by additional transitions, inhibitor arcs and loop connections (connect place with transition and transition with place). In output connections we can use capacities of places or inhibitor arcs. Reset arc may be used to reduce elements and improve readability of net.
transition procedure	-	-	not use	-	object token [21]	We can use outer procedures but existing Petri nets analysers may be unusable. However we can use tools for object tokens [8].
resolution location	conflict structure	returns token (single or attrubuted) or not	may return coloured token, return value cannot be dependent from previous events, returned value may be randomised or coming from outside of the evaluation net	firing is not deterministic	for decision of firing: Controlled Petri nets [24], Timed Petri nets [11], Stochastic Petri nets [12], Labeled Petri nets [25], etc.	Resolution location has resolution procedure. Resolution procedure running in enabling phase and result of it determine decision of firing. Automatic transformation to Petri nets without using external procedures is very difficult. For verifying the use cases we can use Timed Petri nets. For time benchmarking we can use Stochastic Petri nets.
triangular location [15]	connection	two interlinked triangular locations replaces long unreadable connection	-	-	-	It is only graphical notation. We can make extension for Petri net editor for supporting invisible connection and corresponding label.
transition with sender [14]	connection	-	it is sender triangular location and transition in one and when it fired then it sends coloured token from corresponding input location	-	-	Similarly to "triangular location." This item has been introduced for the improve readability of evaluation nets. Token was sent without depending of *transition schema*.

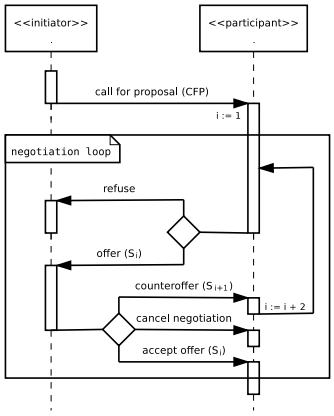


Fig. 2. Negotiation protocol in sequence diagram.

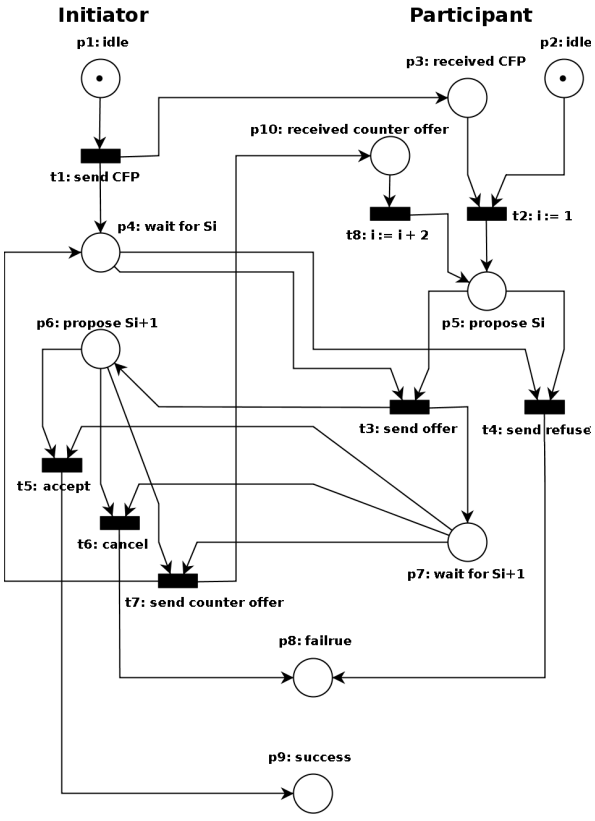


Fig. 3. Negotiation protocol in Petri net.

net goes to l_{12} (`Wait`) and expects him to make a decision (request l_{13}). As a result of the decision, a coloured token appears in l_{13} and the colour represents the `offer` (submission of proposals initiator) or `refuse` (breaking off of negotiations). The coloured token is processed by a_6, which decides whether to proceed according to evaluation net state l_{15} (`Wait`) or l_{17} (`fail`). Then the token is placed in the message S_i. The participant is in state l_{15} and is expecting a message from the initiator to be received by l_{14}. If the message does not appear

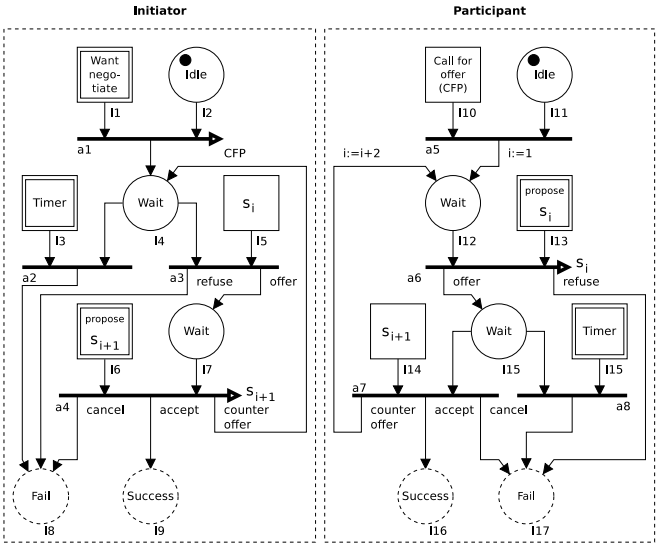


Fig. 4. Negotiation protocol in evaluation nets (our variation). S_i token can have *offer* or *refuse* enumeration value. S_{i+1} token can have *cancel*, *accept* and *counter offer* enumeration value. $i := 1$ and $i := i + 2$ have no effect on the evaluation net flow.

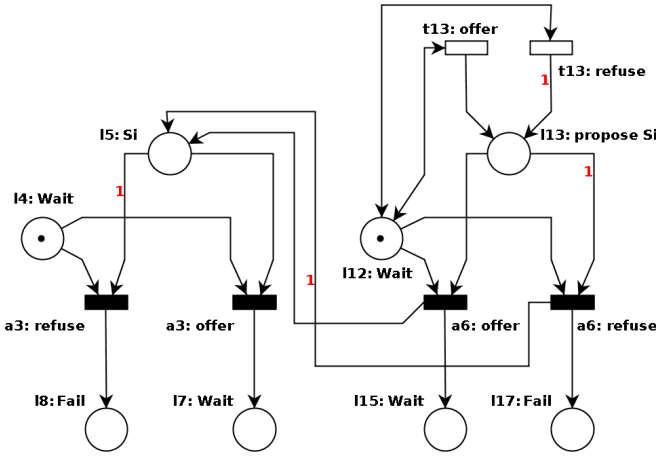


Fig. 5. Equivalent of l_4, l_5, l_7, l_8, l_{12}, l_{13}, l_{16}, l_{17}, a_3 and a_6 from Fig. 4 in coloured Petri nets. Inhibitor arcs are skipped.

TABLE II
NUMBER OF OLD ELEMENTS/ADDITIONAL ELEMENTS BY CONVERTER

Primitive	Places	Transitions		Connections		Colours
		Orig.	Lab.	Orig.	Inh.	
T-TR. Fig. 7 (a)	2/0	1/0	0	1/0	1	1/0
T-TR. Fig. 7 (b)	2/1	1/0	-	1/0	-	-
F-TRANSITION	3/0	1/0	0	3/0	2	1/0
J-TRANSITION	3/0	1/0	0	3/0	1	1/0
X-TRANSITION	4/0	1/1	2	4/6	4	1/1
T-TRANSITION	4/0	1/3	4	4/16	8	1/1
Switch: Fig. 5	8/0	2/2	2	8/12	-	2/0

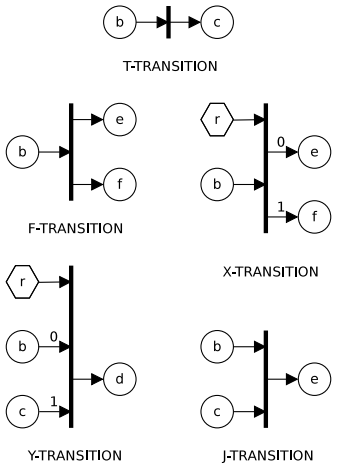


Fig. 6. Primitive *E*-Nets from [18].

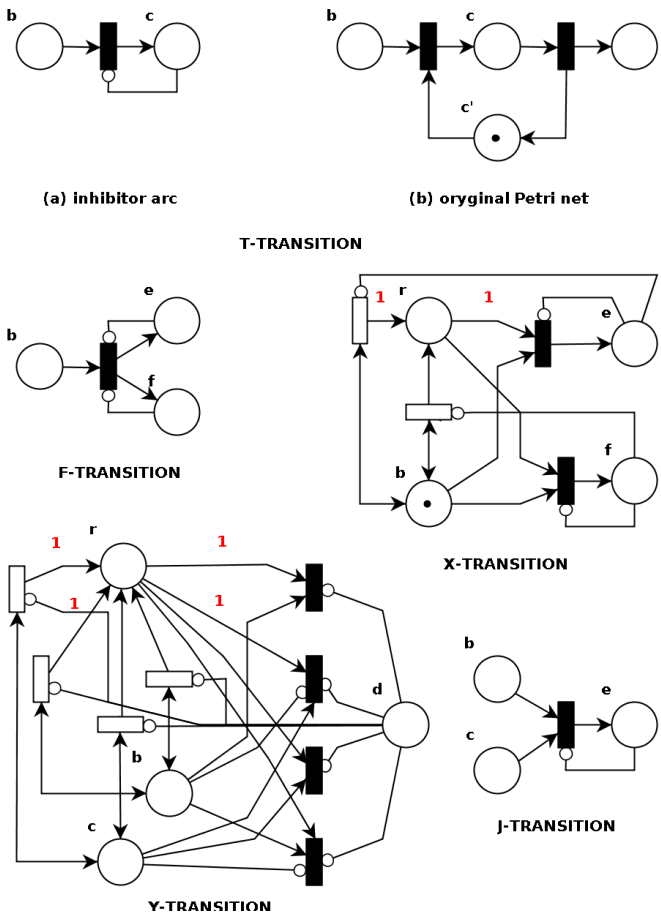


Fig. 7. *E*-Nets primitives from Fig. 6 transformed to Petri nets with extension: coloured, controlled and inhibitor arc. Additionally `T-TRANSITION` was transformed in two versions: (a) Petri net with extensions and (b) original Petri nets (b).

within a given time, the emergency procedure similar to that associated with l_3 will start in the initiator.

After receiving the message, initiator $\mathtt{S}_i$ is processed by a_3, which acts in a similar way to a_6. If the message contains the `refuse` attribute, the network is referred to the state of l_8 (`fail`), but when the message contains the attribute `offer`, the evaluation net goes into standby l_7 to make a decision by the initiator (decision will appear in the l_6).

Next, the attributed token is processed by a_4 and by a_7, which receives it through error $\mathtt{S}_{i+1}$ passed through a_4 and l_{14}.

We have tried to show the same example in Figure 3 using a controlled Petri net. This chart was prepared by hand. We tried to maintain maximum readability and so have omitted the emergency 'time out' in waiting for a message from the other side. We also abandoned the isolated, separate subnet Petri net for initiator and participant (such as in the evaluation net) to increase the number of transitions and arcs.

However, we provided standby. Cost reduction readability in this case was small because the whole Petri net is simple. Attributed tokens and resolution locations could be replaced with controlled transitions. In the former case, the initiator and participant received from their evaluation nets a request for attributed tokens. In this case, the initiator and participant will request a specific, enabled fired controlled transition.

Thus, this example shows that the use of an evaluation net instead of a Petri net makes sense visually. However, the lack of tools for analysis means that the evaluation drawings of the usefulness of the net are slightly larger than usual. This condition can be adjusted by using the described concept, which, however, entails the generation of a significant amount of additional elements. Only a portion of the changes resulting from the transformation in Fig. 5 is shown in Fig. 4.

IV. Conclusion

Evaluation nets are useful for some communication protocols e.g. the negotiation protocol from Fig. 2. However, they are not so popular, and tools for making and analysing them do not exist. Fortunately, evaluation nets with some restrictions can be transformed into Petri net.

The Petri net has many tools including mathematical tools, analysers and graphic editors. It is possible to extend the existing editor to support evaluation nets, transform them into Petri net and then use Petri net analysis tools.

Moreover, it is impossible to transform all things. The *transition procedure* and *resolution procedure* are programming instructions that operate on the attributes of tokens.

For this reason, we must either abandon them and possibly replace the *resolution procedure* function that returns a random result, or take the code of these procedures outside the Petri net and give them control over the firing transitions. In this article, we focused on the first solution. In the future, we want to focus on the other. There is a third solution, which is based on Petri net generating source code [28]. Maybe then it would be able to make a converter that is 100% compatible with the evaluation nets.

Another problem is that transformation adds a lot of new elements. Additional elements may require some modifications to analysers, but this exists only for the analyser and not for the user. For example, without a special filter on the analyser result, we saw our elements and new elements added by the converter. Without a filter our elements may disappear among the others.

Therefore, future work should focus on adjusting existing analysers. There are plans to write an interpreter, that will be able to perform net evaluation in a way similar to the BPEL interpreter.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the dean grant no. 15.11.230.082, AGH University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications. Author Michał Niedźwiecki would like to thank EFS of POKL 4.1.1 (POKL.04.01.01-00-367/08-00), European Union programme for support.

REFERENCES

[1] C. A. Petri, "Kommunikation mit automaten," Ph.D. dissertation, Universität Hamburg, 1962.

[2] F. Ahmad, H. Huang, and X. Wang, "Analysis of the petri net model of parallel manufacturing processes with shared resources," *Information Sciences*, vol. 181, no. 23, pp. 5249 – 5266, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025511003665

[3] I. Rivera-Rangel, A. Ramírez-Treviño, and E. López-Mellado, "Building reduced petri net models of discrete manufacturing systems," *Mathematical and Computer Modelling*, vol. 41, no. 8–9, pp. 923 – 937, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0895717705001548

[4] A. Merabet, "Synchronization of operations in a flexible manufacturing cell: The petri net approach," *Journal of Manufacturing Systems*, vol. 5, no. 3, pp. 161 – 169, 1986. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0278612586900439

[5] J. Will and M. Heiner, *Petri Nets in Biology, Chemistry, and Medicine: Bibliography*, ser. Computer science reports. Inst. of Computer Science, 2002. [Online]. Available: http://books.google.pl/books?id=6DB8GwAACAAJ

[6] A. Eckleder and T. Freytag, "Woped – a tool for teaching, analyzing and visualizing workflow nets," *Petri Net Newsletter*, vol. 75, Dec. 2008. [Online]. Available: http://www.informatik.uni-augsburg.de/pnnl/PDFs/PNNL75_WWW.pdf

[7] K. van Hee, O. Oanea, R. Post, L. Somers, and J. M. v. an der Werf, "Yasper: a tool for workflow modeling and analysis," in *Proceedings of the Sixth International Conference on Application of Concurrency to System Design*, ser. ACSD '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 279–282. [Online]. Available: http://dx.doi.org/10.1109/ACSD.2006.37

[8] M. Westergaard and L. Kristensen, "The access/cpn framework: A tool for interacting with the cpn tools simulator," in *Applications and Theory of Petri Nets*, ser. Lecture Notes in Computer Science, G. Franceschinis and K. Wolf, Eds. Springer Berlin Heidelberg, 2009, vol. 5606, pp. 313–322. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02424-5_19

[9] N. J. Dingle, W. J. Knottenbelt, and T. Suto, "Pipe2: a tool for the performance evaluation of generalised stochastic petri nets," *SIGMETRICS Perform. Eval. Rev.*, vol. 36, no. 4, pp. 34–39, Mar. 2009. [Online]. Available: http://doi.acm.org/10.1145/1530873.1530881

[10] K. Jensen, "A brief introduction to coloured petri nets," in *Proceedings of the Third International Workshop on Tools and Algorithms for Construction and Analysis of Systems*, ser. TACAS '97. London, UK, UK: Springer-Verlag, 1997, pp. 203–208. [Online]. Available: http://dl.acm.org/citation.cfm?id=646481.691443

[11] C. Ramchandani, "Analysis of asynchronous concurrent systems by timed petri nets," Cambridge, MA, USA, Tech. Rep., 1974.

[12] G. Balbo, "Introduction to stochastic petri nets," in *Lectures on Formal Methods and PerformanceAnalysis*, ser. Lecture Notes in Computer Science, E. Brinksma, H. Hermanns, and J.-P. Katoen, Eds. Springer Berlin Heidelberg, 2001, vol. 2090, pp. 84–155. [Online]. Available: http://dx.doi.org/10.1007/3-540-44667-2_3

[13] G. J. Nutt, "The formulation and application of evaluation nets," Ph.D. dissertation, University of Washington, 1972.

[14] M. Żabińska and K. Cetnarowicz, "Application of m-agent multi-profile architecture and evaluation nets to negotiation algorithm design," in *ISSPIT' 2002 : the 2nd IEEE International Symposium on Signal Processing and Information Technology : December 18–21*, K. Y. M. Amin, Ed., 2002, pp. 214–219.

[15] J. D. Noe and G. Nutt, "Macro e-nets for representation of parallel systems," *Computers, IEEE Transactions on*, vol. C-22, no. 8, pp. 718–727, 1973.

[16] G. Rozenberg and J. Engelfriet, "Elementary net systems," in *Lectures on Petri Nets I: Basic Models*, ser. Lecture Notes in Computer Science, W. Reisig and G. Rozenberg, Eds. Springer Berlin Heidelberg, 1998, vol. 1491, pp. 12–121. [Online]. Available: http://dx.doi.org/10.1007/3-540-65306-6_14

[17] R. Zurawski and M. Zhou, "Petri nets and industrial applications: A tutorial," *Industrial Electronics, IEEE Transactions on*, vol. 41, no. 6, pp. 567–583, 1994.

[18] G. J. Nutt, "Evaluation nets for computer system performance analysis," in *Proceedings of the December 5-7, 1972, fall joint computer conference, part I*, ser. AFIPS '72 (Fall, part I). New York, NY, USA: ACM, 1972, pp. 279–286. [Online]. Available: http://doi.acm.org/10.1145/1479992.1480030

[19] F. André and C. (Grup), *Distributed Computing Systems: Communication, Cooperation, Consistency*. Elsevier Science Pub., 1985. [Online]. Available: http://books.google.pl/books?id=dH8EAQAAIAAJ

[20] K. Reinhardt, "Reachability in petri nets with inhibitor arcs," *Electron. Notes Theor. Comput. Sci.*, vol. 223, pp. 239–264, Dec. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.entcs.2008.12.042

[21] M. Köhler, *Object Petri Nets: Definitions, Properties, and Related Models*. Univ., Bibliothek des Fachbereichs Informatik, 2003.

[22] S. Christensen and N. D. Hansen, "Coloured petri nets extended with place capacities, test arcs and inhibitor arcs," in *Applications and Theory of Petri Nets, volume 691 of LNCS*. Springer Verlag, 1992, pp. 186–205.

[23] C. Dufourd, A. Finkel, and P. Schnoebelen, "Reset nets between decidability and undecidability," in *Automata, Languages and Programming*, ser. Lecture Notes in Computer Science, K. Larsen, S. Skyum, and G. Winskel, Eds. Springer Berlin Heidelberg, 1998, vol. 1443, pp. 103–115. [Online]. Available: http://dx.doi.org/10.1007/BFb0055044

[24] L. Holloway and B. Krogh, "Controlled petri nets: A tutorial survey," in *11th International Conference on Analysis and Optimization of Systems Discrete Event Systems*, ser. Lecture Notes in Control and Information Sciences, G. Cohen and J.-P. Quadrat, Eds. Springer Berlin Heidelberg, 1994, vol. 199, pp. 158–168. [Online]. Available: http://dx.doi.org/10.1007/BFb0033544

[25] M. Cabasino, A. Giua, M. Pocci, and C. Seatzu, "Discrete event diagnosis using labeled petri nets. an application to manufacturing systems," *Control Engineering Practice*, vol. 19, no. 9, pp. 989 – 1001, 2011, <ce:title>Special Section: DCDS'09 – The 2nd {IFAC} Workshop on Dependable Control of Discrete Systems</ce:title>. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967066111000025

[26] R. Smith, "The contract net protocol: High-level communication and control in a distributed problem solver," *IEEE Transactions on computers*, vol. 29, no. 12, pp. 1104–1113, 1980.

[27] M. Niedźwiecki, K. Rzecki, and K. Cetnarowicz, "Complex negotiations in the conclusion and realisation of the contract," July 2013, international Conference on Computational Science.

[28] J.-B. Voron and F. Kordon, "Transforming sources to petri nets: a way to analyze execution of parallel programs," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, ser. Simutools '08. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 13:1–13:10. [Online]. Available: http://dl.acm.org/citation.cfm?id=1416222.1416240

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1013–1019

Analyzing Meme Propagation in Multimemetic Algorithms: Initial Investigations

Rafael Nogueras, Carlos Cotta
ETSI Informática, Universidad de Málaga
Campus de Teatinos, 29071 Málaga, Spain
Email: ccottap@lcc.uma.es

***Abstract*—Multimemetic algorithms (MMAs) are a subclass of memetic algorithms in which memes are explicitly attached to genotypes and evolve alongside them. We analyze the propagation of memes in MMAs with spatial structure. For this purpose we propose an idealized selecto-Lamarckian model that only features selection and local improvement, and study under which conditions good, high-potential memes can proliferate. We compare population models with panmictic and toroidal grids topology. We show that the increased takeover time induced by the latter is essential to improve the chances for good memes to express themselves in the population by improving their hosts, hence enhancing their survival rates.**

I. Introduction

FOUR decades ago, Richard Dawkins [3] put forward the definition of meme in analogy to the biological concept of gene. Memes were broadly characterized as units of imitation, that is, ideas or pieces of knowledge that jump from brain to brain, striving and proliferating in some cases and becoming extinct in others. Even more interestingly, memes are not static objects but dynamic entities that mutate during their lifetime; these mutations can make them more strong/interesting/useful/... thus boosting their propagation, or can have the opposite effect, making that particular mutation fade away. This plasticity explains their comparatively faster rate of adaptation with respect to biological genes.

Inspired by this notion of meme, Moscato [13] conceived a new optimization paradigm: memetic algorithms (MAs). MAs are a family of population-based optimization techniques that blend together ideas of different metaheuristics, most notably the orchestrated interplay between global (population-based) search and local (individual-based) search. The most popular incarnation of MAs features an evolutionary search engine endowed with local search add-ons. The notion of memetic evolution is here captured by the Lamarckian lifetime learning to which solutions are subject to, via the use of some local search operators. Incidentally, it has been suggested [17] to use the term *agent* rather than individual or solution in this context, to emphasize the fact that they are active entities that purposefully try to optimize the problem under consideration. We refer to [7], [12], [14], [15], [16] for a broad discussion on MAs.

This work is partially supported by Spanish MICINN under project ANYSELF (TIN2011-28627-C04-01), and by Junta de Andalucía under project DNEMESIS (P10-TIC-6083).

While memes are typically fixed in classical MAs (i.e., they are given by the particular choice of local search operators), several models trying to make them change during the optimization process have been proposed. This can be accomplished at a variety of levels. A simple possibility is the so-called 'meta-lamarckian learning' [18] in which the MA has a collection of local search operators (memes) available a priori, and some mechanism is used to decide which of them is applied on which solution and when (notice the connection with hyperheuristics [2]). A more complex approach features self-adaptation of the memes themselves. An example of this kind of self-adaptation is provided by multi-memetic algorithms (MMAs), in which each solution carries "genes" that indicate which local search has to be applied on it. These can range from simple pointers to existing local search operators to the parametrization of a general local search template [10] or even to the definition of a grammar to specify new complex local search operator [9], [11].

An interesting issue that arises in the context of MMAs is how memes propagate and spread over the population. While population dynamics has been well-studied in the case of evolutionary algorithms – e.g., [1], [6], [19], [20], the scenario is more complex in the case of memes: unlike genotypes (which correspond to solutions and thus can be evaluated according to the problem under consideration), memes can be only indirectly assessed via the effect they have on genotypes. Furthermore, memes evolve in MMAs alongside with solutions by attaching to them. Since this attachment is part of the self-adaptive process, it is up to the algorithm to discover good fits between individual pairs of genotypes and memes, and this is commonly done using only information about the genetic quality of solutions (i.e., fitness information). This work is aimed to study how memes propagate in such an environment driven by genetic selection and spatial structure. To this end, we consider and analyze an idealized model of MMAs. This model is described in next section.

II. Model Description

Let us consider an abstract model of MMAs in which each agent is characterized by a pair $\langle g, m\rangle \in D^2$, for some $D \subset \mathbb{R}$. The first member of the pair $-g-$ represents the genotype, which we equate to fitness for simplicity. As to the second member $-m-$ it represents a meme. More precisely, this value captures the *improvement potential* of that meme, that is,

978-83-60810-53-8/$25.00

a measure of how good solutions can get by applying the meme. We assume there is a monotonically increasing function $f : D^2 \to D$ encapsulating the application of a meme to a genotype, i.e., the effect of a single epoch of Lamarckian learning. Thus, an agent $\langle g, m\rangle$ becomes $\langle f(g,m), m\rangle$ after the application of the meme, where

$$\lim_{n\to\infty} f^n(g,m) = m \quad \text{if } g < m \tag{1}$$

$$f(g,m) = g \quad \text{if } g \geqslant m \tag{2}$$

Here $f^n(g,m)$ is the n-th composition of the function on the first argument, i.e., $f(f(\cdots f(f(g,m),m),\cdots,m),m)$. It must be noted that while this is very idealized characterization of the potential of a meme (since in general this potential is not going to be absolute but may depend on a complex match between the meme, the genotype and the problem landscape) it serves as an initial approximation to study several issues related to meme propagation in the agent pool.

The population P of the MMA is thus a collection of μ such agents, $P = [\langle g_1, m_1\rangle, \cdots, \langle g_\mu, m_\mu\rangle]$, endowed with a spatial structure that constrains agent communication. Let this spatial structure be characterized by a $\mu \times \mu$ Boolean matrix S, where S_{ij} is `true` if, and only if, the agent placed in the i-th location can communicate with the agent placed in the j location. Since we are interested in observing the dynamics of propagation of memes, we consider a extension of the selection-only model of evolutionary algorithms (i.e., using only selection/replacement and no variation operator) in which we add the local improvement stage of memetic algorithms. A scheme of the model is shown below in Algorithm 1.

Algorithm 1: Selecto-Lamarckian Model

```
for i ∈ [1 ··· μ] do
    INITIALIZE⟨g_i, m_i⟩;
end
while ¬ CONVERGED (P) do
    i ← URAND(1, μ) // Pick random location
    ⟨g, m⟩ ←SELECTION(P, S, i);
    g' ← f(g, m) // Local improvement
    P ← REPLACE(P, S, i, ⟨g', m⟩);
end
```

After initializing suitably the contents of the population, the algorithm engages on a cycle of selection plus improvement until the agents converge. Convergence is here approached from a memetic perspective, that is, we terminate the algorithm when the population comprises a homogeneous collection of memes (regardless of whether there is still diversity at the genetic level or not). As to the inner working of the algorithm, it resembles the uniform choice update strategy of cellular automata [21], in which the next location to be activated is selected uniformly at random with replacement.

III. Meme Propagation

Having defined the general model in the previous section, let us consider some qualitative features of meme propagation that can be extracted from it. Let us assume that selection is done by binary tournament, i.e., once a location i selected, a neighboring location j is selected from $\mathcal{N}(i) = \{j \mid S_{ij}\}$, and the agent with the best fitness is retained. As to replacement, let us assume that the improved agent replaces the agent that lost the previous tournament.

We are interested in analyzing the number of copies of each meme in the meme pool, so let us denote by $N(m,g,t)$ the number of instances of meme m attached to genotype g at time t, assuming for simplicity that D is some discrete domain. If we divide this quantity by the pool size μ we obtain $p(m,g,t)$, the fraction of the population comprising meme m attached to genotype g at time t. In each passing iteration of the system the number of copies can be estimated as

$$N(m,g,t+1) = N(m,g,t) + C(m,g,t) - D(m,g,t) \tag{3}$$

where $C(m,g,t)$ and $D(m,g,t)$ represent the expected number of copies of meme m attached to genotype g that are created or destroyed at time t. The creation of a new copy can be accomplished by the combined effect of selection of a suitable agent with meme m and the application of the meme to the corresponding genotype. Let us express this as:

$$C(m,g,t) = \sum_{g'} \sigma(m,g',t) p(g' \xrightarrow{m} g) \tag{4}$$

where $\sigma(m,g',t)$ is the probability of selecting an agent carrying meme m and genotype g' at time t and $p(g' \xrightarrow{m} g)$ is the probability that the application of meme m on genotype g' results in genotype g. The first quantity can be computed as the probability that the binary tournament picks two agents with meme m and genotype g or only one agent with this structure but better fitness than its competitor:

$$\sigma(m,g,t) = p(m,g,t)^2 + \\ + 2\left\{p(m,g,t)\left[1 - p(m,g,t)\right]\right\} \cdot \frac{\sum_{m'}\sum_{g'<g} p(m',g',t)}{1 - p(m,g,t)} \tag{5}$$

where the last factor is the probability that the fitness of the competitor is worse than g provided it is not a $\langle m,g\rangle$ agent. This expression assumes that the global distribution of memes/genotypes across the whole population is the same as for local neighborhoods. Obviously, this holds for the panmictic case in which any two agents are neighbors so we can assume this case initially, and consider it a first approximation to more general situations.

As to the destruction of a copy of a particular pair meme/genotype, it can arise via the selection of such a pair and the subsequent application of local improvement (which will alter the genotype) or via replacement by an agent of higher fitness. The first case also requires that the other agent chosen in the tournament be a copy of the same pair, so that it is later substituted by the improved agent. Thus,

$$D(m,g,t) = \sum_{g' \neq g} p(m,g,t)^2 p(g \xrightarrow{m} g') + \tilde{\sigma}(m,g,t)\ . \tag{6}$$

The replacement probability $\tilde{\sigma}(m,g,t)$ can be expressed as:

$$\tilde{\sigma}(m,g,t) = 2\left\{p(m,g,t)\left[1-p(m,g,t)\right]\right\}\cdot \cdot \frac{\sum_{m'}\sum_{g'>g} p(m',g',t)}{1-p(m,g,t)} \tag{7}$$

Let us know consider the evolution of the system in the early-term and mid-term, before a particular meme starts to saturate the population. In this situation memes are widely spread across genotypes, so $p(m,g,t) \ll 1$, so we can take quadratic terms $p(m,g,t)^2$ as approximately 0 and terms $1-p(m,g,t)$ as approximately 1. We thus have:

$$\sigma(m,g,t) = 2p(m,g,t)\sum_{m'}\sum_{g'<g} p(m',g',t) \tag{8}$$

$$\tilde{\sigma}(m,g,t) = 2p(m,g,t)\sum_{m'}\sum_{g'>g} p(m',g',t) \tag{9}$$

Substituting back into Eqs. (4) and (6) we get:

$$C(m,g,t) = 2\sum_{g',m'} p(m,g',t)\sum_{g''<g'} p(m',g'',t)p(g' \xrightarrow{m} g) \tag{10}$$

$$D(m,g,t) = 2p(m,g,t)\sum_{m'}\sum_{g'>g} p(m',g',t) \tag{11}$$

Since $p(g' \xrightarrow{m} g) = 0$ for $g < g'$ or $m < g$, Eq. (10) reduces to

$$C(m,g,t) = 2\sum_{g''<g'\leqslant g}\sum_{m'} p(m,g',t)p(m',g'',t)p(g' \xrightarrow{m} g) \tag{12}$$

If $m \leqslant g$ then $p(g' \xrightarrow{m} g)$ is 1 if $g' = g$ and 0 otherwise. Subsequently, the difference $\Delta(m,g,t) = C(m,g,t) - D(m,g,t)$ is in this case

$$\begin{aligned}\Delta(m,g,t) = {} & 2p(m,g,t)\sum_{m'}\sum_{g''<g} p(m',g'',t) - \\ & -2p(m,g,t)\sum_{m'}\sum_{g''>g} p(m',g'',t) \\ = {} & 2p(m,g,t)\cdot \\ & \cdot\sum_{m'}\left[\sum_{g''<g} p(m',g'',t) - \sum_{g''>g} p(m',g'',t)\right]\end{aligned} \tag{13}$$

Focusing on the sign of the difference in the above expression, we essentially obtain that *inert* memes (i.e., memes that can no longer improve their hosts) can strive by hitchhiking, that is, if they attach to agents above the median of the population.

Let us on the other hand consider the case $m > g$. In this situation, $p(g' \xrightarrow{m} g)$ is 1 if $g' = f^{-1}(g,m)$ and 0 otherwise, where we denote by $f^{-1}(g,m)$ the genotype value such that $f(f^{-1}(g,m),m) = g$. Using g^{-m} as a shorthand for $f^{-1}(g,m)$,

$$\begin{aligned}\Delta(m,g,t) = {} & 2p(m,g^{-m},t)\sum_{m'}\sum_{g''<g^{-m}} p(m',g'',t) - \\ & -2p(m,g,t)\sum_{m'}\sum_{g'>g} p(m',g',t) \\ = {} & \sum_{m'}\left[2p(m,g^{-m},t)\sum_{g''<g^{-m}} p(m',g'',t) - \right. \\ & \left. -2p(m,g,t)\sum_{g'>g} p(m',g',t)\right].\end{aligned} \tag{14}$$

The sign of this expression depends on the balance between the goodness of genotypes in the basin of attraction of g and the badness of g itself (in both cases goodness/badness relative to the rest of the population). Notice that in general there is a reinforcement between these quantities in the sense that the better a genotype in the basin of attraction of g, the better we can expect g to be. This does not just mean that *active* memes proliferate more and more when they attach to good solutions as one would expect, but also that memes with high potential can find their way to the final stages of the evolution provided they have enough time to improve their hosts (recall that the goodness of solutions evolves with time as an effect of the application of the meme). This suggests that models with slower genetic convergence can have a beneficial effect on the propagation of good memes, allowing the latter enough time to express themselves in the population and overcome the hitchiking effect of bad memes. Next section provides a more quantitative analysis of this effect via numerical simulations.

IV. Numerical Simulations

The numerical experimentation is aimed to explore empirically the dynamics of meme propagation and how it is affected by factors such as the population size, the relative improvement potential of memes and the underlying spatial structure of the population. Regarding population sizes, we have considered values $\mu \in \{100, 256, 400, 625\}$. These values cover a broad range of population sizes and are also perfect squares, which is important in connection with one of the spatial structures considered, namely a square toroidal grid with von Neumann neighborhood: two locations (i,j) and (i',j') are connected if their Manhattan distance $|i-i'|+|j-j'| \leqslant r$, where r is the neighborhood radius. We have considered $r = 1$ which leads to the traditional North-South-East-West (plus the current location) neighborhood. The other spatial structure considered is the panmictic model in which all locations are connected. In either case, we have considered the function

$$f(g,m) = \begin{cases} g & \text{if } g \geqslant m \\ (g+m)/2 & \text{if } g < m \end{cases} \tag{15}$$

to represent the action of memes. Intuitively, this function provides smaller improvements for increasingly good genotypes much like often happens in practice. All experiments

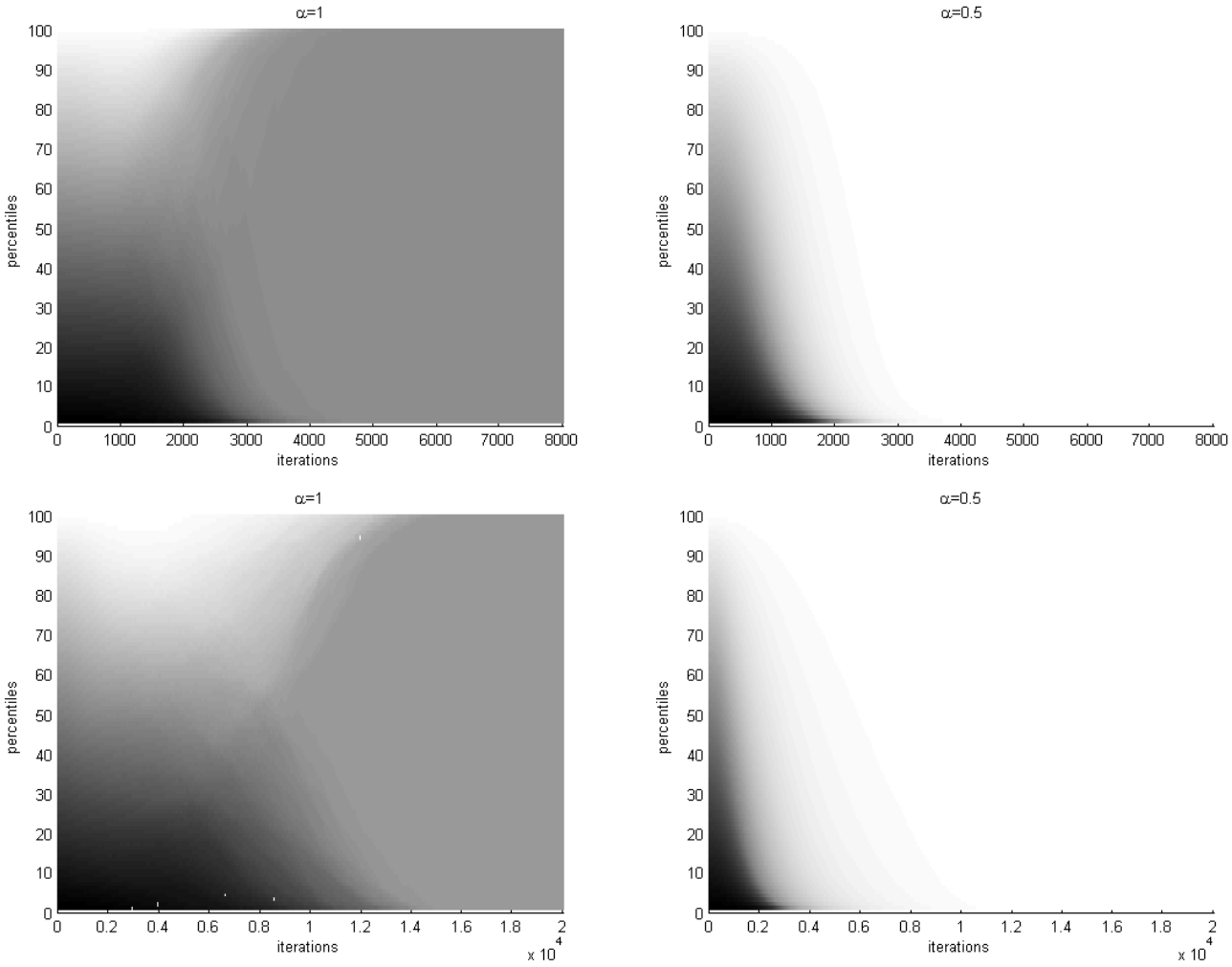


Fig. 1. Meme maps for simulations with $\mu = 625$. The upper row corresponds to panmictic connectivity and the lower row to von Neumann neighborhood. Similarly, the left column corresponds to genotypes initialized in $[0, 1]$ and the right one to initialization in $[0, 0.5]$ (memes are initialized in [0,1] in both cases). Lighter shades of gray indicate higher meme values. The evolution of the algorithm is depicted in each subfigure from left to right, each vertical slice representing the distribution of memes at a certain time t. Notice the different scale in the X-axis.

are averaged over 100 runs in order to obtain representative results. Each run is terminated upon convergence of memes, which for simplicity is determined when all memes are equal to 2 decimal positions.

Let us firstly analyze meme propagation as a function of the relative improvement potential of memes at the beginning of the run. For this purpose, we take $D = [0, 1]$ and consider a scenario in which genotypes and memes are randomly initialized in this range, and another scenario in which genotypes take initial values in $[0, 0.5]$ whereas memes are randomly sampled from $[0, 1]$. Figure 1 shows the distribution of memes at each time step (the lighter the shade of gray, the higher the meme value). Focusing firstly in the upper row (panmictic topology), notice the clearly different behavior depending on genotype initialization. When genes and memes are both initialized in $[0, 1]$ the algorithm does seldom converge to a high-potential meme. Actually, such memes temporarily proliferate in the initial stages of the algorithm but are later driven to extinction by memes hitchhiking on high quality genotypes to which they sticked by chance. The situation is quite different when genotypes are initially drawn from $[0, 0.5]$: in this case the algorithm does gradually converge to the upper part of the meme distribution, with low-potential memes quickly disappearing from the population. A more detailed perspective on this is provided by Figure 2 in which qualified runtime distributions (QRTDs) [8] are provided. These indicate the probability that a certain target (in this case, convergence to a meme in a desired percentile) is reached as a function of the number of iterations. Notice how the probabilities are below 10% for memes above the 95% percentile in the first scenario, whereas this probability is 100% in the second scenario. In the latter a spurious match between a very good genotype and a bad meme cannot happen since these very good solutions do not exist initially. Furthermore, high-potential memes initially

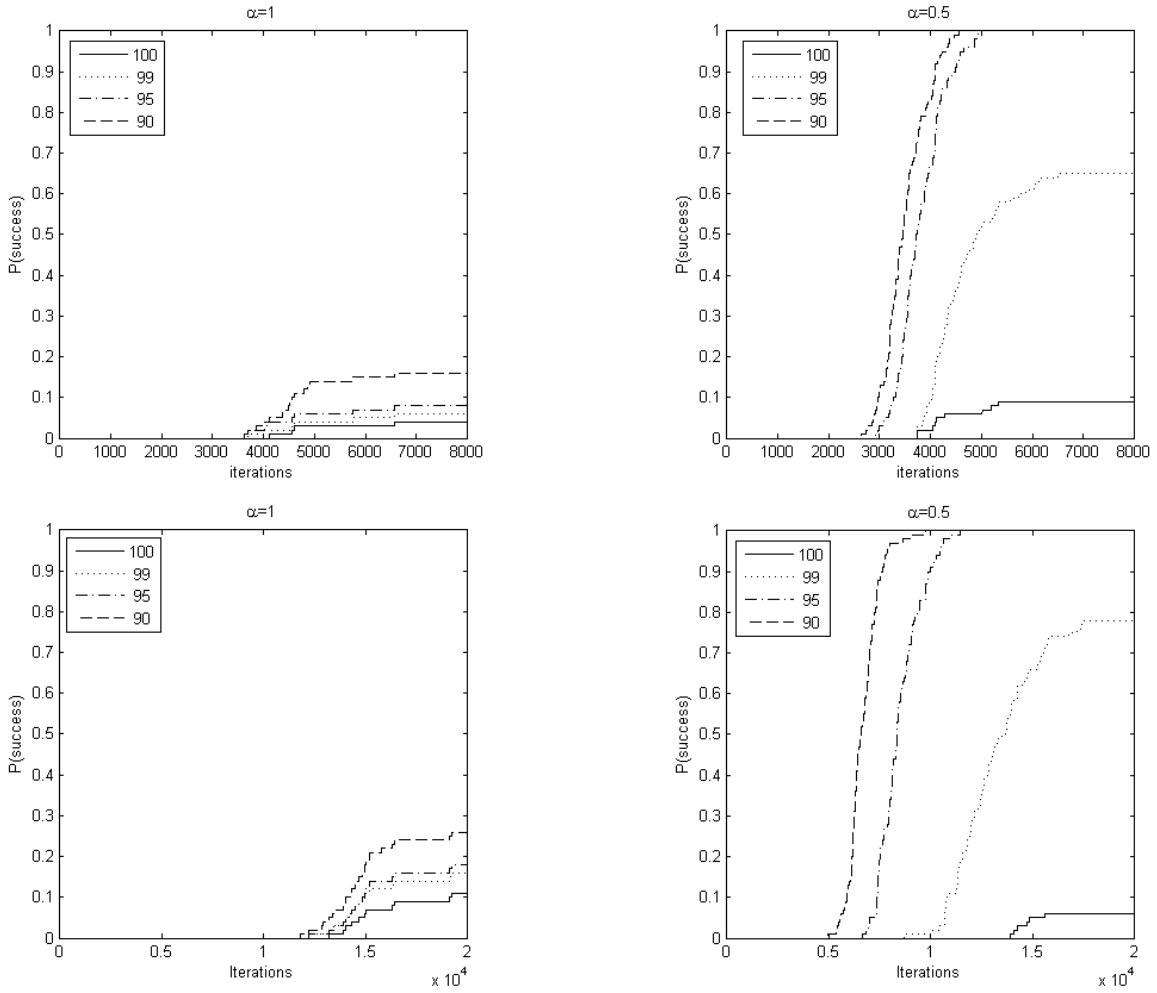


Fig. 2. Qualified runtime distributions for simulations with $\mu = 625$. The upper row corresponds to panmictic connectivity and the lower row to von Neumann neighborhood. Similarly, the left column corresponds to genotypes initialized in $[0, 1]$ and the right one to initialization in $[0, 0.5]$ (memes are initialized in [0,1] in both cases). The curves indicate the probability that the population converges to a meme in the i-th initial percentile of the population as a function of the number of iterations. Notice the different scale in the X-axis.

attaching to bad genotypes can highly improve the quality of the latter in the initial steps, thus increasing their chances of survival.

Let us know turn our attention to the effect of the spatial structure. The bottom row of Figure 1 shows the distribution of memes for the case of von Neumann neighborhood. Notice how a similar pattern as in the panmictic case is observed with respect to genotype initialization. A more detailed inspection indicates several differences though. Firstly, notice how the convergence is slower in this case (e.g., the scale in the X-axis is larger). This is a well-known effect of the use of spatial structure and is commonly exploited in the context of evolutionary algorithms for promoting diversity and thus decreasing the chances of getting stuck in local optima [4], [23]. In the case of MMAs this has an additional advantage, namely the fact that a slower convergence increases the lifespan of individual memes, thus giving them more chances to improve their hosts if they have the potential to do so. Hence, the algorithm is more robust and can better cope with hitchhiking inert memes. This can be seen in the meme map in the bottom row of Figure 1 by a larger prevalence of lighter-gray areas, and more clearly in the QRTDs (bottom row of Figure 2), e.g., the 95% percentile is reached with nearly 20% probability in the case of $[0, 1]$-initialization (cf. below 10% in the panmictic case), and the 99% percentile is reached with nearly 80% probability for $[0, 0.5]$-initialization (cf. about 65% in the panmictic case). A signrank test [24] indicates that the difference in the final percentile reached is statistically significant in both cases ($\alpha = .05$).

Finally, we consider the takeover time, namely the time required for a meme (not necessarily the best one as shown previously) to dominate complete the population. Figure 3 shows the growth curves, depicting the percentage of the meme pool occupied by the most repeated meme (notice that the most repeated meme need not be the same throughout a run; we simply count the number of copies of the most repeated one

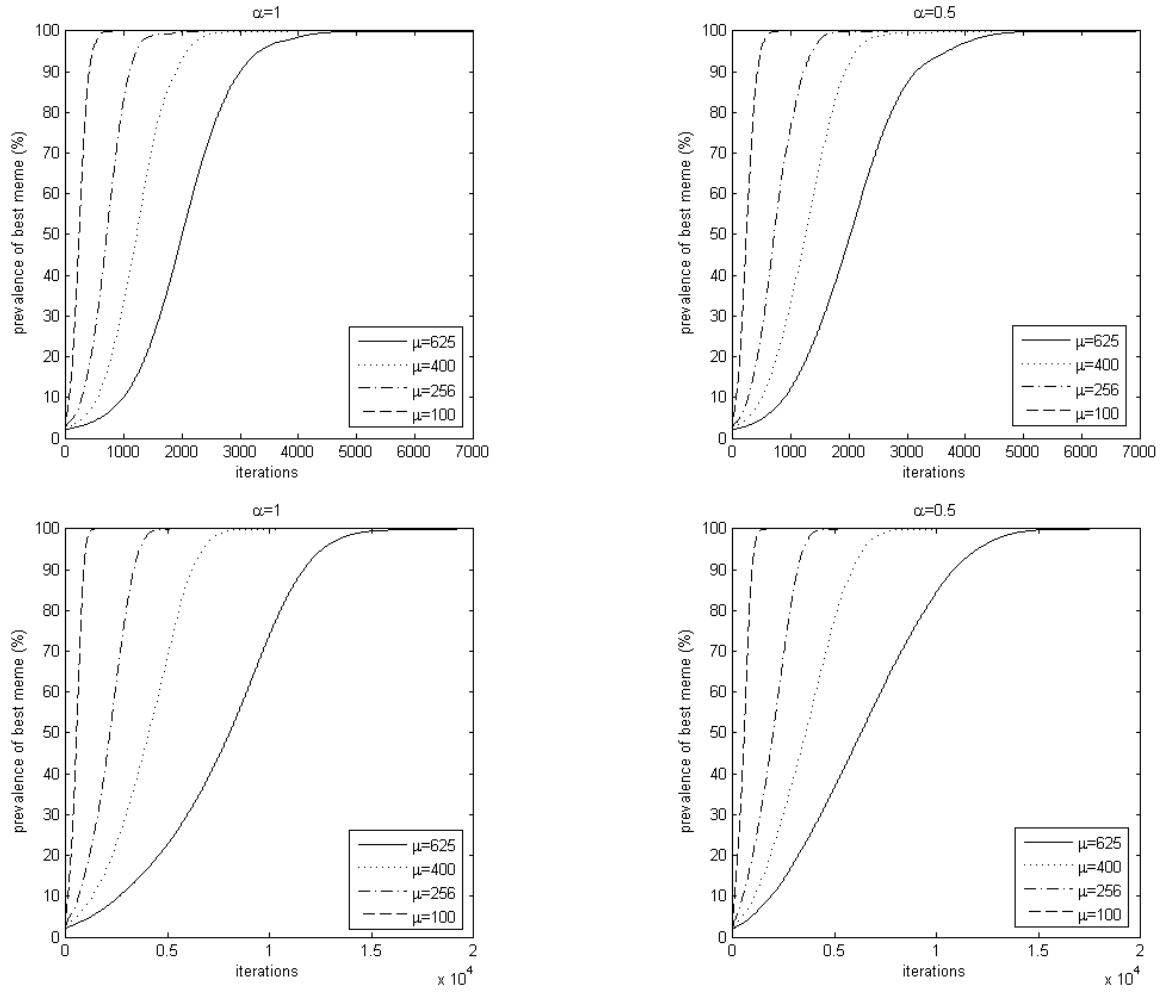


Fig. 3. Growth curves for different population sizes. The upper row corresponds to panmictic connectivity and the lower row to von Neumann neighborhood. Similarly, the left column corresponds to genotypes initialized in $[0, 1]$ and the right one to initialization in $[0, 0.5]$ (memes are initialized in [0,1] in both cases). Notice the different scale in the X-axis.

TABLE I
FITTING GROWTH CURVES TO A LOGISTIC FUNCTION. FOR EACH ALGORITHM CONFIGURATION THE SCALE PARAMETER α AND THE MEAN SQUARED ERROR IS SHOWN.

		population size							
		$\mu = 100$		$\mu = 256$		$\mu = 400$		$\mu = 625$	
topology	G	α	mse	α	mse	α	mse	α	mse
panmictic	0.5	81.628878	0.000042	210.791430	0.000069	324.649846	0.000064	521.933265	0.000017
	1.0	77.708284	0.000008	187.904342	0.000030	297.902782	0.000027	462.866673	0.000010
von Neumann	0.5	186.815036	0.000366	578.914610	0.000554	1046.429102	0.000361	1974.995403	0.000318
	1.0	168.567695	0.000322	585.804249	0.000532	1057.519394	0.000537	1945.625376	0.000705

at each time step). These curves exhibit the typical shape of the well-known logistic model $f(t) = 1/(1 + Ke^{-t/\alpha})$. Indeed, such a model was proposed early in the literature by Sarma and De Jong [20] in the context of spatially structured evolutionary algorithms. While by no means the unique alternative – e.g., see [5]– it serves as a good starting approximation to quantify the growth of the dominant meme. Qualitatively, we observe as expected the well-known pattern of slower convergence for increasing population sizes and for the von Neumann topology [6] as opposed to the panmictic population. From a quantitative point of view, we have fitted the growth data to a logistic curve to identify the scale factor α that renders the number of iterations dimensionless. The resulting data is shown in Table I. As it can be seen, the fit is quite good, yielding very low mean squared errors. The scale parameters are quite similar for variants with the same topology, and are about 2-5 times larger for the von Neumann topology than for the panmictic case, in correspondence with the relative takeover time which

can be seen in Figure 3. With respect to the population size, the increase in the scale parameter admits a linear interpolation $\alpha = a + b\mu$ yielding values of $b = 0.84$ and $b = 0.74$ for the panmictic case and $b = 3.43$ and $b = 3.40$ for grid topology with von Neumann connectivity.

V. Conclusions

We have presented some initial steps in the line of analyzing meme propagation in MMAs. Using an idealized model of genotypes and memes we have shown that the selection intensity plays a very important role in allowing high-potential memes to proliferate. In a panmictic model, good memes will dominate the final population when the starting solutions have a substantial improvement margin on average. When this margin is smaller, average memes can hitchhike their way to the final stages of the evolution and make other comparatively better memes become extinct. In the presence of a spatial structure inducing longer takeover times (in our case a toroidal square grid with von Neumann topology), this hitchhiking effect is somewhat mitigated, allowing good memes to express themselves and increasing their chances for making it to the final population. An interesting line of future research focuses on the consideration of other topologies and study their effect on meme propagation. Work is in progress in this area. Looking beyond, another topic for further research is the extension of this analysis to coevolutionary memetic algorithms [22] in which memes are detached from genotypes and co-evolve alongside the latter in a separate population.

References

[1] E. Alba and G. Luque. Growth curves and takeover time in evolutionary algorithms. In K. Deb, editor, *Genetic and Evolutionary Computation Conference – GECCO 2004*, volume 3102 of *Lecture Notes in Computer Science*, pages 864–876, Seattle, WA, 2004. Springer-Verlag.

[2] P. Cowling, G. Kendall, and E. Soubeiga. A hyperheuristic approach to schedule a sales submit. In E. Burke and W. Erben, editors, *PATAT 2000*, volume 2079 of *Lecture Notes in Computer Science*, pages 176–190, Berlin Heidelberg, 2008. Springer-Verlag.

[3] R. Dawkins. *The Selfish Gene*. Clarendon Press, Oxford, 1976.

[4] B. Dorronsoro and E. Alba. *Cellular Genetic Algorithms*, volume 42 of *Operations Research/Computer Science Interfaces*. Springer-Verlag, 2008.

[5] M. Giacobini, E. Alba, and M. Tomassini. Selection intensity in asynchronous cellular evolutionary algorithms. In Erick Cantú-Paz et al., editors, *Genetic and Evolutionary Computation Conference – GECCO 2003*, volume 2723 of *Lecture Notes in Computer Science*, pages 955–966, Chicago, IL, 2003. Springer-Verlag.

[6] M. Giacobini, M. Tomassini, A. Tettamanzi, and E. Alba. Selection intensity in cellular evolutionary algorithms for regular lattices. *IEEE Trans. Evolutionary Computation*, 9(5):489–505, 2005.

[7] W.E. Hart, N. Krasnogor, and J.E. Smith. *Recent Advances in Memetic Algorithms*, volume 166 of *Studies in Fuzziness and Soft Computing*, chapter Memetic Evolutionary Algorithms, pages 3–27. Springer-Verlag, Berlin Heidelberg, 2005.

[8] H. Hoos and T. Sttzle. *Stochastic Local Search: Foundations & Applications*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[9] N. Krasnogor. Self generating metaheuristics in bioinformatics: The proteins structure comparison case. *Genetic Programming and Evolvable Machines*, 5(2):181–201, June 2004.

[10] N. Krasnogor, B.P. Blackburne, E.K. Burke, and J.D. Hirst. Multimeme algorithms for protein structure prediction. In J.J. Merelo et al., editors, *Parallel Problem Solving From Nature VII*, volume 2439 of *Lecture Notes in Computer Science*, pages 769–778. Springer-Verlag, Berlin, 2002.

[11] N. Krasnogor and S.M. Gustafson. A study on the use of "self-generation" in memetic algorithms. *Natural Computing*, 3(1):53–76, 2004.

[12] N. Krasnogor and J.E. Smith. A tutorial for competent memetic algorithms: model, taxonomy, and design issues. *IEEE Transactions on Evolutionary Computation*, 9(5):474–488, 2005.

[13] P. Moscato. On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms. Technical Report Caltech Concurrent Computation Program, Report. 826, California Institute of Technology, Pasadena, California, USA, 1989.

[14] P. Moscato and C. Cotta. A gentle introduction to memetic algorithms. In Fred Glover and Gary Kochenberger, editors, *Handbook of Metaheuristics*, volume 57 of *International Series in Operations Research & Management Science*, pages 105–144. Kluwer Academic Press, New York, USA, 2003.

[15] F. Neri and C. Cotta. Memetic algorithms and memetic computing optimization: A literature review. *Swarm and Evolutionary Computation*, 2:1–14, 2012.

[16] F. Neri, C. Cotta, and P. Moscato. *Handbook of Memetic Algorithms*, volume 379 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin Heidelberg, 2012.

[17] M.G. Norman and P. Moscato. A competitive and cooperative approach to complex combinatorial search. In *Proceedings of the 20th Informatics and Operations Research Meeting*, pages 3.15–3.29, Buenos Aires, 1989.

[18] Y.-S. Ong and A.J. Keane. Meta-lamarckian learning in memetic algorithms. *IEEE Transactions on Evolutionary Computation*, 8(2):99–110, 2004.

[19] G. Rudolph and J. Sprave. A cellular genetic algorithm with self-adjusting acceptance threshold. In *1st IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications*, pages 365–372, London, UK, 1995.

[20] J. Sarma and K. De Jong. An analysis of local selection algorithms in a spatially structured evolutionary algorithm. In T. In Bäck, editor, *7th International Conference on Genetic Algorithms*, pages 181–186. Morgan Kaufmann, 1997.

[21] B. Schönfisch and A. de Roos. Synchronous and asynchronous updating in cellular automata. *BioSystems*, 51:123–143, 1999.

[22] J.E. Smith. Coevolving memetic algorithms: A review and progress report. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(1):6–17, 2007.

[23] M. Tomassini. *Spatially Structured Evolutionary Algorithms*. Natural Computing Series. Springer-Verlag, 2005.

[24] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1:80–83, 1945.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1021–1027

Fair and truthful multiagent resource allocation for conference moderation

Adam Połomski
Warsaw University of Technology,
ul. Nowowiejska 15/19, Warsaw, Poland
Email: A.K.Polomski@elka.pw.edu.pl

***Abstract*—Multiuser voice conferencing platforms are more and more popular. Internet bandwidth is becoming very accessible, what makes voice over IP used on an everyday basis. Being able to communicate with multiple people at the same time can be beneficial, but on the other hand increases the need of coordination mechanisms. Determining a moderation scheme which is fair and efficient is not a trivial problem to solve. We define conference moderation as a multiagent resource allocation problem and introduce a process based on Vickrey auctions to solve it. A concept of co-owned communication channel is what stands as a basis of our definition of fairness.**

***Index Terms*—Multiagent systems, Auctions, Social welfare, Moderation, Simulation**

I. Introduction

EXCHANGE of information is a crucial part of our existence. In some situations we have to elaborate with larger groups of people. It might be very valuable, as every member of such group brings some extra knowledge, potentially useful for the whole society. Yet, it comes at a certain price. It requires an additional effort to manage the flow of information as the number of participants grows. In a most general definition, a conference is ”a meeting for consultation, exchange of information, or discussion”. Yet, the conferences may vary a lot in its specifics. A discussion among a group of friends lives by a different rules than a company meeting. An environment in which a large group of random people struggles to exchange some information has the biggest potential of becoming extremely chaotic, thus bringing the flow of information to a minimal level. Open societies with low entry barriers often face problems of disruptive behavior such as flooding or spamming. How do we deal with that? How do we coordinate the flow of information in a way which is efficient and fair? That is a role for moderation mechanism.

Decentralized moderation schemes for large scale social platforms like Usenet or Slashdot is being discussed in a number of papers [1], [2]. Multiagent resource allocation and a concept of bargaining or trading agents is also a common topic of research, with distribution of network bandwidth being one of the possible applications [3].

This is an extended version of the paper in which we first introduced the presented allocation model [4]. We mostly focused on the verification of the model. However, we also provided a new, refreshed view of the problem definition and social welfare, which may be found in the next two sections. We gave some more detailed insight into features a fair and effective social welfare metric should possess. Section V contains a description of Vickrey auction based allocation method. We have also introduced a resale procedure to deal with a problem of choosing a proper allocation time. Next we move on to an overview of verification methodology in Section VI followed by simulation results in Section VII. A JADE based implementation of conferencing agents is presented.

II. Background

The environment with which we are dealing is very specific. It is a huge online audio conference platform connecting people from all parts of the world. It is used by hundreds of thousand users every day. It is very likely that conference participants will not know each other. The discussion topics may vary and cannot be anyhow limited or managed. The only thing that can be assumed or enforced is that all participants in a single discussion share the same language. Also the number of concurrent ongoing conferences may be as high as tens of thousands. The model of a single conference is very simple though. All participants connect to a single device—media server, which is responsible for handling the voice stream. Strictly speaking it broadcasts the voice stream transmitted by a single participant to all others. It also has a steering protocol which allows controlling it to some extent. What is most important, it allows specifying which participants are allowed to transmit voice signal in the given moment of time and which ones are only allowed to receive it.

A. Moderation

For this sort of audio conversation platform to function successfully, a moderation mechanism is required. User experience would suffer otherwise. It is hard to strictly define what would be seen as ”good” or ”bad” conversation by the participants. There is a basic rule that definitely needs to be fulfilled in order to achieve a fair discussion.

No participant will be able dominate the discussion.

Moderation is a mean to fulfill those requirements. If the ability to speak is distributed properly among participants over time, it should be possible to maximize welfare of the whole group. There are a few standard, commonly known moderation mechanisms, which can be observed:

- No moderation—For example a group of friends talking at the cafeteria will not require any moderation to get the most of their discussion. It is important that the group

978-83-60810-53-8/$25.00

is relatively small. The fact that participants know each other well and have no point in dominating the discussion also helps.
- Discussion rules—For example a lecture at the university. Both, the lecturer and the students are aware of discussion rules up front and will respect them.
- Human moderator—A designated person is responsible for leading the discussion by granting/revoking voice to participants. It is also very important that moderator understands and follows the discussion, as it is crucial to pick the right people to speak in a given time slot.
- Queuing—All participants queue up and are the voice is granted in a "round robin" scheme.

All of the above moderation models except queuing cannot be applied in this specific environment. That is mostly because of size of the system. Queuing is the only model that does not require manual management and puts no trust that users will apply to some rules without forcing them into it. A huge drawback of queuing in this case is the maximum pessimistic waiting time. It is possible that a participant who needs the voice most will be forced to wait until everyone else uses the granted time slot. That is an area, where introducing a multi agent solution could bring better results.

B. Multiagent resource allocation

Multiagent resource allocation is a process of distributing a number of items (resources) among a number of agents [5]. This brief definition, however, does not fully describe the problem.

Resources might differ in characteristics. The whole range of resource types is substantial and each might require different allocation technique. For instance, we can distinguish divisible goods (like network bandwidth) and indivisible. Also it may, or may not be allowed to share an indivisible resource among a number of agents.

Agents may have preferences over different allocation outcomes. Not only may they have preferences over resource bundles they receive, but also over bundles received by others. Preferences can be represented in a various ways, like utility function or binary relation on alternatives. Moreover, agents may or may not be truthful while reporting their preferences.

Allocation can be performed with the use of various allocation procedures, which can be either centralized or distributed. Typical examples of centralized procedures are auctions or voting mechanisms, with a central entity empowered to decide on the final allocation. In distributed solutions agents try to come to a common agreement through a sequence of local negotiation steps. In both cases, the objective is to find an allocation which is feasible or optimal. What stands behind the concept of optimal depends on the specific multiagent resource allocation scenario.

III. Problem definition

We formulate conference moderation task as a resource allocation problem. The following definition is general and may find application whenever multiple agents compete over a non divisible resource across multiple time periods.

Let $N = \{a_1, a_2 \dots a_n\}$ be a finite set of participating agents. Each agent takes part in $T \in \mathbb{N}^+$ consecutive resource allocation runs, each for a separate time period. A resource can only be allocated to one agent at a time, therefore the set of feasible allocations $\Delta = N$. Let δ_t denote the allocation outcome for t allocation run $\forall t \leq T$. There is also a special null allocation δ^- which represents a situation when no agent holds a resource. Each of the participating agents has its preference regarding every allocation represented by the utility function $u_{i,t} : \Delta \rightarrow R, \forall t \leq T, a_i \in N$. Let vector $x = (\delta_1, \delta_2 \dots \delta_T)$ hold all allocations across all time periods. We call $S(x)$ a social welfare function. $S : \Delta^T \rightarrow R$. We will discuss it in details in section IV. For now it is only important that it determines an overall happiness of the whole society N for a given allocation vector x. Objective is to find x^*, which will $\max S(x)$.

Utility functions are not known upfront. While deciding upon an allocation δ_t, our knowledge consists of:
- current and past utility functions for all participating agents
- allocations that have been performed up to this point

Note that the shapes of current utility functions u_{it} are very likely to depend on the whole history of allocations up to this point in time $(\delta_1, \delta_2, \delta_3 \dots \delta_{t-1})$ and what those allocations brought. In other words, every allocation decision can affect the way agents shape their utilities in the future. It might also have an impact on the total number of consecutive allocations T.

IV. Social welfare function

Social welfare function S ranks every feasible allocation vector x. This ranking represents a welfare of the whole society N if an allocation x took place. Since S determines whether we choose one allocation vector over the other, it also defines what we consider as fair. It is not a straightforward task to rule what is fair in these specific conditions. Moreover, according to Arrows impossibility theorem [6] there exists no reasonably consistent social welfare metric. It needs to be carefully chosen to fit the specific scenario.

For a given application Nash social welfare metric has advantages over Utilitarian or Egalitarian viewpoints [7]. It leverages between fairness by equalizing utility distribution among agents and higher overall utility. However, we cannot define the social welfare metric as in Equation 1.

$$S_N(x) = \prod_{a_i \in N, t \leq T, t \in \mathbb{N}^+} u_{it}(\delta_t) \tag{1}$$

Consider the scenario presented in Table I and two allocation vectors $x_1 = (a_1, a_1, a_2), x_2 = (a_2, a_2, a_1)$. Both vectors are equally valuable in terms of fairness, as each of the two agents gets his share of resources and the division is as equal as possible. After applying Nash social welfare metric we get $S_N(x_1) = 250 > S_N(x_2) = 200$, thus it clearly favors x_1 over x_2. From the utilitarian perspective we have

TABLE I
SAMPLE WELFARE VALUES

t	δ_t	$u_{1,t}$	$u_{2,t}$
1	a_1	5	1
1	a_2	1	10
2	a_1	5	1
2	a_2	1	2
3	a_1	10	1
3	a_2	1	10

$S_U(x_1) = 23 < S_N(x_2) = 25$, what means that overall utility is lower in case of allocation x_1.

An ideal social welfare metric for this environment should in the first place ensure that all rules of fair discussion proposed in Section II-A are fulfilled. Once this is secured, higher overall profit should be promoted. We propose the conference metric given by equation 2, which mixes the concepts of utilitarian social welfare and Nash product together.

$$S_M(x) = \prod_{a_i \in N} \sum_{t \leq T, t \in \mathbb{N}^+} u_{it}(x_t) + \epsilon \tag{2}$$

A. Unanymity

Unanimity principle is the most important concept of welfare economics [8]. It says that the chosen utility vector should not be Pareto inferior to any other feasible utility vector. Metric defined by Equation 2 fulfills the rule of unanimity, as it prefers allocations which bring strongly Pareto optimal utility vectors. Let x_1 be a feasible allocation vector preferred by the metric. Let x_2 be a different feasible allocation dominating x_1 in terms of Pareto domination. Therefore we have:

- $u_{i,t}(x_{1,t}) \leq u_{i,t}(x_{2,t}), \forall a_i \in N, \forall t < T, t \in \mathbb{N}^+$
- $\exists a_i \in N, \exists t < T, t \in \mathbb{N}^+, u_{i,t}(x_{1,t}) < u_{i,t}(x_{2,t})$

It is easy to see, that $S_M(x_1) < S_M(x_2)$. This means that x_2 would be a preferred allocation vector by the metric.

Note that it does not take place if we remove ϵ from the equation. For instance, if one of the agents had zero utility for every possible allocation, all allocation vectors would be seen as equal.

B. Anonymity

Anonymity (symmetry across the agents) is another important social welfare metric feature. It is especially significant while considering fairness, as it indicates if any member of the society is in a privileged position. S_M is anonymous to some extent. For any choice of agents a_1 and a_2, switching their utility functions, so that $u'_{1t} = u_{2t}, \forall t \leq T$ and $u'_{2t} = u_{1t}, \forall t \leq T$ will not change the rating of any allocation vector. This only holds true if we perform so for all values of t. All allocations $\delta_j, \forall j < t$ and utility functions $u_{n,t}, \forall j < t$ act as an allocation history, which has a significant impact while deciding on δ_j. This historical data makes some agents more privileged then others. That is all in line with our understanding of fairness.

V. FAIR ALLOCATION MODEL

The proposed model is designed based on two assumptions:

- The whole resource is cofounded by each of the participating agents, therefore each member of the society owns an equal share of rights.
- Every agent may only grant some utility to owning the resource himself—$u_{it}(a_j) = 0, \forall i \neq j$. A more general allocation model with no such restriction has been proposed in [4], but it is out of scope of this paper.

The resource is indivisible and can only be utilized if fully owned. No one can use just a part of resource, yet agents can negotiate over the price and purchase or sell it to each other. For this purpose each agent a_i is associated with r_{it}, which might be interpreted as agent's wallet for time period (allocation number) t. Initially $r_{i1} = R, \forall a_i \in N$, where R is a positive constant value to initialize all the wallets. The conference is divided into T shorter periods, at the beginning of each agents may express their desire to obtain full rights to transmission channel. Resource allocation is then performed with the use of Vickrey auction mechanism [9]. The allocation pattern for period t is following:

1) Each participant a_i issues a bid with the valuation v_{it}. The bid cannot be higher than the actual wealth of agent at that time, therefore $v_{it} = \min(u_{it}(a_i), r_{it})$
2) The winning agent a_k and the price to pay p_t is determined with the use of Vickrey auction.
3) Price to pay is deducted from the winner's wallet $r_{k,t+1} = r_{k,t} - p_t$.
4) All agents which sell their resource rights to a_{kt} are rewarded $r_{i,t+1} = r_{i,t} + \frac{p_t}{|N|-1}$.
5) The whole resource is allocated to $\delta_t = a_k$ for time period t

According to the introduced model, resource allocation time is constant and defined upfront. It is a serious limitation given that demands might vary in terms of allocation length. If allocation time was kept constant and set too long, we would waste the resource due to excess allocation period. On the other hand, setting this time too short would cause allocations which only partially meet the demands. Moreover, there is no guarantee that agents assign any utility at all to allocations which only partially meet their demands.

We deal with this problem by setting an allocation time long and introducing a resale procedure. Whenever an agent owns a resource which he no longer requires while there is still some time left before exclusive rights period ends, he can request for an earlier reallocation. If a demand for the resource exists, he might get a fraction of the price he paid back.

Let $\delta_t = a_i$ and $\delta_{t+1} = a_j$. Agent a_i decided to request a resale after utilizing the resource for time l out of full allocation time L. When a_j wins:

- $r_{i,t+2} = r_{i,t+1} + \frac{l}{L} \min(p_t, p_{t+1}) + \frac{p_t - \frac{l}{L} \min(p_t, p_{t+1})}{|N|-1}$
- $r_{k,t+2} = r_{k,t+1} + \frac{p_t - \frac{l}{L} \min(p_t, p_{t+1})}{|N|-1}, \forall k \neq i, j$

It is important that an agent will never gain from the resale procedure. It is only possible to get a fraction of an allocation

price back. Otherwise, it would motivate members of the society to purchase a resource with the hope of future resale at a higher price.

By leveraging Vickrey auction, we gain all characteristics of this auction mechanism. Performing allocation is quick and does not require a lot of overhead network traffic. This is a very important feature, as it is crucial to finish negotiations before the beginning of conference period affected by this allocation. Agents are also highly encouraged to bid their true valuations, as it is in line with the dominant auctioning strategy. It is important to mention here that in the dominant strategy equilibrium is a weak equilibrium for VCG process in case of asymmetrical bid ranges [10]. For the richest agent it is equally reasonable to bid anything from the full value of his wallet to a wallet value of the second richest agent. In this case however, it has no effect on the choice of winner or on the price.

Unfortunately Vickrey auction has a couple of drawbacks. It is vulnerable to bidder collusion agreements. A group of agents with the highest valuations may settle not to vote their true valuations in order to lower the resulting price. The model is also exposed to lying auctioneers. Agents may bid shill votes in order to inflate the price and increase the income.

VI. Methodology

In order to verify the proposed model and how it operates as conference moderation mean, we performed a whole range of simulated discussions. We implemented two types of agents playing a role in the conference—coordinator and participant. Implementation was done with the use of JADE platform [11]. Coordinator is responsible for keeping track and deciding whenever the society should transition to a next allocation phase. It is also up to the coordinator to decide upon the resource allocation based on all data (utilities) collected from participants. Participant's main responsibility is to declare its current utility of holding the resource whenever coordinator announces the bidding phase.

A. Protocol

Naturally, interaction among agents requires a specified communication language. FIPA standard defines a huge set of protocols, which come implemented out of the box with JADE. We have chosen some of as a language for our agents.

"Subscribe" protocol—defined by the FIPA standard. Should be used, whenever "Initiator" agent wishes to monitor the state of an object owned by the "Receiver". For the purpose of this simulation, subscribe protocol is used by the participants to monitor state changes of the communication channel. This information is broadcasted by the coordinator after each performed allocation.

"Allocation" protocol—FIPA standard does not provide a Vickrey auction protocol. We designed the protocol of our own as shown in Figure 1. Every auction is initiated by the coordinator agent, who collects bids from all participants and performs the allocation according to the implemented model.

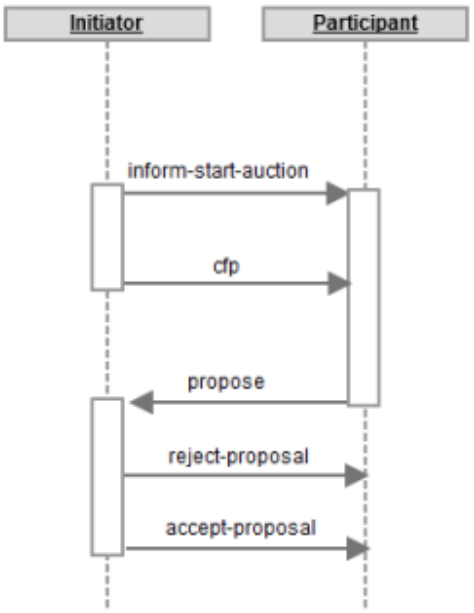


Fig. 1. Allocation protocol

"Request" protocol—defined by the FIPE standard. Used whenever an "Initiator" agent asks the "Receiver" to perform some action. In the discussion simulation, this protocol is used by a participant agent to request for an earlier reallocation procedure.

B. Behaviours

The whole process begins with a subscription phase when coordinator awaits for all participants to submit their conference subscription requests. After a certain time limit the discussion moves on to the main phase, which consists of two alternate behaviors:

- Auction—Collecting bids (valuations) from all conference participants. Making an allocation decision based on collected bids. Informing participants about the auction result.
- Allocation—Granting the resource to the new owner. Collecting and distributing the payment. Informing participants about the resource ownership change.

The whole auction process along with a decision upon the next allocation change is performed in advance (auction speedup time) to the actual change, in the background of the previous allocation time. This procedure is designed to eliminate allocation time fluctuations caused by the auction procedure, as it might involve passing a substantial number of messages over the network or performing timely calculations.

C. Demands

The main task of participant agent is to compete over a resource (communication channel) according to his preferences. Determining current utility is performed based on the agents set of demands. A single demand object is shown in the Figure 2. It is characterized by the following features

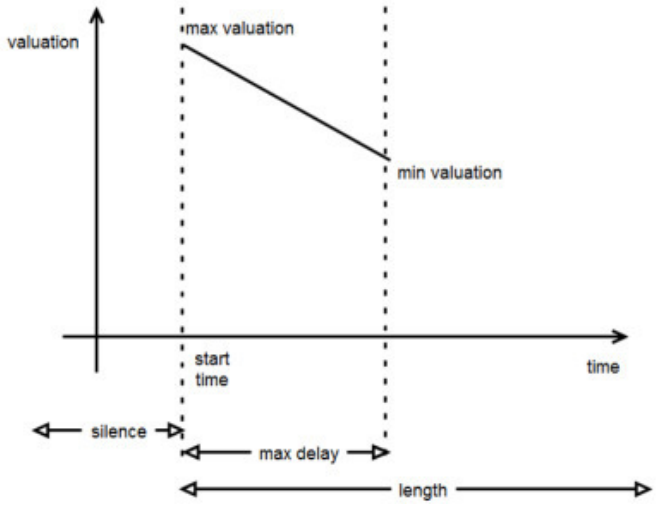


Fig. 2. Demand object

- Start time—the time since when the demand is active and has influence on agent's utility
- Maximal delay—maximal time in which the demand is still active
- Length—allocation time required to satisfy this demand
- Maximal valuation—valuation for an allocation at the demand start time
- Minimal valuation—valuation for an allocation with a maximal delay
- Silence time—time interval from the previous demand

In order to determine the current resource valuation, an agent scans through the whole set of active demands and picks the one with highest active valuation. If a demand is not satisfied right when it arises (start time), valuation deteriorates linearly up till maximal delay.

D. Dynamic demands

Aiming to increase similarity to real live conversational environment, we implemented semi intelligent agents capable of dynamically shaping their utility functions based on discussion history. Human conversation is a process driven by certain rules [12]. Among others, these rules describe how people choose and change the discussion topic:

- RULE 3: In introducing a new topic of conversation, the topic should be chosen so that both speakers have some knowledge and interest in its discussion.
- RULE 6: The topic of conversation may drift to a subject where the conversational participants share a great amount of knowledge.
- RULE 13: Each participant in the conversation has the conversational goal of saying things that are important to the other participant.

Having that in mind, we have developed an agent capable of reacting (adjusting his utility function) to the discussion flow. Such agent is supplied with a set of topic preferences and discussion memory to record the occurrence frequencies of all topics. Given all that as an input data, whenever an agent takes active part in a discussion, it will first choose a theme possibly interesting for all parties and shape it's utility function accordingly. Every topic from the list might be chosen with the probability directly proportional to agent's preferences and the frequency of occurance in the discussion so far.

E. Allocation models

We performed simulations for four different allocation models acting as a moderation procedure. Apart from the "Fair allocation model" proposed in Section V, we use three other allocation schemes for the purpose comparison:

- Queuing—Participants reporting need for the resource (positive valuation) are put into FIFO queue. Resource is always allocated to the first agent in the queue. There is no risk of dominating the discussion, yet it does not give any preference to allocations which increase the overall social welfare.
- Choosing maximal valuation—Resource is allocated to an agent reporting highest valuation at the given time. This model does not encourage agents to bid their true valuations nor does it put any preference to fair allocations. The dominating strategy is to bid just a tiny bit above the highest bidding participant.
- Maximizing social welfare metric—Chooses the allocation which maximizes social welfare metric defined by Equation 2. Prefers allocations which are both fair and increase the overall social welfare. However, participants might try to increase their profits or dominate the discussion by faking their bids. Under the assumption that allocation decision does not have any impact on the future shape of agents' utility functions, this model guarantees to maximize the social welfare.

VII. Evaluations

We came up with three discussion participant profiles which vary in characteristics of their demands set. All demand attributes were generated from uniform distribution on a given range.

- Regular:
  - Silence time: 3 to 15 seconds
  - Maximal delay: 3 to 7 seconds
  - Length: 3 to 10 seconds
  - Maximal valuation: 2 to 10
  - Minimal valuation: 0 to half of maximal valuation
- Aggressive:
  - Silence time: 3 to 3,5 seconds
  - Maximal delay: 3 to 7 seconds
  - Length: 3 to 10 seconds
  - Maximal valuation: 10
  - Minimal valuation: 9 to 10
- Dynamic: (as described in Section VI-D):
  - Silence time: 3 to 15 seconds
  - Maximal delay: 3 to 7 seconds

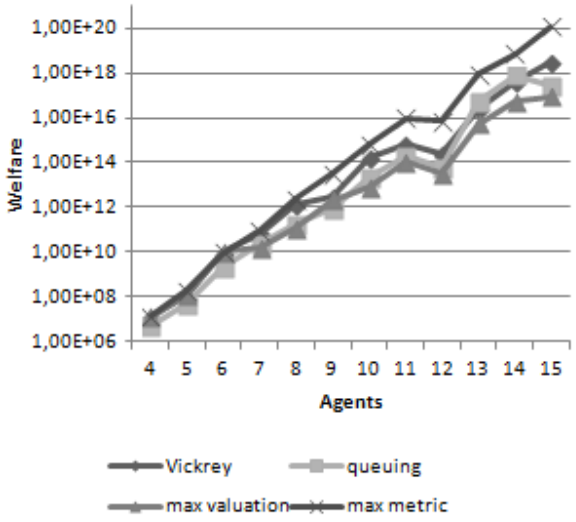


Fig. 3. Regular demands

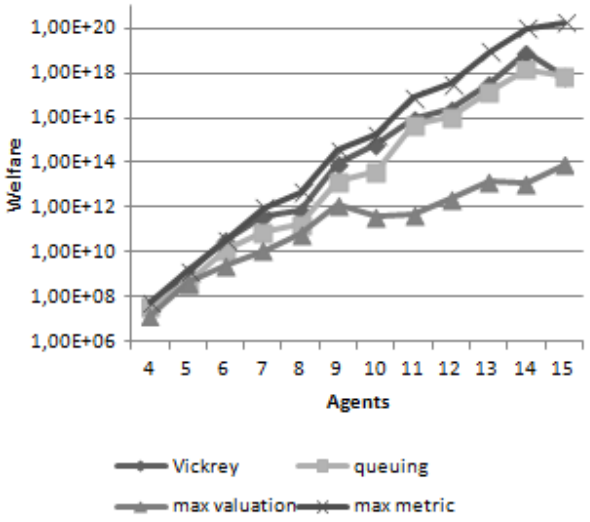


Fig. 4. Regular demands with 10 percent aggressive

- Length: 3 to 10 seconds
- Maximal valuation: 10 * topic preference
- Minimal valuation: 0 to half of maximal valuation
- Topics of interest: 5 randomly chosen topics out of total 10
- Topic preference (for each topic of interest): 0 to 1

We performed model evaluations by simulating a number of discussions with the following parameters:

- Discussion time—5 minutes
- Allocation time—10 seconds
- Auction speedup time—3 seconds
- Number of participants—3 to 15

For a chosen distribution of demand profiles among participants, we performed 10 simulations for every number of participants from the range and every moderation model. Welfare has been calculated according to metric defined by Equation 2 and averaged over those 10 runs. Each agent had his set of demands generated before every single simulation.

Figure 3 contains results for the most ideal scenario, where every agent has a ”regular” demands profile. All allocation methods show decent behaviour, as there is was no special need to bother about fairness or overall welfare. ”Maximize metric” model outperforms all others, as expected.

In the next scenario, we had chosen the ceil of 10% of all agents and set their profiles to ”aggressive”. Results are shown in Figure 4. ”Choosing maximal valuation” clearly prefers allocation vectors which bring lower social welfare. It promotes aggressive agents over others and allows them to dominate the whole process.

Figure 5 shows results for a scenario where all participants represent ”dynamic” demands profile. In such environment ”Queuing” moderation model performed very poorly. It is due to the way it provides fairness. Allocating a resource to the first agent from the queue might very often prevent participants

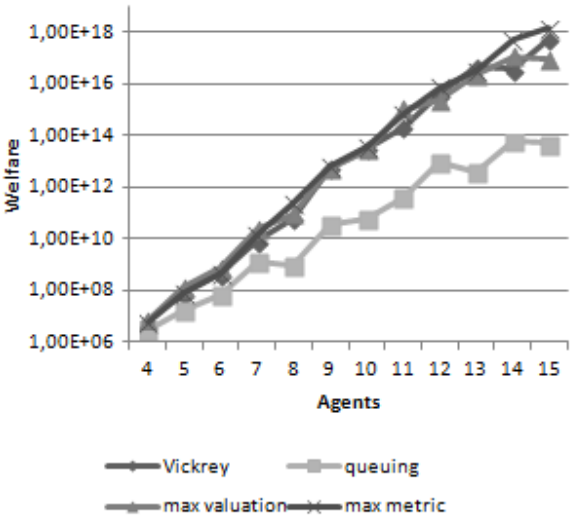


Fig. 5. Dynamic demands

from finding a common set of preferred topics. ”Maximizing welfare metric” is no longer strictly superior to other allocation model, as the choice of allocation might have an impact on participants’ future utility functions.

The last scenario we simulated is a mixture of the previous two. Among the participants with ”dynamic” demands the chosen ceil of 10% of agents have their profiles set to ”aggressive”. This environment highlights the weaknesses of both ”queuing” and ”choosing maximal valuation” which perform visibly worse than the remaining two.

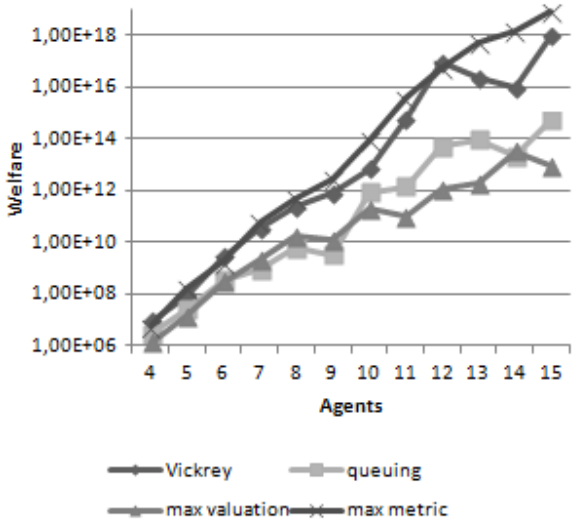


Fig. 6. Dynamic demands with 10 percent aggressive

VIII. Conclusions

In this paper we proposed a social welfare metric to determine the quality of a conference taking fairness aspects into consideration. Further on, we introduced a multiagent resource allocation scheme which embraces the described concept of fairness while getting as much out of overall social welfare as possible. We have performed a whole series of simulated discussions in order to verify the model's quality. Empirical results show that this resource allocation procedure is very much in line with the social welfare metric defined by Equation 2. It generates an allocation which is fair and highly valued by the whole society in a hostile environment with agents trying to dominate. We have also observed good results with semi-intelligent conversation aware agents.

Future work includes testing the model in a real life scenario by deploying to a broad public. We would also like to lay a more solid theoretical foundation to our definition of fair allocation. Analyzing whether our concept is in line with fair dominance [13] should be a good starting point. Another area of theoretical research are budget bound auction mechanisms [14] and investigate the impact on the introduced Vickrey auction based process.

References

[1] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, J. Riedl, and H. Volume, "Grouplens: Applying collaborative filtering to usenet news," *Communications of the ACM*, vol. 40, pp. 77–87, 1997.

[2] C. Lampe and P. Resnick, "Slash(dot) and burn: distributed moderation in a large online conversation space," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI '04. New York, NY, USA: ACM, 2004, p. 543550.

[3] T. Hasselrot, "Fair bandwidth allocation in internet access gateways - using agent-based electronic markets," SICS, Tech. Rep., 2003.

[4] A. Polomski, "Multiagent scheme for voice conference moderation," in *FedCSIS*, 2012, pp. 1215–1220.

[5] Y. Chevaleyre, P. E. Dunne, U. Endriss, J. Lang, M. Lematre, N. Maudet, J. Padget, S. Phelps, J. A. Rodrguez-aguilar, and P. Sousa, "Issues in multiagent resource allocation," *Informatica*, vol. 30, p. 2006, 2006.

[6] K. J. Arrow, *Social Choice and Individual Values, Second edition (Cowles Foundation Monographs Series)*, 2nd ed. Yale University Press, Sep. 1970.

[7] J. M. Vidal, "Fundamentals of multiagent systems," 2006. [Online]. Available: http://www.multiagent.com/fmas

[8] H. Moulin, *Axioms of Cooperative Decision Making (Econometric Society Monographs)*. Cambridge University Press, Jul. 1991.

[9] Y. Narahari, D. Garg, R. Narayanam, and H. Prakash, *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[10] M. H. Rothkopf, "Thirteen reasons why the vickrey-clarke-groves process is not practical," *Oper. Res.*, vol. 55, no. 2, pp. 191–197, Mar. 2007.

[11] F. Bellifemine, G. Caire, A. Poggi, and G. Rimassa, "JADE: A White Paper," *EXP in search of innovation*, vol. 3, no. 3, pp. 6–19, 2003.

[12] J. G. Carbonell, "Intentionality and human conversations," in *Proceedings of the 1978 workshop on Theoretical issues in natural language processing*, ser. TINLAP '78. Stroudsburg, PA, USA: Association for Computational Linguistics, 1978, pp. 141–148.

[13] W. Ogryczak, "Bicriteria models for fair and efficient resource allocation," in *SocInfo*, ser. LNCS, vol. 6430, 2010, pp. 140–159.

[14] H. Varian, "Position auctions," *International Journal of Industrial Organization*, vol. 25, no. 6, pp. 1163–1178, Dec. 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1029–1035

Verifying data integration agents with deduction-based models

Radosław Klimek*, Łukasz Faber* and Marek Kisiel-Dorohinicki*
*AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {rklimek,faber,doroh}@agh.edu.pl

***Abstract*—The paper shows how an agent-based system can be subjected to formal verification using a deductive approach. The particular system for gathering open source intelligence is considered, which is build on a framework for data integration. Techniques allowing for automatic extraction of logical specifications are described with emphasis on pattern-based and rule-based approaches. An example illustrates how the proposed method works in a scenario with iterated agent tasks combining these two approaches.**

I. Introduction

THE key concept in multi-agent systems (MAS) are intelligent interactions (coordination, cooperation, negotiation). Thus multi-agent systems are ideally suited to representing problems that have multiple problem solving methods, multiple perspectives and/or multiple problem solving entities [1]. Yet this variety of perspectives often makes the design and implementation of software MAS a really difficult task.

Formal methods enable the precise formulation of important artifacts and the elimination of ambiguity. There are two well established approaches to formal reasoning and system verification [2]. The first is based on the state exploration ("model checking") and the second is based on deductive reasoning. Model checking is an operational rather than analytic approach [3]. On the other hand, deduction-based formal verification is essential for sustainable verification leverage, characterized by intuitiveness, a top-down way of thinking, logic-based reasoning, coverage of infinite computations, etc. Temporal logic is a well established formalism which allow to describe properties of reactive systems. The semantic tableaux method, which might be descriptively called "satisfiability trees", seems intuitive and may be considered as a goal-based formal reasoning.

The contribution is based on an agent-based framework dedicated to acquiring and processing distributed, heterogeneous data collected from the various Internet sources [4]. Data processing in such a system is structuralized by means of dynamic workflows based on agents' interactions. Our goal is to provide a formal description of these interactions to make sure the system works properly. Since logical specifications are difficult to specify manually, a method for an automatic extraction of logical specifications, considered as a set of temporal logic formulae, was proposed in [5], or for a building requirements models during the software requirements elicitation in [6]. Here, a case of iterated agent tasks is considered and illustrated by a more complex scenario in a slightly different application area. The case includes both an active model and a state model for agent systems.

In the first part of the paper essential logical background is provided and the method of specification generation based on workflow describing agents' interactions is described. Then the general structure of the framework for data integration is presented. This constitutes a base for the discussion of the scenario of the particular system, which shows how the approach works in practice.

II. Logical preliminaries

Logical background which is temporal logic, semantic tableaux, and the deduction-based verification system are discussed below. *Temporal logic* TL is a formal and logical system for the specification and verification of software models and systems [7]. It introduces symbolism (unary and dual operators are $\Diamond$ for "sometime in the future" and $\Box$ for "always in the future") for representing and reasoning about the truth and falsity of formulas throughout the flow of time and taking into consideration changes of their valuation. It allows to describe both temporal relations between reached states and to specify expected properties. The attention is focused on *propositional linear-time temporal logic* PLTL, i.e. the time structure constitutes a linear and unbounded sequence. Each element in the mentioned sequence corresponds to a propositional world, i.e. *atomic propositions* AP are valued in every point of the sequence. Temporal logic and their syntax and semantics are discussed in many works, e.g. [8], [7]. Considerations in the work are limited to the *smallest temporal logic*, e.g. [9], [10], which is extension of a classical propositional calculus' to the axiom $\Box(\Psi \Rightarrow \Phi) \Rightarrow (\Box\Psi \Rightarrow \Box\Phi)$ and the inference rule $\vdash \Psi \Longrightarrow \vdash \Box\Psi$. The minimal logic, also called the K logic, is sufficient to define many system properties (liveness, safety), The following formulas may be considered as examples of this logic: $action \Rightarrow \Diamond reaction$, $\Box(send \Rightarrow \Diamond ack)$, $\Diamond live$, $\Box\neg(event)$, etc.

Semantic tableaux is a decision procedure, based on formula decomposition, for checking formula satisfiability. Even though it is known in classical logic, it can also be applied in temporal logic [11]. At each step of a well-defined procedure, some logical connectives are removed and formulas are decomposed. The method is a proof by contradiction, i.e. after negation of the initial formula, finding a contradiction in all

978-83-60810-53-8/$25.00

branches means that the inference tree is *closed*, and there are no valuations that satisfy a formula. It leads to the statement that the formula before the negation is true. The method provides, through so-called *open* branches of the semantic tree, information about the source of an error, if one is found. The work [12] provides an example of the inference tree.

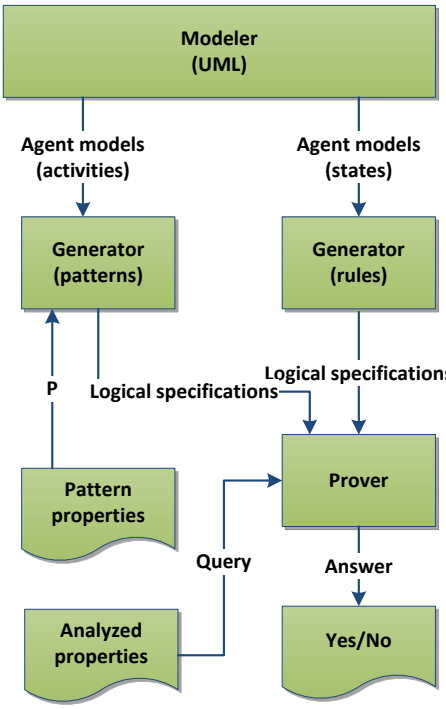


Fig. 1. A deduction-based verification system for agent models

The outline architecture of the proposed deduction-based system for agent models is presented in Fig. 1. There are two generation components. The first one works using the algorithm Γ_1 (described in the next section) and is designed for models expressed in activity diagrams using predefined workflow patterns. The second one works using the algorithm Γ_2 (described in the next section) and is designed for models expressed in state diagrams. The outputs of these generation components are logical specifications understood as sets of temporal logic formulas. The combined specification is treated as a conjunction of all formulas $p_1, \ldots, p_n = L$ and every p_i is a specification formula generated during the extraction Γ_1 or Γ_2. These formulas constitute a logical specification L. The Q formula (query) is a desired system property for the analysed software model.

Both the final specification of a system and the examined properties constitute an input to the prover component, which works using the semantic tableaux method. It enables the automated reasoning. The input for this component is the formula $C(L) \Rightarrow Q$, where $C(L)$ means conjunction of all extracted formulas, or, more precisely:

$$p_1 \wedge \ldots \wedge p_n \Rightarrow Q \tag{1}$$

After negation of formula (1), it is placed at the root of the inference tree and decomposed using the semantic tableaux method's well-defined rules. The work [12] provides an example of the inference tree.

The whole verification procedure can be summarised in the following way:

1) Automatic generation of a logical specification based on design patterns (Γ_1);
2) Automatic generation of a logical specification based on extraction rules (Γ_2)
3) Introduction a property Q as a query for the considered model;
4) The automatic inference using semantic tableaux for the whole formula 1.

Steps 1 to 4, taken as a whole or individually, may be processed many times, whenever models are changed (step 1 or step 2) or there is a need for a new reasoning due to the revised system's property (step 3).

III. Logical specifications

Two generation methods for extraction logical specifications are considered in this section. The first one which is based on predefined workflows is discussed in a more detailed way. However, the rule-based approach is an interesting alternative for generating logical specifications.

A. A pattern-based approach

Presentation of the approach needs to introduce some basic notions and definitions. An *elementary set* of formulas over atomic formulas $a_{i,i=1,\ldots,n}$ is denoted $pat(a_i)$, or simply $pat()$, as a set of temporal logic formulas $\{f_1, ..., f_m\}$ such that all formulas are syntactically correct. The examples of elementary sets are $Pat1(a,b) = \{a \Rightarrow \Diamond b, \neg a \Rightarrow \neg b, \Box\neg(a \wedge b)\}$ and $Pat2(a,b,c) = \{a \Rightarrow \neg\Diamond b \wedge \Diamond c, \Box\neg(b \vee c)\}$. The logical expression enables representing nested and complex structures elementary sets. The *logical expression* W_L is a structure created using the following rules [13]:

- every elementary set $pat(a_i)$, where $i > 0$ and every a_i is an atomic formula, is a logical expression,
- every $pat(A_i)$, where $i > 0$ and every A_i is either
  - an atomic formula a_j, where $j > 0$, or
  - a set $pat(a_j)$, where $j > 0$ and a_j is an atomic formula, or
  - a logical expression $pat(A_j)$, where $j > 0$

  is also a logical expression.

The example of logical expression is $Seq(Flow(a,b,c), Switch(d,e,f))$ which is intuitive, in that it shows the sequence of a parallel split (flow) and then conditional execution (switch) of some activities.

Workflow patterns constitute a kind of primitives and enable the automation of the generation process for logical specifications. It leads to the mapping of workflow patterns to logical specifications. The proposed approach is based on the assumption that the entire activity diagrams are built using only predefined workflow patterns. The assumption is

not a restriction since it enables receiving correct and well-composed systems. *Activity diagrams* of UML enable modelling workflow activities. They support choice, concurrency and iteration. The important goal of diagrams is to show how an activity depends on others [14].

```
Sequence(a1,a2):    /* ver. 13.04.2013
in={a1} / out={a2}
a1 => <>a2 / []~(a1 & a2)
SeqSeq(a1,a2,a3):
in={a1} / out={a3}
a1 => <> a2 / a2 => <> a3
[]~((a1 & a2) | (a2 & a3) | (a1 & a3))
Flow(a1,a2,a3):
in={a1} / out={a2,a3}
a1 => <>a2 & <>a3 / []~(a1 & (a2|a3))
Switch(a1,a2,a3):
in={a1} / out={a2,a3}
a1 & c(a1) => <>a2 / a1 & ~c(a2) => <>a3
[]~((a1 & a2) | (a1 & a3) | (a2 & a3))
Loop-While(a1,a2):
in={a1} / out={a1,a2}
a1 & c(a1) => <> a2 / a1 & ~c(a1) => ~<> a2
[]~(a1 & a2)
```

Fig. 2. Predefined set of pattern temporal properties

Logical properties of all design patterns are expressed in temporal logic formulas. They are stored in the predefined and fixed *logical properties set* P. An example of such a predefined set P for the UML activity diagrams is shown in Fig. 2. Most elements of the P set, i.e. two temporal logic operators, classical logic operators, etc. are not in doubt. a_1, a_2 and a_3 are atomic formulas and constitute a kind of formal arguments for a pattern. The slash allows to place more than one formula in a single line. $c(a)$ means that the logical condition associated with the activity a has been evaluated and is satisfied. The pattern $SeqSeq$ means the concatenation of sequences as a sequence of three arguments. Variables in and out provide information about activities for a pattern that are the first and the last to be executed, respectively. They enable representing pattern to be considered as a whole. All formulas describe both safety and liveness properties for every pattern [15]. Summing up, the predefined set of pattern temporal properties consists of the following elements $\{Seq, SeqSeq, Flow, Switch, LoopWhile\}$ the meaning of which seems intuitive, i.e. sequence, sequence of a sequence, concurrency, choice and iteration.

Generating a logical specification is not a simple summation of formula collections resulting from a logical expression. The generation algorithm Γ_1 sketch for obtaining a set of temporal formulas is given below.

1) At the beginning, the logical specification is empty, i.e. $L := \emptyset$;
2) Patterns are processed from the most nested pattern to be located more towards the outside and from left to right;
3) If the currently analysed pattern consists only of atomic formulas, the logical specification is extended, by summing sets, by formulas linked to the type of pattern analysed, i.e. $L := L \cup pat()$;
4) If any argument is a pattern itself, then the logical disjunction of all elements that belong to in and out sets, is substituted in a place of the pattern;

The algorithm is a modification of the similar one presented in [13]. All patterns of the logical expression are processed one by one and the algorithm always halts. All parentheses are paired. The example of the algorithm is provided in the section V-B.

A logical specification L_1 consists of all formulas obtained from a logical expression using the algorithm Γ_1, i.e.

$$L_1(W_L) = \{f_i : i > 0 \wedge f_i \in \Gamma_1(W_L, P)\} \tag{2}$$

where f_i is a PLTL formula. The sketch of the generation algorithm is presented below. The generation process has two inputs. The first one is a logical expression and the second one is a predefined set of logical properties. The output is a set of logical formulas.

B. A rule-based approach

However, the rule-based approach for generating logical specifications is an interesting alternative for the previous one, i.e. presented in the section III-A. The approach seems suitable for the UML state diagram when considering set of states (nodes) and transitions (edges). The discussion is limited to some basic situations which are defined in terms of temporal logic formulas. Considering all transitions one by one, a logical specification understood as a set of temporal logic formulas is obtained using the following rules which constitute the generation algorithm Γ_2:

- R.1 (Sequence) – the state b is enabled when the state a is reached and an event e occurred, i.e.:
  $\{(a \wedge e) \Rightarrow \Diamond b, \Box\neg(a \wedge b)\}$;
- R.2 (Split) – when the state a is reached and an event e occurred then single thread of control is splited into two threads of control to enable parallel reaching the state b and the state c, i.e.:
  $\{(a \wedge e) \Rightarrow \Diamond b \wedge \Diamond c, \Box\neg(a \wedge (b \vee c))\}$;
- R.3 (Synchronization) – when two states a and b are reached in parallel and two events $e1$ and $e2$ occurred, respectively, then threads of control are transformed and synchronised into a single thread of control to enable reaching the state c, i.e.:
  $\{(a \wedge e1) \wedge (b \wedge e2) \Rightarrow \Diamond c, \Box\neg((a \vee b) \wedge c)\}$.

A logical specification L_2 consists of all formulas obtained using the algorithm Γ_2, i.e.

$$L_2 = \{f_i : i > 0 \wedge f_i \in \Gamma_2\} \tag{3}$$

where f_i is a PLTL formula. The input for the generation algorithm is a state diagram. The output is a set of logical formulas.

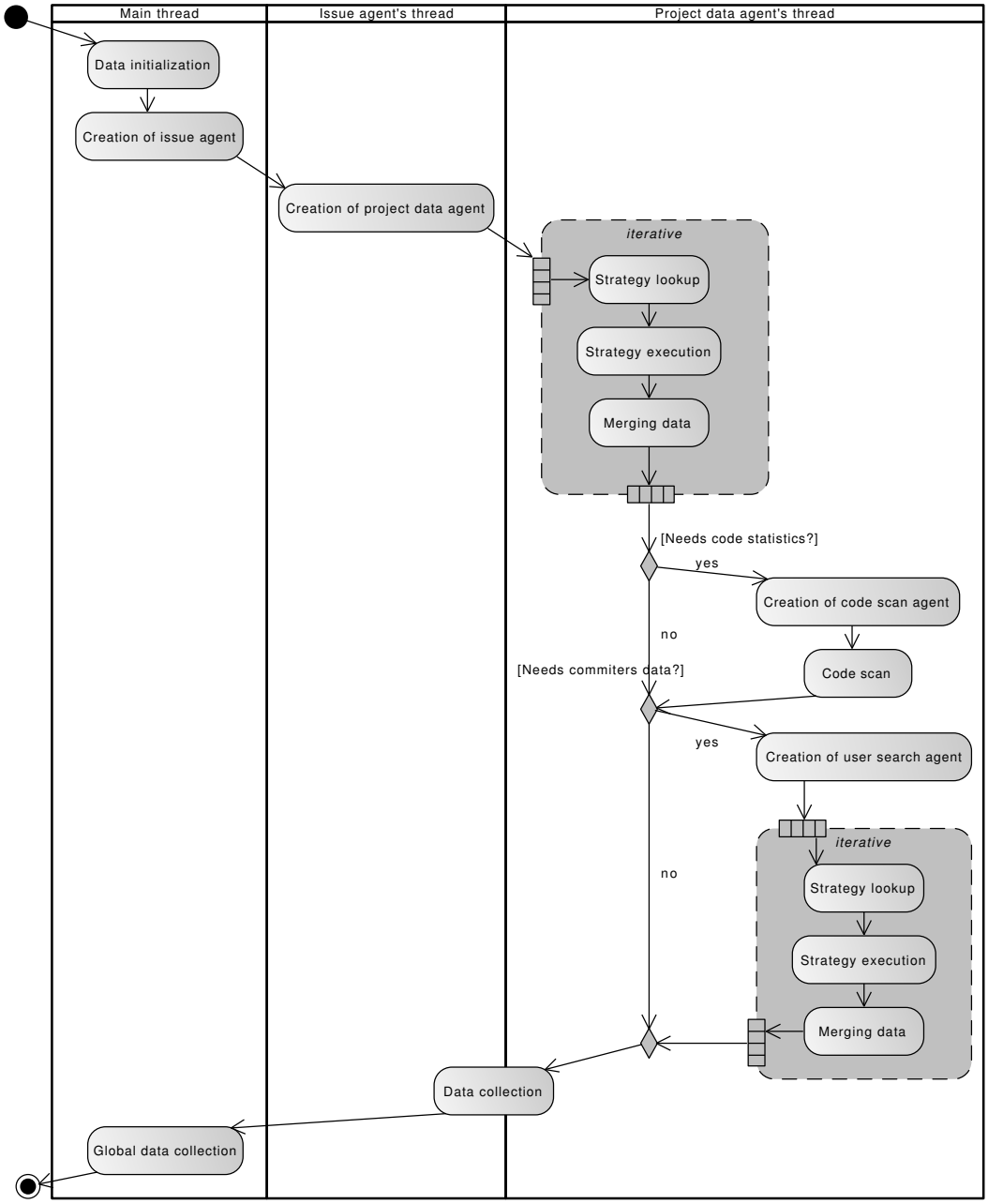


Fig. 3. Activity diagram of the search scenario presented in the section V-A.

IV. Agent-based framework for data integration

The vast amount of information available in the global network calls for complex systems able to perform various analyses with respect to data coming from various, often heterogeneous sources. The described framework provides the data- and task-oriented workflow for collecting and integrating data from a wide range of diverse services [4], [5]. Fig. 4 presents the layered structure of the framework:

- Presentation – the usage of different views is possible. They communicate with the rest of the system and allows the user to interact and control the act of the data integration.
- Middleware – the logical processing of the data is performed here. This part of the system uses the agent paradigm to delegate different parts of the data integration and processing to other parts of the system (or the external software, frameworks, etc.) that are represented also as agents.
- Services and data – wrapping of the various external data sources (and data processing capabilities) takes place here. Interfaces of the systems are adapted and described

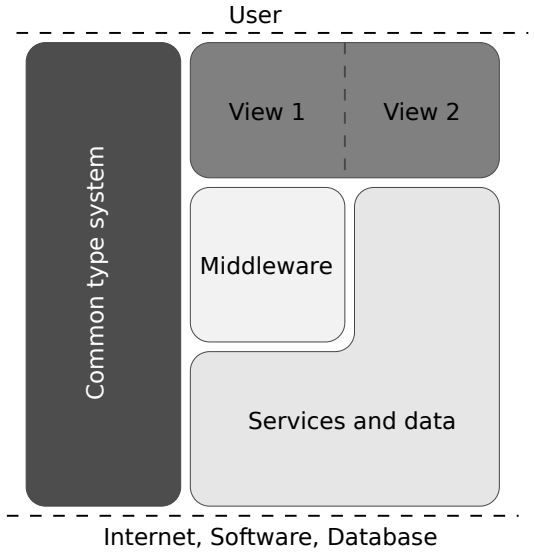


Fig. 4. Layered structure of the implementation

in terms of the common type system.
- Common type system – the whole framework uses the types described here to annotate the processed data, data sources and external systems interfaces or façades.

Queries created by the user are put into the agent system that performs two types of operations: the management of data (by inspecting queries and delegating them to other agents) and execution of demanded actions (including their selection, configuration and fault recovery). The system can divide processing into *issues*. An *issue* is a separate part of data processing, usually centred around an instance of a query object and all related (i.e. found) data.

The current implementation defines three possible functional roles for agents.

- **System agents** provide a bootstrapping and basic functionality. They create new issues, handle errors, monitor the system.
- **Issue Agents** care about a single issue (i.e. user query) and delegate initial tasks to specialised action agents. An Issue Agent retrieves a query from the pool, inspects it and then requests (with messages) a chosen data agent to resolve a query specified in the task.
- **Action (Data) Agents** implement the actual executive part of the search functionality. Upon receiving the task from an issue agent they obtain appropriate strategies and then executes them to answer the query.
  However, they are not constrained only to strategies. They can perform any action on data: merge, simplify, verify, etc.

Issue Agents are identified by a runtime-generated issue identifier that represents an issue hey are taking care of. The Action Agents are described during creation (implementation) with tasks they can perform (called "capabilities") and data types they can operate on.

V. Data collection scenario and its formal analysis

We consider an act of gathering data about open-source projects as one of the possible use-cases. This kind of data can be easily mined from popular infrastructure-providing websites like GitHub, SourceForge, etc. Moreover, it is usually very well interlinked and it makes it possible to gather much broader information about e.g., people working on the project, organisations involved, etc.

A. Base Scenario

The scenario has three sample steps:

1) Gathering data about an open-source project from all available sources.
   The user inputs a query comprising of, for example, a name of the project or the link to its website into the system. The system performs a look-up of possible matches and returns a match of a list of possible matches. In the latter case, there may be a requirement to choose one (the best) match. It can be done manually or delegated to an agent that can rate each result and select the best one.
2) Scanning of the source code of the project.
   If possible, a user may expect the project's code to be scanned to obtain particular statistics. If expected, an agent responsible for project data may create another agent that can scan the code. Such an agent will perform the scan and merge the results back into the data handled by the project data agent.
3) Gathering information about commiters and authors.
   It may be further required to gather data about commiters and authors. The project data agent would create a user search agent that is able to collect users data from project website and other related sources.

On the system level our scenario is implemented as follows:

- The types like Project, Commiter, Code are introduced.
- The action agents that performs operations on types are implemented: Project Data Agent, Code Scan Agent, User Search Agent.
- Strategies for each external service can be created: Project Data Search and Code Scan for e.g., GitHub, SourceForge and User Data Search for, e.g., LinkedIn or Facebook.

Fig. 3 shows the activity diagram of an actual execution of the scenario. The user prepares a query that consists of the initial data to operate on (e.g. a name of the project). The task is placed into the system (*Data Initialization* action). Then, available issue agents are notified about it or a new one is created. The issue agent locates an implementation of a so-called action agent that can handle the specified task (Project Data Agent), instantiates it, and delegates the task execution to this agent.

Project Data Agent inspects the query and calls relevant strategies (to locate and gather project data). Then it may choose to instantiate the Code Scan Agent in order to scan the project's code. The same goes for commiters data – the

Project Data Agent can instantiate a specialised agent that will look up personal data.

Data collection is performed in two phases: first, the data from strategies is merged by a responsible agent (e.g., data about projects by the Project Data Agent); second, all data from an agent is inserted into the object provided by an agent higher in the hierarchy (e.g., Project Data Agent provides a Project object that has a field *commiters* that is filled by the User Search Agent). The query execution is finished by putting results to the pool presented to the user.

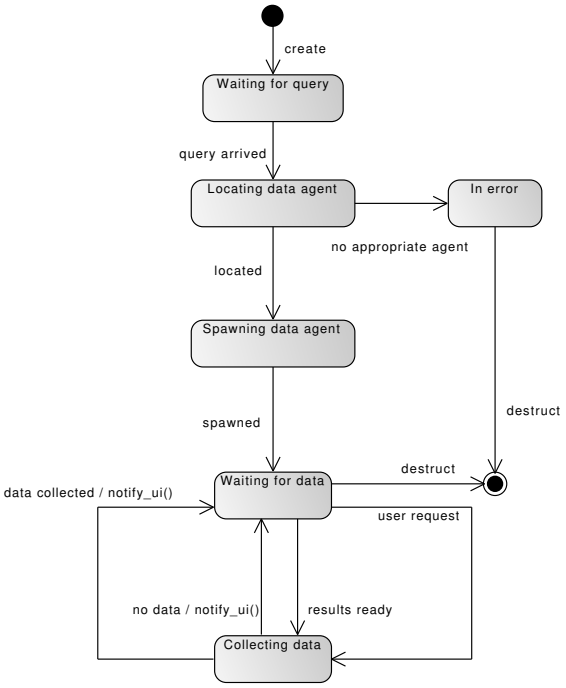


Fig. 5. State diagram of the Issue Agent in the search scenario presented in the section V-A.

Fig. 5 presents sample states of the Issue Agent that is responsible for initiating search for data and forwarding data to the user.

After creation and initialisation, an agent waits for a query from the user. When the query arrives, the agent tries to locate an implementation of a specialised data agent (in the scenario described earlier — Project Data Agent). If such an agent cannot be found it results in a fatal error as it is usually caused by a wrong configuration.

If the data agent is found it is instantiated by the Issue Agent and the query is forwarded to it. Then, the Issue Agent waits for either results from the data agent or request from the UI to get that data. It puts the data, if available, into the user pool and notifies the UI about it. If no data has been found, it generates an error for the UI.

B. Formal analysis and verification

Let us consider the activity diagram shown in Fig. 3. Diagram activities represent propositions which are used for modelling behaviour. Firstly, letters of the Latin alphabet are substituted in a place of propositions. Replacing atomic activities by Latin letters is a technical matter and is suitable only for the work because of its limited size. (In the real world, the original names of activities are used.) The following substitutions are made: a – DataInitialisation, b – CreationIssueAgent, c – CreationProjectDataAgent, d – ProjectDataLookup, e – NeedsCodeStatistics, f – CreationCodeScanAgent, g – CodeScan, h – NeedsCommitersData, i – CreationUserSearchAgent, j – UserSearchLookup, k – DataCollection, and l – GlobalDataCollection. (To simplify considerations, a single proposition instead of a loop is used.) The logical expression W_L for the activity diagram is

$$SeqSeq(SeqSeq(a,b,c), SeqSeq(d, Switch(e, Seq(f,g), N1), Switch(h, Seq(i,j), N2)), Seq(k,l)) \quad (4)$$

Activity diagrams constitute one of two inputs for the deduction system shown in Fig. 1. Two activities $N1$ and $N2$ are introduced since the diagram in Fig. 3 contains two switches without activity (null activity). A logical specification L_1 for the logical expression W_L is built using the algorithm Γ_1 presented in the section III-A. The logical specification is

$$\begin{aligned} L_1 = \{ & f \Rightarrow \Diamond g, \Box\neg(f \wedge g), i \Rightarrow \Diamond j, \Box\neg(i \wedge j), \\ & e \wedge c(e) \Rightarrow \Diamond(f \vee g), e \wedge \neg c(e) \Rightarrow \Diamond N1, \\ & \Box\neg((e \wedge (f \vee g)) \vee (e \wedge N1) \vee ((f \vee g) \wedge N1)), \\ & h \wedge c(h) \Rightarrow \Diamond(i \vee j), h \wedge \neg c(h) \Rightarrow \Diamond N2, \\ & \Box\neg((h \wedge (i \vee j)) \vee (h \wedge N2) \vee ((i \vee j) \wedge N2)), \\ & d \Rightarrow \Diamond e, e \Rightarrow \Diamond(j \vee N2), \\ & \Box\neg((d \wedge e) \vee (e \wedge (j \vee N2)) \vee (d \wedge (j \vee N2))), \\ & a \Rightarrow \Diamond b, b \Rightarrow \Diamond c, \Box\neg((a \wedge b) \vee (b \wedge c) \vee (a \wedge c)), \\ & k \Rightarrow \Diamond l, \Box\neg(k \wedge l), (a \vee c) \Rightarrow \Diamond(d \vee i \vee j), \\ & (d \vee i \vee j) \Rightarrow \Diamond(k \vee l), \\ & \Box\neg(((a \vee c) \wedge (d \vee i \vee j)) \vee ((d \vee i \vee j) \wedge (k \vee l)) \vee \\ & ((a \vee c) \wedge (k \vee l)))\} \end{aligned} \quad (5)$$

Formula 5 represents the output of the generator component in Fig. 1.

Let us consider the state diagram shown in Fig. 5. Firstly, letters of the Latin alphabet are substituted in a place of states. The following substitutions are made: m – WaitingQuery, n – LocatingDataAgent, o – InError, p – SpawningDataAgent, q – WatingData, r – CollectingData, and s – Stop. Next, the following substitutions for events are made: $e'a$ – QueryArrived, $e'b$ – Located, $e'c$ – NoAgent, $e'd$ – Spawned, $e'e$ – Destruct, $e'f$ – UserRequest, $e'g$ – DataCollected, $e'h$ – NoData, and $e'i$ – ResultsReady. State diagrams constitute one of two inputs for the deduction system shown in Fig. 1. A logical specification L_2 is built using the algorithm Γ_2 presented in the section III-B. The logical specification is

$$\begin{aligned} L_2 = \{ & (m \wedge e'a) \Rightarrow \Diamond n, \Box\neg(m \wedge n), (n \wedge e'b) \Rightarrow \Diamond p, \\ & \Box\neg(n \wedge p), (n \wedge e'c) \Rightarrow \Diamond o, \Box\neg(n \wedge o), \\ & (o \wedge e'e) \Rightarrow \Diamond s, \Box\neg(o \wedge s), (p \wedge e'd) \Rightarrow \Diamond q, \end{aligned}$$

$$\Box\neg(p \wedge q), (q \wedge e'e) \Rightarrow \Diamond s, \Box\neg(q \wedge s),$$
$$(q \wedge e'i) \Rightarrow \Diamond r, \Box\neg(q \wedge r), (q \wedge e'f) \Rightarrow \Diamond r,$$
$$(r \wedge e'h) \Rightarrow \Diamond q, (r \wedge e'g) \Rightarrow \Diamond q\} \quad (6)$$

Formula 6 represents the output of the generator component in Fig. 1.

Logical specifications which are built in the above way, i.e. using the proposed algorithms, can be verified formally. Formal *verification* is the act of proving correctness properties of a system. When the semantic tableaux method is used, then the whole reasoning process can be summarised as a process of the verification whether an entailment $F_1, ...F_n \models G$ it suffice to prove that $\{F_1, ..., F_n, \neg G\}$ is unsatisfiable. The *liveness* property of a system means that the computational process achieves its goals, i.e. something good eventually happens. The *safety* property of a system means that the computational process avoids undesirable situations, i.e. something bad never happens. The liveness property for the above model can be

$$b \Rightarrow \Diamond l \quad (7)$$

which means that **if creation of issue agent is satisfied then sometime global data collection is reached**, formally $CreationIssueAgent \Rightarrow \Diamond GlobalDataCollection$. When considering property expressed by formula 7 then the whole formula to be analysed using semantic tableaux, providing a combined input for the prover component in Fig. 1, is

$$C(L_1) \wedge C(L_2) \Rightarrow (b \Rightarrow \Diamond l) \quad (8)$$

where $C(L_x)$ means logical conjunctions of all formulas which belong to L_x, c.f. formula 5 and 6. Presentation of a full inference tree exceeds the size of the work. All branches of the semantic tree are closed, i.e. formula 7, is satisfied in the considered model. The method is easy to scale-up, i.e. extending and summing up logical specifications for other activity diagrams and state diagrams. Then, it is possible to examine logical relationships (liveness, safety) for different activities and states coming from different activity and state diagrams.

VI. Conclusions

The discussed agent-based data integration framework is based on AgE platform [16], which allows to introduce a clear separation of concerns and makes it possible to rapidly build topic-related versions of the system with little configuration. The framework is completely independent of the data it processes and all scenario-related components are provided by the user during the configuration. That is why the proposed verification approach seems to be particularly important to ensure proper configuration of the system.

Future work may include the implementation of the logical specification generation module and the temporal logic prover. Important results might be a CASE software for both the workflow modelling and the deduction-based formal verification. Considering graph transformations [17] is encouraging for models involving distributed representation of knowledge and their efficient implementation. The example of such an implementation is discussed in [18].

Acknowledgment

This work was supported by the AGH UST internal grant no. 11.11.120.859.

References

[1] N. R. Jennings, K. Sycara, and M. Wooldridge, "A roadmap of agent research and development," *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 1, no. 1, pp. 7–38, 1998.

[2] E. Clarke, J. Wing, and et al., "Formal methods: State of the art and future directions," *ACM Computing Surveys*, vol. 28 (4), pp. 626–643, 1996.

[3] E. Clarke, O. Grumberg, and D. Peled, *Model Checking*. MIT Press, 1999.

[4] E. Nawarecki, G. Dobrowolski, A. Byrski, and M. Kisiel-Dorohinicki, "Agent-based integration of data acquired from heterogeneous sources," in *Proc. of 5th Int. Conf. on Complex, Intelligent and Software Intensive Systems - CISIS 2011*, 2011.

[5] R. Klimek, Ł. Faber, and M. Kisiel-Dorohinicki, "Deduction-based modelling and verification of agent-based systems for data integration," in *Proceedings of 3rd International Conference on Man-Machine Interactions (ICMMI 2013), 22–25 October 2013, The Beskids, Poland [paper accepted]*, ser. Advances in Intelligent Systems and Computing, T. Czachórski, A. Gruca, and S. Kozielski, Eds. Springer Verlag, 2013.

[6] R. Klimek, "From extraction of logical specifications to deduction-based formal verification of requirements models," in *Proceedings of 11th International Conference on Software Engineering and Formal Methods (SEFM 2013), 25–27 September 2013, Madrid, Spain*, ser. Lecture Notes in Computer Science, R. Hierons, M. Merayo, and M. Bravetti, Eds., vol. 8137. Springer Verlag, 2013, pp. 61–75.

[7] F. Wolter and M. Wooldridge, "Temporal and dynamic logic," *Journal of Indian Council of Philosophical Research*, vol. XXVII(1), pp. 249–276, 2011.

[8] E. Emerson, *Handbook of Theoretical Computer Science*. Elsevier, MIT Press, 1990, vol. B, ch. Temporal and Modal Logic, pp. 995–1072.

[9] B. F. Chellas, *Modal Logic*. Cambridge University Press, 1980.

[10] J. van Benthem, *Handbook of Logic in Artificial Intelligence and Logic Programming*, ser. 4. Clarendon Press, 1993–95, ch. Temporal Logic, pp. 241–350.

[11] M. d'Agostino, D. Gabbay, R. Hähnle, and J. Posegga, *Handbook of Tableau Methods*. Kluwer Academic Publishers, 1999.

[12] R. Klimek, "Temporal preference models and their deduction-based analysis for pervasive applications," in *Proceedings of 3rd International Conference on Pervasive and Embedded Computing and Communication Systems (PECCS 2013), 19–21 February, 2013, Barcelona, Spain*, C. Benavente-Peces and J. Filipe, Eds. SciTePress, 2013, pp. 131–134.

[13] ——, *Advanced Methods and Technologies for Agent and Multi-Agent Systems*, ser. Frontiers of Artificial Intelligence and Applications. IOS Press, 2013, vol. 252, ch. A Deduction-based System for Formal Verification of Agent-ready Web Services, pp. 203–212. [Online]. Available: http://ebooks.iospress.nl/publication/32843

[14] T. Pender, *UML Bible*. John Wiley & Sons, 2003.

[15] B. Alpern and F. B. Schneider, "Defining liveness," *Information Processing Letters*, vol. 21 (4), pp. 181–185, 1985.

[16] L. Faber, K. Piętak, A. Byrski, and M. Kisiel-Dorohinicki, "Agent-based simulation in age framework," in *Advances in Intelligent Modelling and Simulation*, ser. Studies in Computational Intelligence, A. Byrski, Z. Oplatková, M. Carvalho, and M. Kisiel-Dorohinicki, Eds. Springer Berlin Heidelberg, 2012, vol. 416, pp. 55–83. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-28888-3_3

[17] L. Kotulski, "Supporting software agents by the graph transformation systems," in *International Conference on Computational Science*, ser. Lecture Notes in Computer Science, V. N. Alexandrov and et al., Eds., vol. 3993. Springer, 2006, pp. 887–890.

[18] I. Wojnicki, L. Kotulski, and S. Ernst, "On scalable, event-oriented control for lighting systems," *Frontiers in Artificial Intelligence and Applications: Advanced Methods and Technologies for Agent and Multi-Agent Systems*, vol. 252, pp. 40–49, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1037–1040

Agent Based System for Assistance at Industrial Process Control with Experience Modeling

Gabriel Rojek
AGH University of Science and Technology
Faculty of Metals Engineering and Industrial Computer Science
Department of Applied Computer Science and Modelling
al. A. Mickiewicza 30, 30-059 Kraków, Poland
Email: rojek@agh.edu.pl

***Abstract*—The problem of automatic or computer aided control is still unsolved in many areas of real production processes. In such cases the only one solution is to employ human operator that uses his experience and knowledge in order to manually control parameters of the process. Such approach has many disadvantages relating to characteristics of human work what is the main ground for presented here research. As the solution a methodology is proposed, which follows the decision processes of human operator using his experience. In order to predict capabilities of proposed methodology a test system is designed and implemented with the use of agent technology.**

I. Introduction

THE MAIN genesis of presented here work is an effort to design and to build a computer system supporting control of an industrial process, that is difficult to control with known computational techniques. The main reason for this difficulty is the lack of proper analytical model of such industrial process. This lack prevents obtaining proper values of process parameters by simply computing of optimal values from determined dependence in the form of mathematical equations. Possible solutions can be found in main areas of intelligent control which are fuzzy control, neural networks, expert systems and genetic algorithms [1]. Presented methodologies have some restrictions, which prevent against their use in every case of control problem. One of such restrictions is need for specifying rules or knowledge about proper control (in case of fuzzy control or expert system), what is impossible in many domains of control of industrial processes.

An example of an industrial process, that analytical model is unknown, is the oxidizing roasting process of sulphide zinc concentrates. It is also impossible to formulate knowledge or rules, which specify proper control of this process. Possible solution for control of this process was realized with the use of genetic algorithms and neural network, which predicts the value of fitness function on the basis of previous registered process signals [2]. This approach leads to interpolation of some signals, which are measured with very low frequency. Because some signals are measured once per minute and other only once per day, such interpolation results in obtaining of unreal values of signals, what can be a source of faults and errors and can have disadvantageous influence on process control. The goal of presented here work is presentation of a methodology, that eliminates need for data pre-processing in order to avoid adding of missing data, what is unnatural for decision taking by human workers. The proposed system, as the implementation of presented methodology, should follow decision process of human using his experience, that coincides with case-base reasoning (CBR) methodology (presented in e.g. [3], [4]).

This paper contents 6 sections. In the 2nd section the oxidizing roasting process of sulphide zinc concentrates is presented. The 3rd section presents remarks on experience using and gathering, what is the basis for design of the multi-agent system that is presented in the 4th section. The 5th section presents implementation aspects and tests of developed system, which tests enable to predict usefulness of proposed solutions. The last section presents conclusions.

Financial support of the AGH 11.11.110.085 is acknowledged

II. An Example of Industrial Process

The oxidizing roasting process of sulfide zinc concentrates is the first stage of industrial zinc production. During the roasting process, the aim is to obtain a minimum content of sulphide sulfur in the composition of the product. The generic scenario of control made by a human worker during one production day (that is full production cycle) is presented onto Fig. 1. Shown scenario of control indicates differences in frequency of signal measuring, however all signals have hypothetical influence on quality of products. All signals can be classified into one of three generic groups: independent, controllable and dependent signals.

Independent signals (I) are these parameters, which cannot be modified or changed during production cycle, so it is impossible to change their values in direct or indirect way. In the case of analyzed industrial process independent signals are parameters of chemical composition of raw materials, that are measured only once per a production day. **Dependent signals** (D) are these parameters, which cannot be directly modified. Value of a dependent signal is a hypothetical function of other process parameters and possible time delay. This function is unknown in a case of analyzed process. An example of dependent signal is the pressure or temperature at the upper part of a furnace. **Controllable signals** (C) are these, which can be directly changed or updated during the controlled

978-83-60810-53-8/$25.00

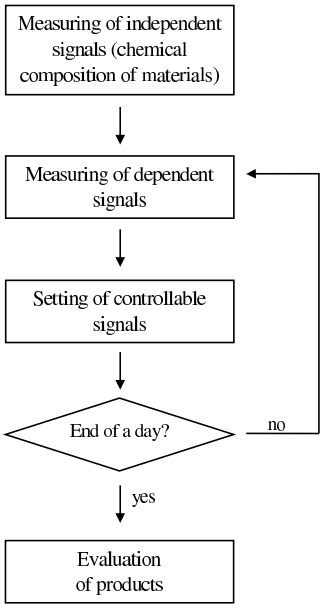


Fig. 1. The generic scenario of process control performed for one day of production

process. The rotary speed of a fan cooling the controlled process can be given as an example of a controllable signal in a case of analyzed process.

In a case of the oxidizing roasting process of sulfide zinc concentrates dependent and controllable signals are measured very often comparing to frequency of independent signal measure. The low frequency of some measures at industrial process control is caused by measured technique - the chemical composition has to be measured manually.

A. Decision Problem

The goal of industrial process control is obtaining products that are characterized by optimal properties. In a case of analyzed industrial process the goal is to obtain minimal concentration of sulphide sulphur in the roasted products, so the **quality evaluation** (Q) is average concentration of sulphide sulphur in the roasted products during production day. Formal definition of decision problem can be presented in the form of statement: how to choose values of controllable signals (C) knowing values of independent (I) and dependent signals (D) in order to obtain best possible quality evaluation (Q). In other words, the quality evaluation is hypothetical function of all signals $Q = f(C, I, D)$, however during process control only controllable signals (C) can be directly changed. Controllable signals (C) should be adjusted to measured values of all other signals in order to maximize quality criterion (Q).

III. Experience and Its Model

Having a goal to follow decision processes that take place in the mind of a human operator of presented industrial process, his work during a day period should be analyzed. A day period is full production cycle. At the beginning of production day the operator knows values of independent signals (I), it means he knows chemical composition of raw materials used for production. The operator assumes, that this chemical composition is constant for the whole production day due to frequency of independent signals measure, which is done only once at the beginning of every production day. Before the start of process control the operator should decide how to control this process, what means how to set present values of controllable signals (C) taking into consideration currently measured values of dependent signals (D). This decision is based on his experience. The experience contains many episodes from past production, which are referred to as cases. Every case in experience contains information concerning solved problem, how this problem was solved and how production was evaluated.

As it was mentioned in the above paragraph, before the start of process control the operator should decide how to control this process with the use of his experience. Accordingly to presented criterion of evaluation, the human operator first searches for cases that concern the same or similar problems and next chooses one case, which brings the best effect described by value of evaluation criterion. Afterwards the human operator is trying to control the process in the way, how it is remembered by him as the solution in chosen case. It means the human operator is following the way, how he has set values of controllable signals (C) knowing measured values of dependent signals (D). This stage of experience using goes on till the end of current production day. When the current production is ended, the human operator obtains information concerning evaluation of made products. So, this time it is possible to update his experience with the case, which concerns just ended production day (in order to use it in the future).

A. Case-Based Reasoning

Presented model of using and gathering experience coincides with case-base reasoning (CBR) methodology. This methodology is relaying on experiences made in the past during solving of concrete problem situations, instead of using any general knowledge related to a problem domain [3], [4]. From a technical point of view a CBR decision system uses a case-base, which is collection of past made and stored experience items, called past cases, or cases. Each time a new problem has to be solved, a CBR cycle is performed, that consists of 4 sequential processes, which are called also phases:

1) Retrieve the most similar case or cases for a current problem that hast to be solved.
2) Reuse the information in the retrieved case in order to solve the current problem.
3) Revise the proposed solution.
4) Retain the experience (the current problem, its solution and results) in order to use it for future problem solving.

The extensive discussion related CBR methodology and its implementation in the domain of industrial process control can be found in [5].

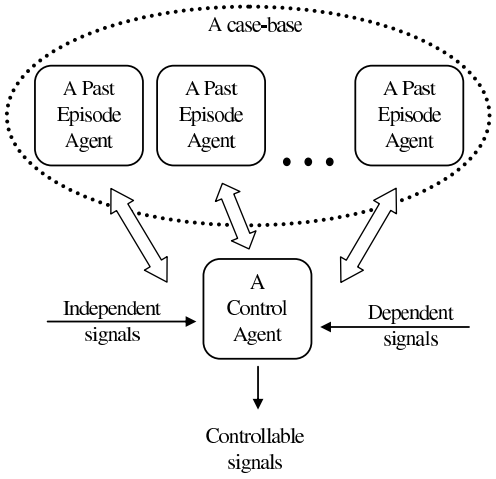


Fig. 2. The structure of proposed multi-agent system

IV. Design of Agent System Modeling Experience

Presented in the previous section assumptions indicate distribution of cases, that are autonomous items of human experience. The autonomy of experience items is the main reason for using of agent technology at design and implementation of this model. Considering autonomy of past cases, it is proposed to design the case-base as a set of autonomous agents, as presented onto Fig. 2. Every agent in this set, called Past Episode Agent should contain all data relating to represented by him episode – a past case of control of the industrial process. Presented onto Fig. 2 Control Agent has a goal to control current production process. The Control Agents communicates with all Past Episode Agents in order to follow processes of experience usage and gathering, which are characteristic for human operator. Referring to CBR methodology, the Control Agent should perform all four phases of the CBR cycle. All interactions in the system are initiated by the Control Agent, so this agent performs also (centralized) management of the whole system. Presented here conception of system does not relate to situational systems.

A. Past Episode Agent

According to the main assumption, that a Past Episode Agent should represent one past case, an agent of this type has to contain data structures related to notion of a past case. A past case should enclose information according solved problem, used solution and evaluation of made products for one past production day. Every Past Episode Agent has to contain:

- single values of independent signals (I) for the whole considered production day (as problem description of represented past case),
- array of values of dependent (D) and controllable (C) signals registered during considered production day (as solution description of represented past case),
- single value of quality evaluation (Q), that is average concentration of sulphide sulphur in made products.

A Past Episode Agent having such data structures filled with proper data models one episode of past production. A Past Episode Agent replays to messages sent by a Control Agents providing information according to stored data.

B. Control Agent

The main goal of a Control Agent is to control current production process. The functioning of the Control Agent starts at the beginning of a current production day. That time values of independent signals (I) characterizing chemical composition of raw materials used for production at the whole current day are known. After the start of its functioning, the Control Agent knows independent signals and starts to sequentially perform phases of CBR cycle: retrieve, reuse, revise, retain.

The main goal of the **retrieve phase** is to find a past case, which concerns similar problem to the current problem of control and contained in this case solution is evaluated as desirable. This goal is realized through below presented interaction between agents:

1) The Control Agent sends request of replay containing values of independent signals (I). This request is sent to all Past Episode Agents existing in the system.
2) Every Past Episode Agent replays to that request by sending back his values of independent signal and the Control Agent receives all replays concerning independent signals.
3) The Control Agent chooses a number of Past Episode Agents representing similar values of independent signals (I). This similarity is based on the Euclidean distance between values of independent signals measured for the current problem and the solved problems represented by Past Episode Agents.
4) The Control Agent sends request of replay containing value of quality evaluation (Q), that is average concentration of sulphide sulphur in the products. This request is sent only to previously chosen Past Episode Agents.
5) Past Episode Agents replay to request by sending back proper value, which indicates evaluation of production. The Control Agent receives all replays concerning evaluation of production. The Control Agent chooses one Past Episode Agent, which represents the best evaluation of made products (what means the smallest value of average concentration of sulphide sulphur in the products).

Execution of above presented interaction scenario results in selecting of one Past Episode Agent that represents a past case used at next phase of CBR cycle.

In the **reuse phase** the solution represented by the previously selected Past Episode Agent should be applied to the current problem of control. The past case contains description of solution in the form of values of dependent signals (D) and values of controllable signals (C). An artificial neuron net can be used to model relation between dependent signals (D) and controllable signals (C). The Control Agent in the reuse phase follows below stated steps:

1) The Control Agent sends to chosen Past Episode Agent request of replay containing array of values of dependent (D) and controllable (C) signals.
2) After receiving of replay, the Control Agent models relation between dependent signals (D) and controllable signals (C). This step the artificial neuron net is created and learned.
3) The Control Agent obtains current values of dependent signals (D) and on this basis predicts values of controllable signals (C) with the help of the learned neuron net. The predicted values of controllable signals should be applied in current process control. This step is repeated till the end of current production day (the reuse phase continues till the end of current production day).

The Control Agent during the **revise phase** obtains evaluation of products made during current production day. This evaluation is in the form of single value of average concentration of sulphide sulphur in the products. Result of the evaluation cannot influence control done by this agent, because production was ended before quality measures.

The **retain phase** enables learning in the CBR cycle, what is analogy to experience gathering by a human operator. This phase starts, when the current problem was solved and evaluation of this solution is known. The current case contains already the description of the problem, description of the applied solution and the evaluation. The goal of Control Agent is now to retain this information by adding a past case that relates to just ended production day. Because every past case is represented by a Past Episode Agent, the Control Agent creates new agent of Past Episode Agent type.

V. System Implementation and Testing

Presented in the previous section description of a system was implemented as a test application, that operates on archival data of an industrial plant. On this stage of research it was impossible to deploy the application in order to influence real production. The lack of mentioned deployment disables to obtain evaluation of real products made under control of developed system, what is the reason for implementation problems. Those problems relate to the revise and retain phase, which require the real evaluation of products. Despite mentioned problems the whole system gives solution for control of the current production day.

Implemented Control Agents uses a neural network, which is a multilayered perceptron, composed of neurons with sigmoid function. All neurons are located in 4 layers composed of 9, 13, 11, 7 neurons. By the modeling step a supervised learning is used. Developed system works in the batch mode and was implemented using Java, JADE and Neuroph.

A. Performed Tests

Performed tests were done with real industrial data, but only 19 days of production were available (every day is full production cycle). The available industrial data was transformed to the case-base of presented system. In the result 19 Past Episode Agent were created (every individual agent represents one past case).

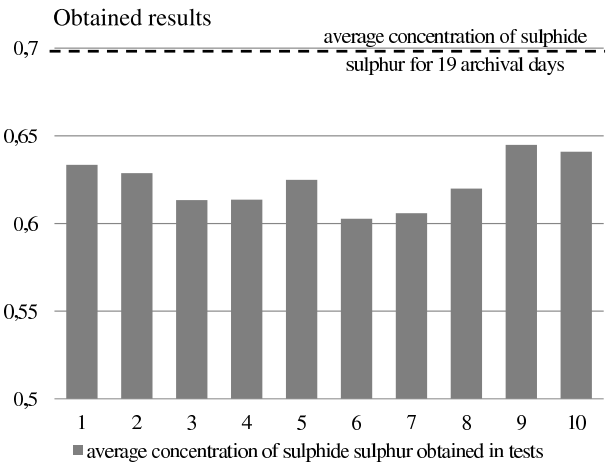


Fig. 3. Results obtained for 10 runs of the implemented system

Presented onto Fig. 3 results concern 10 tests of developed system. Every test relates to one production day, that is full production cycle. The quality measure for performed test was evaluated with the use of an external application, which predicts average concentration of sulphide sulphur in products made during considered production day. As it is shown onto Fig. 3 developed system enabled to obtain more optimal control for every test day, than is was done manually for 19 archival days, because the aim is to minimize concentration of sulphide sulphur in the products.

VI. Conclusions

Presented work shows that it is possible to design and implement an agent system, which follows mechanisms of experience using and gathering. Those mechanisms are consistent with case-base reasoning (CBR) approach and can be used as mechanisms for control of chosen industrial process. Made tests of presented system appoint, that developed system good reflects analyzed mechanisms and obtained results are estimated better than production, which was done in the past by a human worker.

References

[1] K. M. Passino, "Intelligent Control: An Overview of Techniques" in T. Samad, Ed., *Perspectives in Control Engineering: Technologies, Applications and New Directions*, IEEE Press, NY, pp. 104-133, 2001
[2] Ł. Sztangret, Ł. Rauch, J. Kusiak, P. Jarosz and S. Małecki, "Modelling of the Oxidizing Roasting Process of Sulphide Zinc Concentrates Using the Artificial Neural Networks", *Computer Methods in Materials Science*, vol. 11, pp. 122–127, 2011
[3] A. Aamodt and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches", *AICom – Artificial Intelligence Communications*, vol. 7, pp. 39–59, 1994,
[4] R. Bergmann, K. D. Althoff, M. Minor, M. Reichle and K.Bach, "Case-Based Reasoning – Introduction and Recent Developments", *Kunstliche Intelligenz: Special Issue on Case-Based Reasoning*, vol. 23, pp. 5–11, 2009
[5] G. Rojek and J. Kusiak, "Case-based Reasoning Approach to Control of Industrial Processes", *Computer Methods in Materials Science*, vol. 12, pp.250–258, 2012

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1041–1046

Agent-based Architecture and Situation-based Scenario for Consistency Management

Thao Phuong Pham
L3i Laboratory
University of La Rochelle, France
Email: phuong-thao.pham@univ-lr.fr

Mourad Rabah
L3i Laboratory
University of La Rochelle, France
Email: mourad.rabah@univ-lr.fr

Pascal Estraillier
L3i Laboratory
University of La Rochelle, France
Email: pascal.estraillier@univ-lr.fr

***Abstract*—During interactions, system actors may face up to misunderstandings when their local visions contain inconsistent data about a same fact. Misunderstandings in interaction are likely to reduce interactivity performances (deviation or deadlock) or even affect overall system behavior. In this paper, we present agent-based architecture and scenario-structuring approach to deal with such misunderstandings and consistency. It is based on the notion of "situation" that is an elementary building block dividing the interactions between actors into contextual scenes. This model not only supports the scenario execution, but the consistency management as well. In order to organize and control the interactions, a "situation" contextualizes system's actors' interaction and activity, and includes prevention and tolerances mechanisms to deal with the misunderstandings and their causes. We also simulation experimentation on an Online Distance Learning case study.**

***Keywords*—Interactive system; adaptation; misunderstanding; situation- based scenario; consistency management**

I. Introduction

IN INTERACTIVE SYSTEMS, as games and simulators, the users and the internal agents can modify systems content and progress in real time through input adjustments. The interactive systems may adapt the system execution not only to user's actions, but also to user's profile and behavior, making these systems adaptive... In order to perform the adaptativity, the system must capture users' behaviors from their interactions. Then, according to system's logic and designer's logic, the system adjusts its execution to what it perceives of user's logic. Du to user's actions unpredictability, the execution process of an interactive system is also not predictable.

One of the important problems in interactive system is the potential misunderstanding between the users and the system and more generally between system's actors, virtual or physical. If the system does not capture correctly or confuses user's actions, or if the users do not understand what the system expects, that may lead to an erroneous interpretation of their behavior and an erroneous adaptation of system execution. This misunderstanding may concern user-system interactions, but it can also appear in any kind of interaction between any system's actors. It can be due to the incomplete actors' data or the non-determinism of actors' behavior and cause the interaction deadlock or application failure.

In our recent works, [1], [2], we have defined the misunderstanding in interaction as: when two or more system's actors have incoherent data in their local visions about the same fact f and these data is used during their interaction, that can cause an interaction deviation from the planned scenario. An actor may be human user or virtual system's agent. The local vision is actor's own knowledge about its external world (virtual environment, system's resources...), its relations with the others actors (subset of their states) and its own profile (internal state). So, our work focuses on the management of the consistency between the actor's behaviour logic and the system's logic and the consistency between the actor's local visions in order to handle the potential misunderstandings in interactions.

To handle misunderstandings we propose to contextually structure the application execution into interaction sequences called "situations" and including misunderstanding prevention and tolerance mechanisms. Each situation corresponds to a contextual resource-centered sequence of activities and events and is characterized by preconditions and postconditions. That allows the system to control the execution and to establish the casual links between the situations. This model confines actor's interactions in a given context in order to control them and manage the execution consistency. The consistency handling mechanisms, are inspired by techniques from dependability domain since there is an analogy between the misunderstandings in interactive systems and the errors handled in fault tolerant systems [7].

II. Misunderstanding in Interactions and Related Work

In the recent research, we can find several works dealing with the user-system dialogue where the communication is done through a real human language [3]–[5]. According to Rapaport [5], negotiation is the key to understanding. A cognitive agent understands by negotiating with the interlocutor or by hypothesizing the meaning of an unknown word from the context... A cognitive agent can negotiate with itself about something external by comparing it perception and internal knowledge in order to change or correct its own misunderstandings. Other works propose to use confidence scores to measure the reliability of each word in a recognized sentence [6]. Besides, Lopez-Cozar proposed to implement a frame correction module, which is independent of speech recognizer [4]. This module corrects misunderstandings in a sentence, caused by the errors in speech recognition, by replacing the incorrect frame with an adequate one. Karsenty and Botherel applied the adaptable and adaptive transparency strategies to TRAVELS project with the goal of helping the users to understand and

978-83-60810-53-8/$25.00

react appropriately to system rejections and misunderstandings [3]. The ability of making system's interpretations explicit and informing the users on how to correct misunderstandings are two ways to help users handle them. This strategy is very effective in misunderstanding detection and raises the rate of appropriate user responses after system rejections. All of these works deal with the problem in speech dialogue where the misunderstandings are the more frequent. But the misunderstanding can be found in other forms of interaction like actions, gesture...

Our purpose is to define how we can treat the misunderstandings between the actors themselves, besides the user-system misunderstandings. It is not easy to recognize such class of misunderstandings. In the dependability domain [7], we find the inconsistency problem between systems and operators. The "automation surprise" is inconsistency error occurring when the system behaves differently than its operators expect [8]. It may be due to a mismatch between the actual system behavior and the operator's mental model of that behavior [9], and it can lead to "confusion mode" and sometimes critical failures. In general, misunderstandings come from the gap between user's logic and designer's logic, all along action planning between the actors. Many works, particularly in interactive storytelling, have been done to solve the mismatch between users' behaviors and system logic [10]–[12] by predicting the user's future actions and detecting the invalid ones that deviate the execution from the planned objectives. In general, prediction approach cost is very expensive, such as a short-term player behavior modeling module implanted in [10] to simulate how the world would change to player's actions. Moreover, this approach seems not well suited to a real-time interactive systems, nor to systems in which user's behaviors cannot be modeled easily by a set of rules.

Our approach will focus on software and component design model to integrate simple prevention and treatment mechanisms. Our solution relies on three points. *First*, we build robust agent-based architecture with specific additional components in charge of misunderstandings in interaction management. *Second*, we organize the actors' interactions as a situation-based scenario to facilate the interaction control. *Third*, we integrate into situations' dynamic execution the consistency management, including data synchronization, misunderstanding detection and treatment inspired and adapted from fault-tolerance techniques. These mechanisms do not try to predict users' behavior but will take into account users' state to adapt the system execution in order to avoid misunderstandings between actors. The system observes and analyzes users' states, detect the misunderstandings or their consequences and act to keep the consistency between actors' logics at the beginning and at the end of interaction sequences.

III. Agent-based General Architecture for Interactive Adaptive System

Several architecture models for interactive systems have been proposed according to the specific purpose of each work. We chose the approach of multi-agent system in [16] as a starting point to build our model. The advantage of this approach is that each agent can be organized and work autonomously and strategically. We added a special agent called script agent besides the adaptation unit to manage the consistency. Figure 1 shows our overall architecture.

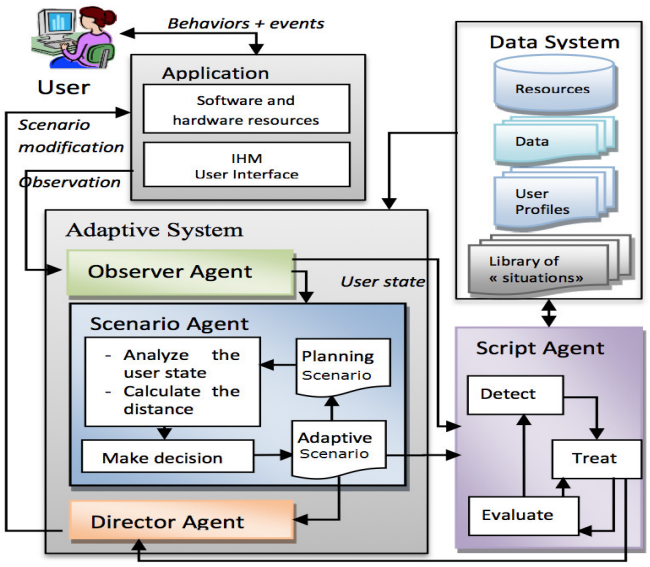


Fig. 1: General agent-based architecture for interactive system

Observer agent: It observes user's behaviors and state, formalizes, normalizes and transfers them to the scenario agent.

Scenario agent: It makes decisions about scenario orientation according to user's state, planned scenario and permanent objective defined by the designer. This agent tries to find the best way to orientate the application execution. Scenario agent takes charge of a library of "situations" planned by the designer. These "situations" (defined in section IV) represent scenario components and are the interaction and the activity sequences that can take place in the application as, for instance, all possible scenes in a theater play.

Director agent: This agent receives the decision taken by the scenario agent. He takes in charge the production of the adaptive scenario and realizes a modification, an answer or an action adapted to the users.

Script agent: Its task is to track inconsistency in 3 steps:

- *Detection*: Detect, confine or partition the inconsistency between situation's actors in order to identify the causes of the misunderstanding.
- *Treatment*: Apply the handling mechanism or strategy to remove the inconsistency and to correct the deflected state that causes the incoherence.
- *Evaluation*: Estimate the efficiency of the treatments in order to improve the applied mechanism for the next time.

IV. Situation-based scenario

A. Approach of interactive storytelling

Our proposition is inspired from the interactive storytelling domain that focuses on scenario execution management.

Interactive storytelling is the unfolding of a story that the player's decisions impact [13], [17]. It also refers to how to generate stories which are both interesting and coherent. We consider that the interactions in an interactive application can be organized, strongly or weakly, as a story scenario. That allows us to adapt ideas from storytelling domain to organizing the interactions.

The scenario in interactive storytelling is represented by a series of actions/events linked together by cause and effect as in [14] or by ordered link as in [10], [15] or by Hierarchical Task Network planning as in [12] where each task is decomposed into subtasks until the primitive actions. But all of these scenario structurings are not suited to built complex interaction sequences where the user's actions are free, non predictable and depending on a great amount of context data. Hence, we propose the notion of "situation" that can be seen as a scene encompassing not only interactions execution but also interactions management and resources use. The situations are the basic narrative elements that facilitate interactions' planning and management by characterizing, contextualizing and confining them.

B. Scenario organizing with situations

1) Situation model: The interactions are split into a set of situations. Each situation is a sequence of interactions between two or more actors in a precise context to achieve a predictive objective, as shows the figure 2. It is characterized by: the preconditions, the postconditions, a set of participating actors and a set of resources. Due to the fact that actors' behaviors, especially human behaviors, are not always precisely modeled, and due to the influence of external events, the progression of a situation can be considered as an execution and adaptation "black box" where the interactions are executed in a non-predictable way. Furthermore, the situation includes consistency management. Itrepresents a set of mechanisms devoted to the prevention, detection and treatment solutions, in order to redress and adjust situation's progression in spite of misunderstanding and inconsistency problems. Consistency management is carried out all along the situation progression

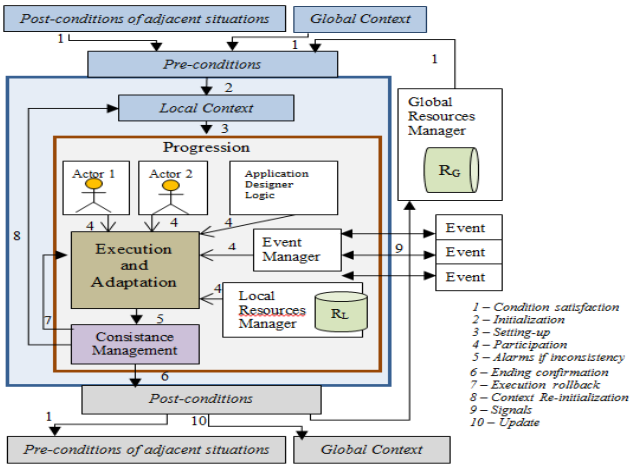


Fig. 2: Elementary Situation Structure

from the local context initialization to the post-condition completion.

2) Situation Graph and Application Execution: The situations are considered as the plot structuring elementary blocks. Each application provide a set of situations defining all the possible interaction sequences that can happen during the application execution. They can be grouped and linked together in order to build the overall application scenario. The scenario is then represented by a directed graph of situations. Each node is a situation and each edge is a transition from one situation to another. The situations graph shows the causal relationships between scenario situations. A scenario may have several beginnings and also some possible endings.

The situation-based scenario approach favorizes the execution control and interaction adaptation. The application progression becomes a scenario unfolding from one starting node to one final node on the predefined situation graph (it is taken in charge by the *scenario agent* in the global architecture). When there is more than one possible situation, the most pertinent one will be chosen by the scenario agent. To increase the adaptability, we can avoid the definition of a predefined graph. In that case, the situation choice is made according to the pre-conditions that best satisfy the global state and decision criteria. This method is flexible, adaptive, and applicable in "real time" during application execution, but it can lead to uncontrollable situation order or infinite loop, if the post-conditions and pre-conditions do not contain sufficient data.

V. Consistency Management Model within Situations

A. Handling Mechanisms

The *consistency management* that we propose consists of a set of specific methods, techniques and mechanisms that aim to handle the misunderstanding problem and to obtain data consistency all along the interactions. They are similar to the dependability techniques [7].

1) Prevention mechanisms: try to suppress misunderstandings occurrence conditions in order to avoid misunderstandings. To avoid data inconsistency, the proposed technique is the explicit declaration of all shared data before situation's interaction sequence start. It aims to identify and share actors' local visions in order to decrease the possibility of interaction deviation. Once the actors have collected the necessary data, they can start the interactions. The data synchronization is another method intending to compare the actors' local visions at a given moment during the interaction sequence in order to avoid the inconsistency of new perceived data. The synchronization can delay the interactions, so it should be done fast and not too frequently to disturb them as less as possible.

2) Tolerance mechanisms: aim to assure interaction continuation despite misunderstanding occurrence via misunderstanding detection and interaction recovery.

Detection: regular check of i) the shared data used during the interactions and ii) the deviation between actors' logics.

Recovery: once a misunderstanding is detected, the system apply one or several of the following techniques : **rollback**

- bringing the system back to a stable state, exempt from misunderstanding, to retry the interactions; **rollforward** - bringing the system to a new misunderstanding free state from which the interactions can go on; **reinforcement** - requiring from one or from all participant actors to do some additional interactions.

3) Removal mechanisms: involve misunderstanding detection and correction, followed by **reinitialisation** of the concerned interaction sequence, or of the whole execution process. The detected misunderstandings will be diagnosed to determine their causes: which data are inconsistent? Which ambiguities exist in the interaction context? Are there protocol faults? After that, an appropriate correction method will be applied to eliminate the related misunderstanding. Finally, the interactions have to be restarted from the last stable point or from the beginning.

B. Inclusion into the Situation Structure

Our situation-based architecture allows the integration of misunderstanding management mechanisms inside the situation in order to control the misunderstandings and their consequences all along situation execution. We define three phases (figure 3).

a) Prologue phase: The explicit declarations of interacting content and data are performed, to synchronize actors' local visions before they start to interact. If the initial data of all actors are identical from the beginning, the possibility of misunderstanding is reduced. If the inconsistency exists, a negotiation step will be performed between the inconsistent actors. Then, one or several of them will modify its/their data, or the divergent data will be isolated/removed and not considered during the interactions.

b) Interaction or Dialogue phase: when the interactions are carried out, the actors will update their local data step by step, as they continuously observe and perceive each other. Despite the initial local vision agreement, misunderstanding may nevertheless occur during the interactions. This is why their local knowledge is synchronized all along the interaction sequence in order to avoid that local data about same facts diverge in actors' local visions. One or several techniques of *reinforcement, rollback, rollforward* can be alternatively used.

c) Epilogue phase: All the interactions are done in the previous phase. If the post-conditions are fulfilled, we can exit the situation with the expected results. But if, for some reason, we do not reach the expected post-condition, the **script agent** has to detect and settle the existing incoherency in order to avoid the propagation of the misunderstandings to other situations. The system may also require actors to do some reinforcing interactions, or if necessary, make a rollback to a last stable state (in this case there must be systematic state saving mechanism), or, even a restart of the whole situation. The main goal of this phase is to quit the situation with the appropriate post-conditions and without latent or active misunderstanding. But, the rollback or reinforcing interactions may not lead the actors towards the planned post-conditions. Therefore, we add in the situation model a special exit point called "exception" that allows the current situation to be stopped at anytime without expected post-conditions and that leads to **exception handling situations**.

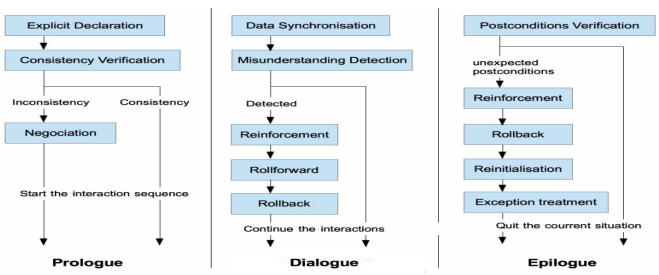


Fig. 3: Consistency management mechanisms

VI. Experiment on Online Distance Learning Case Study

To validate our approach we applied our situation-based methodology in our current online distance learning (ODL) project [2]. The project is devoted to the development of an online distributed platform that simulates a real classroom: teachers and learners carry out learning sessions as in a real life but by interacting through a virtual class environment. The platform integrates an interactive numeric board, camera, microphone and pedagogic tools (as file sharing system or virtual notebook) to support the courses...The figure 4 shows an example of courses scenario based on 6 situations. However, the users may face many difficulties: class supervision, course quality assessment, misunderstandings due to the weak system's interfaces and mechanisms to catch and manage user behaviors. The interactions between the actors in ODL contains numerous factors that may lead to misunderstandings as: multi-meaning or implicit behaviors; supervision tools' observation and interpretation imperfection; system component failures; incomplete, missing, implicit or wrong consigns...

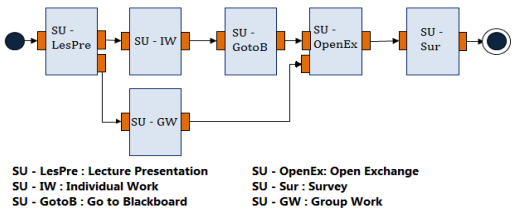


Fig. 4: Situation-based scenario example

A. "Individual Work" Situation Description

To deal with these various misunderstandings, we applied our situation-based solution including consistency management to a particular situation: "Individual Work" (SU - IW in figure 4). Each learner will work individually and has to do the exercises distributed by the system. The system provides additional exercises each time the learners send the previous exercises report. The expected post-condition is that all the learners reach a required knowledge level "*MaxKnowledge*".

Because of the long test duration and development for the real platform prototype, we chose to experiment our misunderstanding management mechanisms and agent-based architecture through a multi-agent simulation with the GAMA

platform[1]. We have 4 types of agents : "Teacher", "Learner", "Observer" and "ODL System". The Observer's role is to observe the state of sent exercises in order to evaluate learners' accumulated knowledge level. The distribution mechanism based on these observations and learners' skill level evaluations is taken in charge by "ODL System" agent that is a combinaison of 3 other agents in our model: scenario, script and director agent (figure 1). Potential misunderstandings in this situation occur when the system distributes the exercises that are incoherent given the learners' skill and expectation. They can result from wrong learners' exercise state observation or from inappropriate distributed exercise level. The misunderstanding handling is done inside the situation during its 3-phase progression (figure 5).

Prologue phase: The system checks each learner's connection status to begin the exercise series distribution.

Dialogue phase: In this situation, the interactions content refers to the exercises distribution and reporting. During learners' work, each observer agent supervises his associated learner's working state and his exercise report to collect data: partial or total termination, work duration, correctness rate. To avoid the wrong estimation of learner's skill and knowledge level, these data have to be synchronized between the observer and his learner after the exercise report is sent and before a new exercise is distributed.

Epilogue phase: To finish the situation the lerners must reach a given skill level after a given number of exercises. If a learner reaches this number without reaching the required skill level, the series will be stopped after a *session deadline* to avoid an abnormal long series. The system sends a *StopSignal* message to all learners to confirm the end of the exercise series after a predefined timeout. It refers to the exception treatment.

B. Experimentation Results

We run the simulation of "Individual Work" situation with the following parameters: 50 learners, 1 teacher, max knowledge level = 25, max difficulty level = 20, session deadline = 250 steps of simulation. We will measure a set of

[1]https://code.google.com/p/gama-platform/

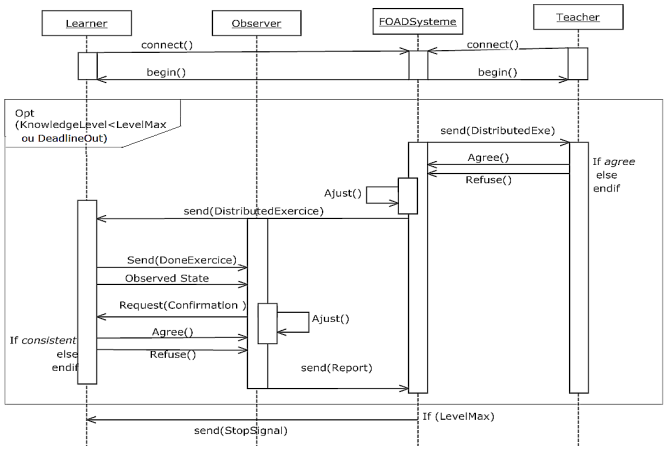
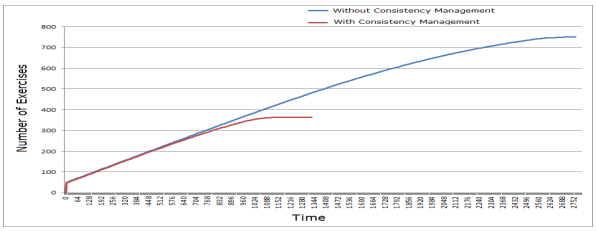


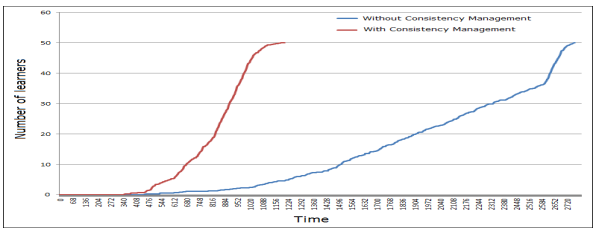
Fig. 5: Agents main interactions in the simulation

important factors influenced by potential misunderstandings:
N_e: total number of distributed exercises;
N_{notend}: total number of real non-finished exercises;
N_{bad}: number of bad observation by all observers;
N_{cor}: number of system observation corrections while detecting the wrong observed states (it refers to the synchroniaation times where consistency management is performed to remove inconherent data);
LI: learners' interest level that increases when the learners succeed and that decreases when they fail their exercises;
T_{total}: total session times (in *steps*) until the last learner has finished his series.

The data are recorded and calculated for the average values from 10 simulations lauching times in each measure. We compare these data between two cases: "with" and "without" the consistency management. The results are summarized in the table I. The total distributed exercises number N_e is twice more in "without" case compared to the "with" case. The average number of not finished exercises in "without" series is higher than in "with" series: 747.4 vs 363.4 also depicted in the figure 6(a). It is obvious that the session duration in "without" case is almost 2 times longer than in "with" case.



(a) Number of distributed exercices.



(b) Number of learners finishing exercise series.

Fig. 6: Comparason between 2 cases "With" et "Without".

The figure 6(b) shows the number of learners that have finished their whole series during the situation execution in the "with" and "without" consistency management cases. The lines shows that the learners work with more exercises and with longer duration T_{total} in the "without" case. We can make the same observation with the average measure values in table I.

Why do we have this difference result? When the consistency management is integrated in the situation execution to handle the potential misunderstandings, the observers have to adjust their observed data according to learners' "disagree" acknowledgments. Hence, the learner's skill level estimation will converge faster to the real value, and the difficulty level of the distributed exercises is more appropriate to his skill. The result is that learners can finish all the exercises and with higher correctness rate. In contrast, if no mechanism is added to control the inconsistency between learners and observers,

TABLE I: Statistical data comparason between 2 cases: With (Wi) et Without (Wo) the consistency management

	N_e		N_{notend}		N_{bad}		N_{obsnon}		N_{cor}		LI		T_{total}	
	Wi	Wo	Wi	Wo	Wi	Wo	Wi	Wo	Wi	Wo	Wi	Wo	Wi	Wo
1	330	735	23	64	83	103	93	114	83	0	78.98	66.1	988	2692
2	363	692	45	28	87	76	110	88	87	0	77.73	76.41	1104	2640
3	361	744	44	55	94	97	114	114	94	0	76.35	69.06	1076	2700
4	383	768	47	73	110	99	129	118	110	0	77.31	65.18	1160	2724
5	347	744	32	60	87	99	109	115	87	0	77.94	67.31	1024	2688
6	360	806	37	65	111	117	122	135	111	0	77.55	64.68	1108	2760
7	392	737	66	59	93	92	118	108	93	0	73.06	66.88	1188	2692
8	379	752	42	66	117	100	131	112	117	0	77.65	65.88	1140	2712
9	353	722	37	69	96	100	111	111	96	0	77.49	67.55	1048	2672
10	361	774	40	62	93	115	117	134	93	0	74.18	65.92	1084	2728
Ave.	363.4	747.4	42.4	60.1	97.1	99.8	115.9	114.9	17.8	0	76.82	67.71	1092	2641

a non-finished exercise can be perceived as finished, and vice versa. The skill estimation is less correct: higher or lower than the real one. There is a higher probability that the ODL system gives to the learners too difficult or too easy exercises. That delays the skill level progression and explains why the learners take more time to terminate the series.

VII. Conclusion

In this paper, we have presented the situation-based design methodology and consistency management mechanisms to handle the misunderstanding in interactions. Our approach is to contextualize the interactions between actors into "situations" and add to these basic narrative blocks consistency management mechanisms split into 3 steps: the *prologue*, data declaration and consistency verification, the *dialogue*, the interaction unfolding, local visions synchronization and misunderstanding treatment, and the *epilogue*, data update and agreement attainment. We do not seek to find a universal algorithm or a solution to deal with all types of misunderstandings in interaction. Our aim is to provide a management pattern that could be systematically used by the application designers or developers and that allow them to incorporate their own verification, synchronization, prevention and tolerance mechanisms adapted to the specific misunderstandings of their applications.

We have applied our methodology to a case study from an Online Distant Learning project. We have built a simulation of the "Individual Work" situation and integrated into it the proposed solutions to show how the consistency management operates on a simulation example. From the experimentation results, we have found out that our mechanisms reduce the incoherent data between learners and observers and improve the performance of exercise distribution: shorter session duration, lower exercise number, faster required level attainment... Even if the simulation is simple and does not cover exhaustively all the possible interactions that can occur in such situation, it illustrates the benefits of misunderstanding management during interaction progression.

The next step of our work is to perform the same experimentation and measures on the prototype under development with live case study and to check the relevance of our approach in other e-Learning "situations".

References

[1] P. T. Pham, M. Rabah and P. Estraillier, "Handling the Misunderstanding in Interactions : Definition and Solution," *in The Annual Int. Conf. on Software Engineering & Applications SEA 2011,* 2011, pp. 47–52.

[2] F. Trillaud, P. T. Pham, M. Rabah, P. Estraillier, and J. Malki, "Online Distant Learning Using Situation-based Scenario," *in the Int. Conf. on Computer Supported Education CSEDU2012,* 2012.

[3] L. Karsenty and V. Botherel, "Transparency strategies to help users handle system errors," *in Speech Communication*, vol. 45, no. 3, Mar. 2005, pp. 305–324.

[4] R. Lopez-cozar, Z. Callejas, N. Abalos, G. Espejo, and D. Griol, "Using Knowledge about Misunderstandings," *in Speech Communication,* 2010, pp. 523–530.

[5] W. J. Rapaport, "What Did You Mean by That? Misunderstanding, Negotiation, and Syntactic Semantics," *in Journal Minds and Machines,* 2000, pp. 397–427.

[6] H. Jiang, "Confidence Measures for Speech Recognition: A survey," *in Speech Communication*, vol. 45, no. 5, 2005, pp. 455–470.

[7] J. C. Laprie, B. Randell, C. Landwehr, and S. Member, "Basic Concepts and Taxonomy of Dependable and Secure Computing," *in IEEE Transactions on Dependable and Secure Computing,* vol. 1, no. 1, 2004, pp. 11–33.

[8] S. Combefis, P. S. Barbe, and C. Pecheur, "A Bisimulation-Based Approach to the Analysis of Human-Computer Interaction Categories and Subject Descriptors," *in EICS '09 Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems,* 2009, pp. 101–110.

[9] G. G. King, "General Aviation Training for "Automation Surprise" ," *in International Journal of Professional Aviation Training & Testing Research,* vol. 5, no. 1, 2011.

[10] B. Magerko and J. E. Laird, "Mediating the Tension between Plot and Interaction," *in AAAI Workshop Series: Challenges in Game Artificial Intelligence,* 2004.

[11] H. Barber and D. Kudenko, "Generation of Dilemma-based Interactive Narratives with a Changeable Story Goal," *in the 2nd Int. Conf. on INtelligent TEchnologies for interactive enterTAINment,* 2008.

[12] R. Paul, D. Charles, M. McNeill, and D. McSherry, "Adaptive Storytelling and Story Repair in a Dynamic Environment," *in The Fourth Int. Conf. on Interactive Digital Storytelling ICIDS,* 2011.

[13] R. Champagnat, G. Delmas, and M. Augeraud, "A Storytelling Model For Educational Games : Heros Interactive Journey," *in International Journal of Technology Enhanced Learning 2,* 2010, pp. 4–20.

[14] B. Karlsson, A.E.M. Ciarlini, B. Feijo, and A.L. Furtado, "Applying a Plan-Recognition / Plan-Generation Paradigm to Interactive Storytelling," *in Workshop on AI Planning for Computer Games and Synthetic Characters,* 2006.

[15] A. Silva, G. Raimundo, and A. Paiva, "Tell me that bit again...' Bringing interactivity to a virtual storyteller," *in Int. Conf. on Virtual Storytelling,* 2003, pp. 1–10.

[16] K. Sahaba, P. Estraillier, and D. Lambert, "Interactive educational games for autistic children with agent-based system," *in 4th Int. Conf. on Entertainment Computing (ICEC'05),* 2005, pp. 422–432.

[17] J. Lebowitz and C. Klug, "Interactive Storytelling for Video Games: A Player-Centered Approach for Creating Memorable Character and Stories", Focal Press, 2011

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1047–1052

Agent-based Resource Management in Tsunami Modeling

Alexander Vazhenin, Yutaka Watanobe, Kensaku Hayashi
Graduate School Department,
University of Aizu,
Aizu-Wakamatsu, 965-8580,Japan
Email: {vazhenin, yutaka, m5161111}@u-aizu.ac.jp

Michał Drozdowicz, Maria Ganzha, Marcin Paprzycki, Katarzyna Wasielewska
IBS PAN, Newelska 6, 01-447
Warszawa, Poland,
Email:{firstname.lastname}@ibspan.waw.pl

Paweł Gepner
Intel Corporation
Pipers Way
Swindon Wiltshire SN3 1RJ
United Kingdom
pawel.gepner@intel.com

***Abstract*—Complexity of tsunami modeling requires designing software system with high level of reusability and interoperability of its components, and flexible resource management. In this paper we investigate how to integrate the tsunami modeling software with an agent-based resource management infrastructure.**

I. Introduction

REFLECTIONS brought about by the Great Japanese Earthquake and Tsunami raise questions how to mitigate the impact of such events at different time scales, from real time tsunami warning, to long-term hazard assessment. This makes the Tsunami Modeling Problem even more important, and requires applying modern software design approaches, including collaboration among distributed clients and computational services. Another challenge is scalability of the approach, which should allow an arbitrary number of users and computational resources to interact in a customizable working environment. Finally, since separate applications and services may be developed for heterogeneous technologies and platforms, reusability and interoperability are important aspects of exposing and combining computational resources and services [1], [2].

Increasing power of personal computers allows their volunteer participation in high-performance computing, by granting computer time for public calculations. For example, the Folding@home project involves distributed computationally intensive simulations of protein folding and other molecular dynamics simulations [3], [4]. Each user can participate in these computations by calculating a part of a problem. Note that volunteer computing does not allow private use of results obtained by a community of users. Furthermore, this project (and many others) is based on applying a single computational kernel, using a BOINC-like client. Unfortunately, tsunami modeling requires more human interactions during the process, to understand the results and on their basis to specify the next round of experiments. Furthermore, as mentioned below, there exists a number of different tsunami models, that often need to be combined to model various phases of the tsunami phenomenon. This makes simple application of the BOIC-like volunteer approach unfeasible. For different reasons, it may not always be possible to use grid-like infrastructures. For instance, in a University with multiple laboratories "belonging to" separate administrators, it may not be easy to combine them into a single grid. This is especially the case when when different resources have different availability schedules.

In response to these challenges, the aim of this note is to outline how to combine tsunami modeling software with an agent-based resource management infrastructure. Here, agents' flexibility and ability for negotiations will allow malevolent use of resources belonging to different administrative domains.

II. State-of-the-art in Tsunami modeling

Let us start from a brief overview of the state-of-the-art in tsunami modeling. In [5], authors suggested that complex mathematical models and high mesh resolution should be used only when and where necessary. They proposed a "parallel hybrid tsunami simulator," based on mixing different models, methods and meshes. Their system was implemented implemented using object-oriented techniques, allowing for easy (re)use of existing codes and adding new ones. Note that the goal was to combine existing approaches to develop high quality hybrid models (rather than high performance).

Authors of [6] experimented with eight parallel tsunami simulators. They applied different programming models; e.g. thread based shared memory, distributed memory, and virtual shared memory. As a result of experiments it was shown that the threading-based approach does not scale well, especially if sufficient "node memory" is not available.

The TsunamiClaw package was developed on the basis of the finite volume method [7]. Here, the solution is represented as piecewise constant and is approximated in discrete computational grid cells. The obtained solution is represented in the form of water depth and momentum. Currently, this project is no longer being actively developed. Instead it has been generalized into the GeoClaw software [24].

The TUNAMI–N2 software [8], uses a hybrid tsunami model with different approaches to deep sea and shallow water / dry land modeling. However, it applies constant grid size in the entire domain. The TUNAMI was developed by Imamura in 1993 by Imamura. The package was written in FORTRAN and had a standard GUI for interaction.

978-83-60810-53-8/$25.00

The Method of Splitting Tsunami (MOST; [9], [10]) was developed at the NOAA (Seattle, USA). It allows real-time tsunami effect forecasting, as it can incorporate data from detection buoys. Furthermore, in the US, the MOST model is used to create inundation maps [11]. Recently, MOST software was combined into a SOA system [1] available through a web enabled interface (ComMIT; [25]).

Summarizing, a number of models exist for tsunami risk mitigation. Typically, they involve: origins of tsunamigenic earthquakes (estimation of magnitude and epicenter location), determination of the initial displacement of the tsunami source, wave propagation, inundation to the dry land, etc. Typical tsunami modeling environment simulates three processes: (1) estimation of residual displacement area, resulting from an earthquake and causing the tsunami, (2) transoceanic propagation of the tsunami through the deep water zones, and (3) contact with the land (run-up and inundation).

III. University of Aizu Tsunami Model

A. Basic Model and Software Tools

Currently, a novel tsunami modeling system is being developed at the University of Aizu. It is based on the principles of the Service-Oriented Architecture (SOA) and follows a Virtual Model-View-Controller pattern (VMVC). The VMVC is an adaptation of the traditional MVC to the SOA ([13]). The starting point for this work was the MOST package ([9], [10]). This method was initially developed in the Tsunami Laboratory of the Computing Center of the USSR Academy of Sciences in Novosibirsk. Subsequently, it was updated in the National Center for Tsunami Research (NCTR, Seattle, USA), and adapted to the models and standards used by the tsunami watch services in the US (and other countries). Here, the propagation of the long wave in the ocean is governed by shallow-water differential equations:

$$\begin{aligned} H_t + (uH)_x + (vH)_y &= 0, \\ u_t + uu_x + vu_y + gH_x &= gD_x, \\ v_t + uv_x + vv_y + gH_y &= gD_y, \end{aligned} \tag{1}$$

where $H(x,y,t) = h(x,y,t) + D(x,y,t)$; h represents water surface displacement, D depth, g gravity, $u(x,y,t)$ and $v(x,y,t)$ are the velocity components along the x and the y axis. Initial conditions should confirm the presence of water in all grid points, except for the tsunami source, where the surface displacement is not equal to zero.

The numerical algorithm is based on splitting in spatial directions the difference scheme, which approximates equations (1). This transforms solution of equations with two space variables to the solution of two one-dimensional equations and allows use of effective finite-difference schemes developed for one-dimensional problems. Moreover, this method permits to set boundary conditions for a finite-difference boundary value problem using a characteristic line method.

B. General Calculation Process

Figure 1 presents the block-diagram summarizing the calculation process. The input data consists of:

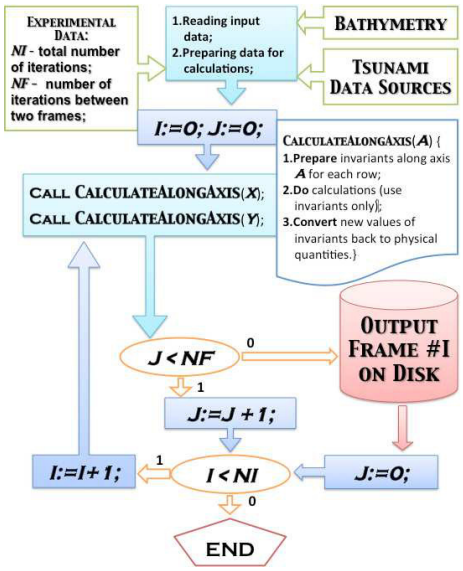


Fig. 1. General Computational Scheme

- bottom topography or bathymetry data,
- initial and boundary conditions,
- other parameters, e.g. time-steps and length of model run.

While running the program implements stores results as a series of frames representing the tsunami propagation process. Parameter NF defines the time interval, during which results are persisted in memory. After this time expires, results of current iteration, containing wave parameters, are stored on the secondary storage devices in the NetCDF format ([11]). The NetCDF format supports the creation, access, and sharing of array-oriented scientific data, while special programs allow its analysis and visualization.

While the original MOST software was implemented in Fortran 90, it was later ported to C/C++. Now, it takes about 3.00 seconds for a single time step on a 4 dual-core (Intel Xeon 2.8GHz) CPUs computer ([13], [14]). Observe that, a typical simulation, consists of about 10000 time steps (8 hours to complete). Therefore, the tsunami modeling needs to be significantly accelerated; especially for real-time tsunami warning guidance. However, speeding up modeling is also crucial for repetitive tsunami simulations.

IV. Service-oriented Architecture and Application Engines

A. Virtual MVC-design Patterns

Variety of methods for tsunami modeling require effective use of heterogeneous components on a variety of platforms and architectures. Furthermore, to achieve reusability, it is more cost effective to integrate applications rather than to rebuild them. Therefore the Virtual MVC design pattern (VMVC) was applied (Figure2). The demarcation of a Functional (View) and an Implementation (Model) task can be achieved by inducing an Integrator (Controller). The Controller can be enriched by encapsulating certain Non-Functional activities

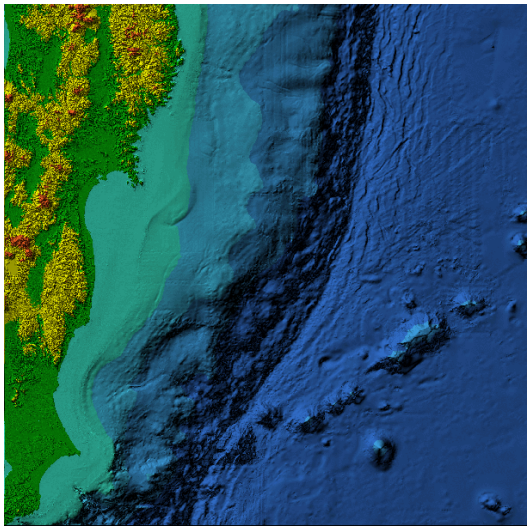


Fig. 3. Visualization of the 2413x2405 gridded relief around NE coast of the Honshu island.

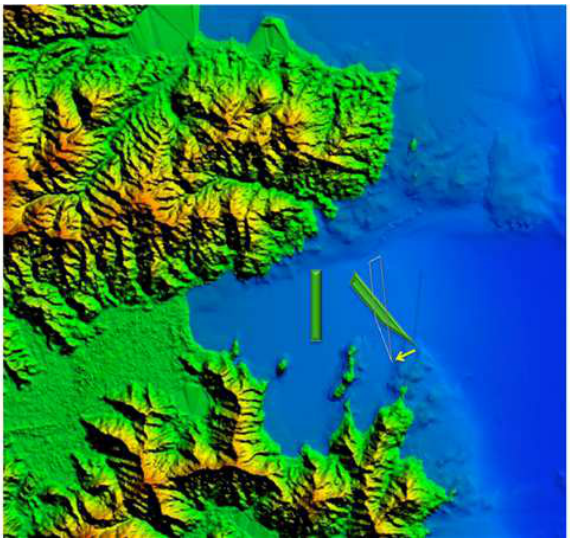


Fig. 4. Bathymetry with artificial objects.

such as security, reliability, scalability, and routing. This enables the separation of Integration Logic from that of Functional Logic (Client Application) and Implementation Logic (Service). The extendable set of application-oriented systems can be realized based on this integration environment including Service-Oriented Multistage Tsunami Modeling [13]. Figure2 shows main elements of this architecture, where the Tsunami Modeling database contains Tsunami Scenarios (input parameters), Results of Modeling, and the Bottom Topography (or Bathymetry) data. The scenarios can be created/edited or downloaded from the Internet. The Tsunami Scenario Loader converts data to the Database format. Similar operations can be provided with the Tsunami Modeling Result Data obtained by external modeling components and published on the Internet.

The quality of the bathymetry data is one of the key parameters defining the accuracy of the model. This data is updated periodically. Here, a special bathymetry was developed, covering the area of the Pacific Ocean adjacent to the northwest parts of the Honshu island (Japan). The gridded digital bathymetry for the numerical modeling was prepared using 500 m resolution bathymetry around Japan [14], and 1 arc sec ASTER Global digital elevation model [15]. A computational rectangular grid of 2413x2405 points includes knots of pre-setup depth values. Length of a spatial step in both directions made 0.0024844 geographical degrees that is about 277 m in a North-South direction and about 221 m in the West-East direction. The bottom relief of the domain is stretching from 34 to 40 degrees of North Latitude and from 140 to 146 degrees of East Longitude, and is shown in Figure 3.

Modeling used tsunami data generated from the Great Japanese Earthquake (38.322°, 142.369°E, Mw = 8.9 at 5:46:23 UTC) on March 11, 2011 [16]. The fault length and width were 400 km $\times$ 150 km. Numerical experiments confirmed the reliability of this technique, and a good fine-grained CUDA acceleration of modeling process [14].

B. Hybrid Tsunami Modeling Combining Natural and Artificial Bathymetry Objects

Lessons from the Great Japanese Tsunami stress importance of reducing impact of tsunamis on different time scales. First, it is necessary to provide real-time tsunami warning. Second, tsunami modeling can be used for long-term hazard assessment (e.g. detailed inundation models across the Japanese sea-line). Third, the well–known "Matsushima effect," the considerable influence of natural geographical objects, like islands and bathymetry, on the wave height and speed of tsunamis, could be considered. Such effect exists in the Matsushima area where presence of islands mitigates effects of tsunamis (while they are absent on the Fukushima coast).

This conjecture is based on [17], where results of a simulation of effects of submarine barriers on tsunami wave propagation were presented. The experiments were conducted in a basin 5 m in length and 10.5 cm in depth. Single and double barriers were used in variable arrangements. Experiments indicated capability of reducing tsunami run-up. It was also shown that parameters of simulations can be translated to natural conditions. Therefore, similar computer simulations could be completed for crucial coastal areas. In this case, aim of the simulation would be to study effects of objects of different shapes, sizes and placement configurations (see, Figure 4). In this way it may be possible to design and build a set of artificial objects (islands) that can be used to protect the coastal areas. In particular, such protection could be of extreme value in highly populated areas, as well as in industrial areas (e.g. nuclear plants, factories, airports, etc.).

Currently, we are enhancing the modeling system by adding the Interactive Bathymetry Editor, supporting object-oriented GUI-editing on bathymetric data, as well as including/removing artificial objects of variable placement, shapes and size, and a plan of changing these parameters during the simulations (see, Figure 5). Here, the modeling process is

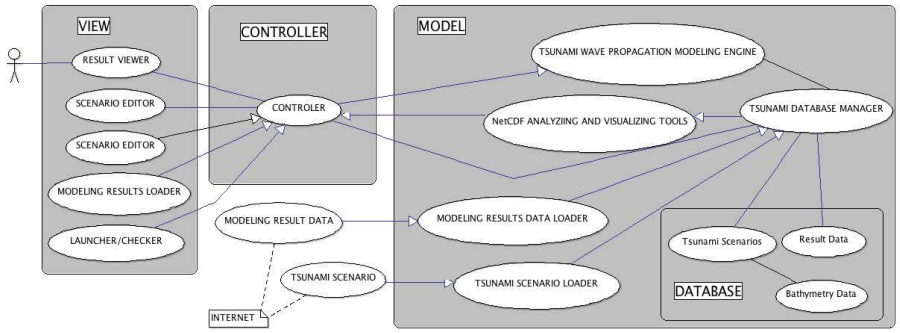


Fig. 2. VMVC-SOA Decomposition of the Tsunami Modeling Environment

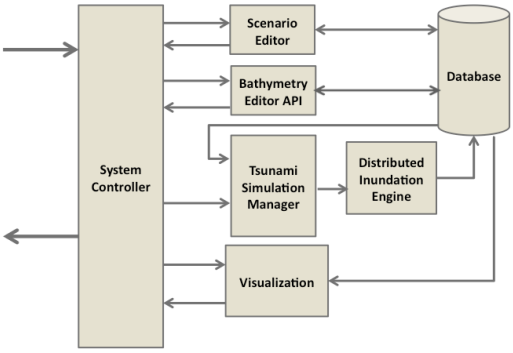


Fig. 5. Components for Hybrid Tsunami Modeling

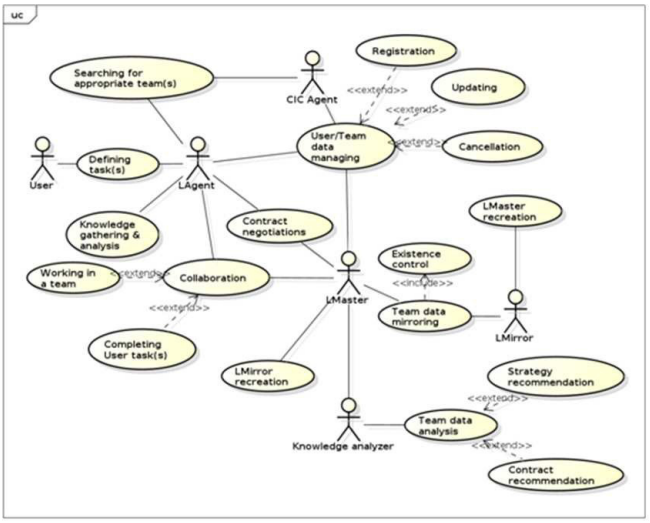


Fig. 6. Use Case diagram of the AiG system

controlled by the Tsunami Simulation Manager, and has as its goal finding suitable number, sizes, and placement of artificial islands and submarine bathymetry objects that minimize the dangerous tsunami wave parameters (height and speed). This problem requires multiple runs of identical tasks with different parameters. While such computing scenario could be realized in a BOINC-like environment, for reasons outlined above, we have decided to attempt at their realization using the agent-based infrastructure.

V. AiG for Tsunami Modeling

The Agents in Grid (AiG) project aims at providing a flexible agent-based infrastructure for managing resources in the Grid ([18], [19]). Application of software agents and semantic technologies makes it well-suited for open, dynamic and heterogeneous environments. The AiG architecture is based on the concept of an open Grid – a network of heterogeneous resources, owned and managed by different organizations. It allows for users to either provide a new resource to the Grid in order to earn money, or to use the Grid to execute a task. Note, that this design nicely fits into the VMVC design pattern, since the Client Application is separated from reusable Services (available resources), while agents representing user and resource manager share characteristics of the Controller. Hence, we can distinguish component responsible for Integration, Functional and Implementation Logic.

The elements of the system have been modeled as software agents: each resource is governed by (and each user is represented by) an *LAgent* and performs its tasks as part of a team, managed by an *LMaster* agent. Teams are registered in a directory service, represented by the *Client Information Center* (CIC) agent, which handles matchmaking of users to teams. The decision, which team to choose to execute the job is a result of autonomous negotiations between the *LAgent* and the *LMaster*s of the appropriate teams. In a similar way, adding a resource to the system involves negotiations with teams looking for new team members. The main features of the system are shown in Figure 6.

In the AiG project all information is represented in ontological format, using the OWL language. The usage of ontologies enables to describe jobs, resources and their relationships in a structured, yet flexible way (for more details, see [20]). Support previously unhandled hardware and software (such as new hardware devices, software libraries or programs) involves only modification of the ontology and does not require any additional customization. As the job descriptions are also defined in OWL, it is possible to specify different parameter

Fig. 7. Example of choosing a property for constraining

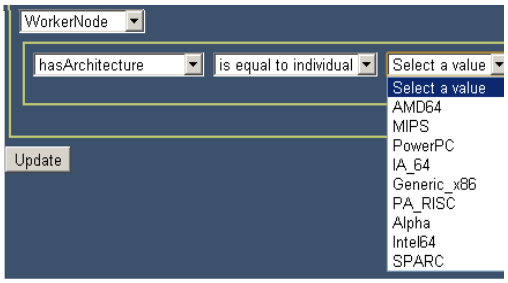


Fig. 8. Example of choosing an individual

sets and required computing resources depending on the type of the task to be performed. This is especially important for the applications described above, where the hardware and software configurations are highly heterogeneous and tools used to perform the experiments may differ from machine to machine.

Attempting at using the AiG architecture at the University of Aizu, one can observe that it is not an open environment (with resources joining and leaving the Grid dynamically). Moreover, there is no economic aspect – if a resource matches the requirements of the job and is available for use, there is no need for negotiations of terms of usage. These differences allow us to simplify its structure and use. As a result, we resign from the notion of resource teams, and place an *LAgent* (playing the role of the *LMaster*) on each computing node. We can also eliminate the scenario of the *LAgent* joining a team. Instead, adding a new resource will mean registering it with the *CIC* as a standalone node (one member team). Finally, we reduce (and re–focus) the negotiations. Their role will be to provide information about current and planned utilization of the resource. This will allow the *LAgent* representing user to decide where to run which job and when.

Let us now describe how the system will allow tsunami modeling using multiple resources in the university laboratory. The user starts by accessing a web based interface, which allows communication with her *LAgent*. Next, she specifies the hardware and software requirements of the job (as constraints on the ontological terms describing needed resources). This task is done using the interface based on the OntoPlay module [21] (its Condition Builder component), giving the user freedom in describing the needed resources, while guiding her through the contents of the ontology without deep knowledge of its structure (of semantic technologies in general). As shown in Figure 7, the Condition Builder is composed of a series of condition boxes used to create constraints on class-property relationships. Depending on the chosen class, the user can select, which class property she wishes to restrict. For example, having selected the *PhysicalMemory* class, the expanded property box will contain properties such as *hasTotalSize* and *hasAvailableSize* (see, Figure 7).

After selecting the class and property the user can choose the required operator and value. Here, she sees only operators applicable to the type of the property. Specifically, this means that for value properties it would be operators such as *equalTo*, *lessThan* or *greaterThan*, while for object, properties the user would be allowed to select, e.g. is equal to individual or is constrained by. Should a user wish to restrict the value of a particular property to a fixed individual from the ontology, Condition Builder lists all individuals that can be used in the context (see, Figure 8).

To illustrate the process, let us assume that the user wishes to find all resources that are running the Linux operating system, have at least 8 GB of memory and a 4-core processor. In this case the user starts with an empty *ComputingElement* specification. First, he would constrain the property isRunningOS of the class *ComputingElement* to any individual belonging to the class Linux. We assume it does not need to be any particular flavor of Linux, so we do not add additional conditions on this class (however, when installing the AiG system at the University of Aizu, we will represent each machine as an individual in an ontology). Second, the user adds condition on the property *hasMemory*, constraining it to the *PhysicalMemory* subclass and adding a nested condition specifying that the *hasTotalSize* property should have a value greater than 8000 MB. Similarly, the user constraints the *hasCPU* to an instance of the CPU class with a value of the *hasCores* property equal to 4.

After user submits the resource requirements, the *LAgent* passes this description to the *CIC*, which performs semantic reasoning on its knowledge base, to find resources satisfying the given criteria and returns a list of matching nodes, including the information how to contact the *LMaster*s at each node.

Note that even in a semi-dynamic environment, such as a university laboratory, there is no guarantee that the resources found by the CIC are, at the moment, available for use. The machine may be offline, used for other purposes, or the agent process may not be running. Therefore, there is a need for additional verification of the availability of resources. This is handled through agent negotiations, albeit in a very simplified form. When the *LAgent* receives the list of *LMaster* addresses, it issues a *Call For Proposal* message to gain confirmation of whether the resources are able to perform the task. The

LMasters confirms (or rejects the proposal), and provides information when it could start executing the job. This helps to handle the case of temporarily occupied / not available nodes. Once the *LAgent* receives offers from the (benevolent) agents, it presents the list to the user, who can choose the nodes on the basis of their availability and parameters.

The final step of submitting jobs is the specification of task parameters. As described in the previous sections, for the tsunami simulations it is necessary to run different algorithms on different data sets and parameters to come up with multiple results. Consequently, user is required to provide multiple job descriptions (one for each model / parameter). The job description is going to be provided using the same Condition Builder mechanism, although using a different part of the ontology and will contain information such as the command or script for running the computation, together with any additional parameters required by the algorithm, such as location of the necessary data and the place for storing the output(s).

Completed job description is sent by the *LAgent* to the respective *LMaster*, which then executes the task. Once the computation is finished, the *LMaster* creates a *JobResult* message, which contains information about the job execution, the outcome and links to the result data and any resources created by the simulation algorithm. The *LAgent*, on the other hand, is responsible for gathering all responses from the nodes taking part in the experiment and presenting them to the user, thus ending the scenario.

VI. Concluding remarks

The aim of this paper was two-fold. First, to introduce important issues that arise in the tsunami modeling and that require a robust resource management to run the needed simulations. Here, the "Matsushima effect" was used as the main grounding scenario. Second, we have argued that the agent-semantic infrastructure, developed within the Agents in Grid, project can provide the needed solution. This is especially in the situation when multiple resources are under administration of separate authorities and may be availale according to their own schedules. Next, we have discussed changes (simplifications) that need to be introduced to the AiG infrastructure to adapt it to the needs of the tsunami modeling, as a task completed using machines available within the University of Aizu. Finally, we have presented a detailed scenario how the user would use the AiG infrastructure to run a simulation. Goal of our future research is to complete the changes outlined in this paper and install the AiG infrastructure on the machines in the laboratory of Prof. Vazhenin and Prof. Watanobe, and run experiments with tsunami modeling.

Acknowledgment

Work of Marcin Paprzycki was completed while visiting the University of Aizu.

References

[1] Th. Erl, *SOA Design Patterns,* Prentice Hall, 2010
[2] M. Kuniavsky, *Smart Things: Ubiquitous Computing User Experience Design,* Elsevier, 2009
[3] Folding@home Distributed Computing, http://folding.stanford.edu/
[4] A. Beberg, D. Ensign, G. Jayachandran, S. Khaliq, "Folding@home: Lessons from eight years of volunteer distributed computing", *in Proc. of IEEE International Symposium on Parallel & Distributed Processing (IPDPS),* Rome, Italy, 2009, pp. 1–8
[5] X.Caiand, and P.Langtangen, "Making Hybrid Tsunami Simulators in a Parallel Software Framework", *LNCS,* vol. 4699, pp. 686–693, Springer–Verlag. 2008
[6] K.Ganeshamoorthy, D. Ranasinghe, K.Silva, and R.Wait, "Performance of Shallow Water Equations Model on the Computational Grid with Overlay Memory Architectures", *Proc. of the Second International Conference on Industrial and Information Systems (ICIIS 2007), IEEE Press,* Sri Lanka, 2007, pp. 415–420
[7] D. George, TsunamiClaw User's Guide, http://faculty.washington.edu/rjl/pubs/icm06/TsunamiClawDoc.pdf/pubs/icm06/TsunamiClawDoc.pdf
[8] N. Shuto, F. Imamura, A. C. Yalciner, G. Ozyurt, TUNAMI N2; Tsunami modelling manual, http://tunamin2.ce.metu.edu.tr/
[9] V. Titov, "Numerical Modeling of Tsunami Propagation by using Variable Grid",*Proc. of the IUGG/IOC International Tsunami Symposium, Computing Center Siberian Division USSR Academy of Sciences, Novosibirsk,* USSR, pp. 46–51, 1989
[10] V. Titov and F. Gonzalez, "Implementation and Testing of the Method of Splitting Tsunami (MOST)",*Technical Memorandum ERL PMEL-112,* National Oceanic and Atmospheric Administration, Washington DC, 1997
[11] J.C. Borrero, K. Sieh, M. Chlieh, and C.E. Synolakis, "Tsunami Inundation Modeling for Western Sumatra",*Proc. of the National Academy of Sciences of the USA,* Vol. 103, N 52, http://www.pnas.org/content/103/52/19673.full, 2006
[12] R. Cortez, A. Vazhenin, "Developing Re-usable Components Based on the Virtual-MVC Design Pattern",*LNCS,* vol. 7813, pp. 132–149, Springer-Verlag, 2013
[13] A. Vazhenin, K. Hayashi, Al. Romanenko, "Service-oriented tsunami wave propagation modeling tools",*in Proc. of the Joint International Conference on Human-Centered Computer Environments (HCCE '12),* Aizu-Wakamatsu, Japan, ACM Publisher, 2012, pp. 131–136.
[14] A. Vazhenin, M. Lavrentiev, A. Romanenko, An. Marchuk, "Acceleration of Tsunami Wave Propagation Modeling based on Re-engineering of Computational Components",*International Journal of Computer Science and Network Security,* vol.13, N 3, pp. 24–31, 2013
[15] http://jdoss1.jodc.go.jp/cgi-bin/1997/depth500_file
[16] http://www.gdem.aster.ersdac.or.jp/search.jsp
[17] http://iisee.kenken.go.jp/staff/fujii/OffTohokuPacific2011/tsunami.html
[18] Katarzyna Wasielewska, Micha l Drozdowicz, Maria Ganzha, Marcin Paprzycki, Naoual Attaui, Dana Petcu, Costin Badica, Richard Olejnik, and Ivan Lirkov. Trends in Parallel, Distributed, Grid and Cloud Computing for engineering. chapter Negotiations in an Agent-based Grid Resource Brokering Systems. Saxe-Coburg Publications, Stirlingshire, UK, 2011
[19] Wojciech Kuranowski, Maria Ganzha, Maciej Gawinecki, Marcin Paprzycki, Ivan Lirkov, and Svetozar Margenov. Forming and managing agent teams acting as resource brokers in the grid–preliminary considerations. International Journal of Computational Intelligence Research, 4(1):9–16, 2008
[20] Micha l Drozdowicz, Maria Ganzha, Katarzyna Wasielewska, Marcin Paprzycki, and Pawe l Szmeja. Using ontologies to manage resources in grid computing: Practical aspects. In Sascha Ossowski, editor, Agreement Technologies, volume 8 of Law,Governance and Technology Series, pages 149–168. Springer Netherlands, 2013
[21] M. Drozdowicz, M. Ganzha, M. Paprzycki, P. Szmeja, K. Wasielewska, OntoPlay – a flexible user-interface for ontology-based systems, http://ceur-ws.org/Vol-918/111110086.pdf
[22] A. M. Fridman, L. S. Alperovich, L. Shemer, L. A. Pustil'nik, D. Shtivelman, An. G. Marchuk,and D. Liberzon, "Tsunami wave suppression using submarine barriers," Physics-Uspekhi, vol. 53, no. 8, pp. 809–816, Aug. 2010
[23] Ian Foster, Nicholas R. Jennings, and Carl Kesselman. Brain meets brawn: Why grid and agents need each other. Autonomous Agents and Multiagent Systems, International Joint Conference on, 1::8–15, 2004
[24] http://depts.washington.edu/clawpack/geoclaw/
[25] http://nctr.pmel.noaa.gov/ComMIT/

11th Conference on Advanced Information Technologies for Management

We are pleased to invite you to participate in the 10th edition of Conference on "Advanced Information Technologies for Management AITM'2012". The main purpose of the conference is to provide a forum for researchers and practitioners to present and discuss the current issues of IT in business applications. There will be also the opportunity to demonstrate by the software houses and firms their solutions as well as achievements in management information systems.

Topics

The topics of interest include but are not limited to:

- Concepts and methods of business informatics
- Business Process Management and Management Systems (BPM and BPMS)
- Management Information Systems (MIS)
- Enterprise information systems (ERP, CRM, SCM, etc.)
- Business Intelligence methods and tools
- Strategies and methodologies of IT implementation
- IT projects & IT projects management
- IT governance, efficiency and effectiveness
- Decision Support Systems and data mining
- Intelligence and mobile IT
- Cloud computing, SOA, Web services
- Agent-based systems
- Business-oriented ontologies, topic maps
- Knowledge-based and intelligent systems in management

Event Chairs

Dudycz, Helena, Wrocław University of Economics, Poland

Dyczkowski, Mirosław, Wrocław University of Economics, Poland

Korczak, Jerzy, Wroclaw University of Economics, Poland

Program Committee

Abramowicz, Witold, Poznan University of Economics, Poland

Alt, Rainer, University of Leipzig, Germany

Andres, Frederic, National Institute of Informatics, Tokyo, Japan

Chmielarz, Witold, University of Warsaw, Poland

Cypryjanski, Jacek, University of Szczecin, Poland

Czarnacka-Chrobot, Beata, Warsaw School of Economics, Poland

Damiani, Ernesto, Università degli Studi di Milano, Italy

De, Suparna, University of Surrey, Guildford, United Kingdom

Dufourd, Jean-François, University of Strasbourg, France

Kannan, Rajkumar, Bishop Heber College (Autonomous), Trichy, India

Kersten, Gregory, Concordia University, Montreal, Canada

Kowalczyk, Ryszard, Swinburne University of Technology, Melbourne, Victoria, Australia

Ligęza, Antoni, AGH University of Science and Technology, Poland

Magoni, Damien, University of Bordeaux – LaBRI, France

Owoc, Mieczyslaw, Wroclaw University of Economics, Poland

Pankowska, Malgorzata, University of Economics in Katowice, Poland

Sikorski, Marcin, Gdansk University of Technology, Poland

Stanek, Stanislaw, General Tadeusz Kosciuszko Military Academy of Land Forces in Wroclaw, Poland

Teufel, Stephanie, University of Fribourg, Switzerland

Ziemba, Ewa, University of Economics in Katowice, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1055–1062

Advancements in Cloud Computing for Logistics

Uwe Arnold, Jan Oberländer
AHP GmbH & Co. KG

Terminalring 13, D-04435 Leipzig-Halle Airport, Germany
Email: {arnold, oberlaender}@ahpkg.de

Björn Schwarzbach
University of Leipzig,
Information Systems Institute,
Grimmaische Staße 12, D-04109 Leipzig, Germany
Email: schwarzbach@wifa.uni-leipzig.de

***Abstract*—Adequate integrated ICT infrastructure and services are a prerequisite for keeping pace with the rapid rise of complexity and service levels in logistics. Recent studies indicate a high attractiveness and impact perspective of cloud computing for logistics service providers within few years in order to cope with the growing IT capacity demands. Within this paper, a comprehensive overview is given on R&D with relation to CC for logistics. Among these, the EU-project LOGICAL is presented in detail since it combines different aspects and benefits of CC for the logistics sector. A generic system of CC use cases in logistics and the corresponding needs for a logistics cloud architecture are discussed and compared with the implementation status of the LOGICAL cloud. Special attention is given to the problem of incompatible data and service interfaces. Instead of following the single-window, single-document concept, a semi-automated on demand interface creation service is presented as an intermediate alternative for the practitioning logistics sector.**

I. Introduction

A. Market Background

THE emergence of new cloud computing services is steadily increasing. More and more companies realize the benefits and opportunities of using IT-resources with unlimited scalability and on-demand services at pay per use conditions over the Internet, as opposed to "classical" on-premise installation and operation. A recent survey of web-hosting and cloud computing specialist *Parallels* [1] indicated that especially small and medium sized enterprises (SME) are drivers of extraordinary growth rates above 20% per year in this market domain worldwide.

The full potential of collaborative business processes, especially for logistics companies, is still not exhausted. The benefits for design and organization of heterogeneously fragmented logistics processes based on new logistics software that is available within minutes and allows an easy integration of customers, suppliers and partners, are about to be appreciated by logistics service providers (LSP). The latest *Logistics Trend Radar* report, published by DHL, ranked cloud computing and supergrid logistics among the trends of highest mid and long term impact perspective [2] due to the expectation that these innovative trends will foster completely new process models and service provider types in logistics of the future.

The work presented in this paper was funded by the CENTRAL EUROPE programme co-financed by the ERDF under the project LOGICAL (www.project-logical.eu)

The trends mentioned above may be seen as a partial facet of a larger trend: *bottom-up economics*, a paradigm change which may lead to a total economic reconfiguration in the 21th century, driven by the Internet. The planning, organization and implementation of complex logistics processes is currently carried out by large logistics companies with complex software systems. Cloud computing fosters the cooperation and collaboration of numerous small and medium-sized logistics enterprises without major capital expenditure in IT hardware and software. An open research question is how to cope with heterogeneous data models and interfaces on one side and how to organize and control these cloud-based collaborative business processes.

B. State of R&D in Cloud Computing for Logistics

The development of cloud computing platforms, services and solutions for various business purposes is driven by different institutions, academic and commercial, both on national and international levels. The project "Future Business Clouds" (FBC) alone lists about sixty different cloud computing R&D-projects which are funded by the EU and its member states [3]. Most of these projects, however, are dealing with general technological aspects of cloud computing for business. Just about. 5% of these business clouds, however, explicitly address the application domain of logistics and supply chain management. In addition to these cloud developments, an impressive number of R&D activities and institutional capacities have been initiated internationally in the EU under the 6th and 7th Framework Programme during the last decade upon the field of information and communication systems in transport and logistics. This R&D-domain is relevant for cloud computing in logistics due to a significant focus on interoperability and standardization of data structures in support of collaborative and smart supply-chain management and resource efficient co-modal transport management, especially for SME. An overview on related EU-projects is given by [4].

The joint efforts to solve the problem of incompatible interfaces and data structures as main obstacles of interoperability in the transport and logistics sector lead to a conceptual *Common Framework for Information and Communication Systems in Transport and Logistics* which follows the single-window, single-document approach to create interoperability by standardization. Finally, a unified single transport document shall be established that can be used for

978-83-60810-53-8/$25.00

all modes of transportation. The framework consists of a definition of different "roles" (stakeholders with unique set of responsibilities), "business processes", related standardized "messages" and common ontology based "data elements". Current plans to connect the common framework community with SMEs and proprietary systems aim at the creation of standard web forms and so-called "connectors" (like "translators between differing formats and data models"). As a part of the EU *Freight Transport Logistics Action Plan* the common framework was developed with a holistic perspective by means of integrating the concepts and results of several EU-projects (e.g. FREIGHTWISE, e-Freight, INTEGRITY, Smart-CM, SMARTFREIGHT, EURIDICE and RISING) in the EU-project DiSCwise [4].

A second integrative initiative upon EU-level is the open innovation network platform ETP (*European Technology Platform on Logistics*) which is the output of the 7th FP EU-project WINN [5], [6] and was launched under the acronym ALICE with participation of global players in logistics and industry in June 2013. Among others, involved R&D institutions are the Dutch Institute of Advanced Logistics DINALOG, the Polish competence centre of logistics ILiM and the German Fraunhofer Institute IML. The major issue of ETP and ALICE is virtual collaboration in supply chains. A case study of a platform (T-Scale) based upon global communication standards, which supports virtual supply chains in real time, is described in [7].

A market-oriented approach to collaboration is matching transport demand and offer by means of virtual market places such as online spot exchange platforms and services. An example of this category is the web-based system for rail freight matching developed in the joint R&D-project CODE 24 of the Rotterdam-Genoa corridor [8], programmed by means of open source tools at Duisburg-Essen University. A second example is the project CloudLogistic [9] of Aachen University and industry partners which develops a cloud platform for matching part loads of trucks (capacities and demands) based upon geo-coordinates. Related issues of the business model, SLAs and billing mechanisms are included in this project.

One of the most prominent examples of establishing a virtual market place by means of cloud computing is the ,*"logistics mall"* of the *Fraunhofer innovation cluster for cloud computing in logistics*, developed by the Fraunhofer institutes IML and ISST and operated by Logata GmbH [10]. Basically, the logistics mall serves as a virtual IaaS and SaaS platform for matching demand and supply of logistics software and related IT services. It comprises both an ASP for running proprietary software and a SaaS engine and SOA-bus for combining atomic services with uniform data model and interfaces. Standardization is achieved by means of a uniform ontology and semantic modeling leading to standard *Business Objects* (BO). The mid-term development perspective is a repository of BOs and granular SaaS components which are selected, linked and orchestrated by means of a *Logistic Process Designer* and an interactive graphics user surface. The developers expect IT-cost reductions up to 50% especially for SME due to the mall and its on-demand services.

The example of the logistics mall illustrates that interoperability of IT services and SaaS components of logistics clouds and platforms are crucial for the capability of generating value added especially for the benefit of SME by means of combining available SaaS components to customized virtual process and supply chains. Numerous developments are characterized by standardization approaches like common ontology, semantic programming (e.g. using the language OWL), federated data management and linked open data concepts. Related projects are for instance CollabCloud [11] and COCKTAIL [12] to mention just a few.

A meanwhile finished R&D-project which among other results produced a uniform ontology (in OWL) for logistics was InterLogGrid [13]. Based upon InterLogGrid the joint R&D-project LOGICAL was initiated in the Central Europe programme in order to integrate several of the issues and benefits of the R&D-activities mentioned before: IT- and business process outsourcing, virtual market place for logistics services, integrative data and collaboration space and platform for the orchestration and optimization of collaborative business especially for the benefit of SME-size LSPs [14].

C. LOGICAL profile

LOGICAL's [14] objective is to enhance the interoperability of logistics businesses of different sizes, to improve the competitiveness of Central European logistics hubs through the development of a modern logistics cloud infrastructure. Beneficiaries of the project are especially small logistics companies that are enabled to use cloud-based logistics software to collaborate with other regional and global players. Cloud computing furthermore enhances the hubs' attractiveness for business activities in logistics.

LOGICAL will be simultaneously implemented at six major Central European logistics hubs: Leipzig (DE), Bologna (IT), Wroclaw (PL), Miskolc (HU), Koper (SI) and Ustí nad Labem (CZ). They represent multi-modal infrastructures such as the Airport of Leipzig/Halle, the freight village Interporto Bologna in Northern Italy, one of the most important sea harbours in the Adriatic Sea (Port of Koper) and the largest logistics centre in Hungary. In this way, cloud computing is used by different companies to organize intermodal transports using innovative cloud services.The project started in May 2011 and ends in October 2014.

The results of a survey in order to determine the initial as is situation among participating LSPs, their information demands, typical business processes and first architecture concepts were described in [15]. In the following the LOGICAL architecture and functionality will be presented in detail both from an application oriented view and in terms of the technical components used. Chapter V covers a special contribution to the issue of connecting data and systems with heterogeneous formats and data models.

II. LOGICAL USE CASES

A. Logistics cloud computing architecture: generic use cases

Based upon the survey findings, the identified user demands and migration requirements, major use cases of cloud computing for logistics were developed from a rather practical point of view. This process, however, cannot be considered to be finished, since understanding the opportunities of cloud computing in the logistics application domain grows with usage experience. Therefore, a two-step methodology was put into practice: at first generic use cases (use case classes) were developed and described in order to cover the utilization potential of a logistics cloud as completely as possible. Afterwards, specific use case instances which originated from communications with the survey participants and project partners were presented. The collection of these specific use cases will never be complete due to the ongoing creative process of finding useful new applications of a logistics cloud by means of ongoing interaction and communication with the growing number of users. For the logistics cloud the following generic use cases were identified and are interrelated in multiple ways (see fig. 1):

1) *Outsourcing* of IT resources and related services, i.e. hardware, software applications, and data pools from local (on-premise) IT-systems into a cloud
2) *Integration, Synchronization and Sharing of data* created and utilized by multiple users
3) *Market Place* for product and service offers and demands, platform for adding e-commerce activities to the corporate business models
4) Platform for the *management and optimization of collaborative business* activities of multiple business partners.

1) Generic Use Case 1: IT-Outsourcing: A meanwhile standard application of web-based systems consists in providing and using web-hosted software applications, either by means of an application service provider or by SaaS.

Typical IT functions which are outsourced in general business environments are accounting software, enterprise resources planning software (ERP), customer relations management software (CRM), document management software (DMS) and project management software (PMS).

Outsourcing of logistics IT services for logistics service providers may include transport management software (TMS), route planning software, fleet management software, tracking & tracing software, warehouse management system (WMS), supply chain management software.

Outsourcing is a method which supports enterprises in concentration upon core competences and cutting down secondary or overhead costs. Consequently, following the step of merely outsourcing the IT services of secondary business processes a higher level of this strategy is reached by completely outsourcing the complete related business process, such as accounting processes e.g. financial accounting, personnel accounting, e-procurement & e-commerce fulfillment.

Usually, for outsourcing just one client (e.g. company) is using the web-hosted application provided by the cloud, even

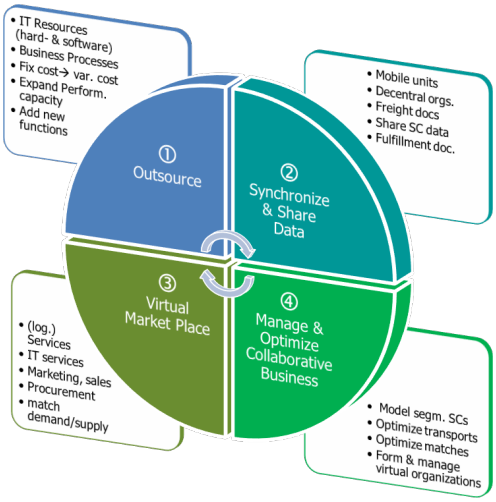


Fig. 1. System of generic use cases of a logistics cloud

if the client is represented by multiple persons (employees, team members). Legally, this relationship can be considered as a 1:1-relation. To find and select web-hosted application or public cloud service a public market place will be used. Another possibility for logistics companies is to develop and use own privte cloud services e.g. with locked data space and encapsulated VM.

The software applications which are offered for IT service outsourcing can be provided as a web-hosted application which instead of running on a local computer is running on a virtual machine. To use *separated instances* of the software, an application service provider (*ASP*) is used as a component of the cloud architecture. To use the *same instances of the software like other users*, the services in the *SaaS runtime engine* are usable.

2) Generic Use Case 2: Synchronize & Share Data: Using cloud computing for the integration, synchronization and sharing of data is one of the original drivers of establishing early cloud systems. A web-hosted managed data space is a basic solution for synchronizing files in simple file-sharing scenarios.

An already well-established representative of this cloud function is the meanwhile widespread DropBox® which offers web-hosted data storage capacity at pay-per-rent conditions (block tariff system based upon booked storage volume independent from actual consumption of the memory space). The Dropbox® is already established with data sharing and access right administration services.

The file synchronize & share function of the cloud can be applied to intra- as well as extra-organization uses of file synchronization and sharing. Typical intra-organizational uses are:

- File synchronization of mobile actors and business units, such as trucks and other vehicles, external service teams, smart devices of employees etc.
- Linkage of subsidiaries, regional or branch offices, ser-

vice posts etc. with the headquarter.
- Business data exchange and synchronization within decentralized organizations.

File Exchange across borderlines of single enterprises and organizations are:

- File exchange with clients
- File exchange among business partners
- File exchange with infrastructure operators such as seaports, airports, intermodal terminals and authorities such as customs authorities

Since data are usually shared among multiple users and the integration space should be a unique one, this relationship can be considered as a n:1-relation. The cloud can be used by clients to give business partners simple access to own files by using a web-hosted storage software suite e.g. DropBox® or OwnCloud.

3) Generic Use Case 3: Market Place: Using the internet as a channel for e-business is state-of-the-art for numerous market participants and traders. Although e-commerce in logistics is still a rather rare phenomenon, a logistics cloud may be the right instrument for adding an online-component to the commercial processes of members of the logistics community. The cloud market place in this context can be a limited access community market or platform open to the public. Since the members of logistics communities cover a wide spectrum of different services, the design of the market place should rather be like a shop of the shops (mall) than a uniform store. Since all functions of the cloud can be considered as a marketable service, the cloud market place can provide access to the whole service repository of the cloud as well as to the complete set of services offered by the logistics communities attached.

A user of the market is addressing to multiple recipients of his sales offer or procurement request. Thus the typical use configuration is a 1:m-relation.

The market place function of a cloud requires the following components of the cloud architecture: e-commerce platform and administration system (affiliate system with purchase monitoring, feed-back system and brokerage provision administration), data base management system, query masks, search & matching engine.

4) Generic Use Case 4: Management platform: The fourth generic use case class represents advanced uses of the cloud which aim at efficiency improvements and value added by means of additional cloud services. Matching demands and supplies in the market place does not automatically mean that a best fit is found. This requires optimization tools, i.e. instruments provided by operations research and systems analysis in order to find an optimum. This optimum may consist in the minimization of cost, carbon footprint, failure risk or a maximum of defined benefit functions. Applications in the logistics domain may be:

- Optimization of transports: best fit of demand and offer of transportation capacities according to predefined goal functions.
- In a generalized form: best fit of any kind of service demand and suitable supplies.

In sophisticated cases, the suitable supply for a service demand may not be offered by a single party but has to composed from the offered capacities of multiple providers as a fragmented sequence of several basic logistic processes, such as transport, storage, cross-docking, transport, intermodal transfer, transport, storage, commissioning, final delivery etc. In such a case, support services are needed for composing the whole process chain and for managing the cooperation of several (heterogeneous) partners. Possible functions of this functionality of collaborative business engineering and management are:

- Composition of suitable logistics process chains (from the online catalogue of single service capacities provided by single partners and covered by the logistics communities)
- Setting up of a "virtual organization" (a special purpose vehicle for logistics projects) of the partnering service providers
- Management and administration of the business processes of the virtual organization with devoted data work space, ERP-service, management tools, job management and billing services and allocation of cost and revenues to the contributing partners.

In these complex cases of multiple actor cooperation, *m* participants are addressed in the composition phase of the fragmented process chain and *n* participants access mutually applied data in the operation phase of the virtual organization. Thus, the use configuration is a m:n-relation.

The collaborative business function of a cloud requires the following components of the cloud architecture: optimization tools, simulation of fragmented logistics process chains, cloud hosted representations and management tools for virtual organizations.

B. Logistics cloud computing architecture: specific use cases

The specific use cases are ordered according to the generic use case classes system presented in the previous chapter.

1) Specific Use Case 1: Logistics software catalog: The logistics software catalog so far contains and provides typical business and logistics software applications. At current state the following applications are available: Standard office software (MS Office 365), ERP software (OpenERP), document management softeare (RICOH DMS), transport management software (PSItms), warehouse management software (*LogBase on Demand*®)

2) Specific Use Case 2: Synchronize & Share: All four use case categories of the LOGICAL cloud require data storage capacity. Thus, use case 2 will be integrated as cloud data space in other use cases.

In addition, the cloud will provide a managed file workspace function for pooling, synchronization and sharing of files in analogy to *Dropbox*®. The related software suite which is going to be used for this function is *OwnCloud*.

Several LOGICAL partners already develop or operate cloud-based systems for the common use and exchange of freight, customs or other official documents and files. These systems and data share functions can be linked to or integrated

into the LOGICAL cloud. Systems to mention in this context are for example:

Logistics Cluster Leipzig-Halle (PP3) and its member SALT Solutions developed a simple software tool for smartphones linked with a web-based data space (e.g. LOGICAL cloud workspace) which helps freight forwarders and other transport service providers to cope with new legal requirements of safety inspection, supervision and documentation. The traffic-manger app of SALT is a direct example of new service products which are developed due to the communication of the R&D projects InterLogGrid and LOGICAL between IT companies and experts and the application community, in our case the logistics service providers organized in the logistics cluster Leipzig-Halle.

Luka Koper (Port of Koper, Slowenia, PP14) developed a planning and scheduling system (TINO) for trucks unloading and loading on the seaport grounds in order to equalize traffic, increase throughput capacity and support freight forwarders as well as the port authorities in planning the logistics processes. In addition, Luka Koper operates a web-hosted information and service platform LUNARIS. Luka Koper now plans to develop a web service and a new module in the cloud platform LUNARIS that would allow registered shipping agents to extract all data from the cloud-solution TINO that are needed to satisfy the customs requirements for the Export Customs manifest.

Another example of using a cloud workspace for the integration of data is the container-information-service provided by port of Koper's platform LUNARIS. The e-zabojnik (e-container) application provides tracking information about containers delivered to, stored within and departing from the seaport grounds.

3) Specific Use Case 3: Market Place: In addition to the logistics IT applications as presented before, the LOGICAL cloud will contain a market place for marketing and matching common logistics services such as transports, warehousing, freight commissioning, and value added services.

As a first step, the offline-partner manual, which was developed in the logistics cluster Leipzig-Halle will be transformed into an online available web solution and extended to all other logistics communities of LOGICAL. This online catalogue of service providers, their available resources (vehicles, technical equipment, warehouses, permits and licenses), logistics competences and frequently served relations as well as features required for international cooperation (language skills, country experiences etc.) can be considered as the online catalogue of the comprehensive logistics service capacities of the LOGICAL community. Parts of the online representation of the web-catalog of partners, competences and capacities shall be publicly accessible for marketing purposes and can be used by shippers for finding appropriate logistics service providers.

The related user and service capacity data are to be stored and managed within the LOGICAL cloud database. In addition to the database itself the LOGICAL cloud surface has to be established with suitable entry- and query masks.

Once the general features of the cloud users are available, the following step will consist in the establishment of the market place open to the public (or only to registered members of the logistics community) where logistics service providers can sell standardized logistics services online via the logistics service market.

The opposite to sales offers, i.e. the placement of logistics service demands by shippers, 4PL-providers and other logistics clients in order to carry out online-tenders for required services has to be introduced as an inverted version (service demand) of the data objects representing offered services (service offers).

The final development level of the logistics service market place will offer a semi-automatic matching service for suitable pairs of matching demand and supply items. This service of the cloud will require a (fuzzy) matching engine.

4) Specific Use Case 4: Management Platform: One of the objectives of logistics communities is to foster cooperation among community members and to develop new forms of collaborative business. The fourth use case of the LOGICAL cloud is meant to combine the functions of the preceding three and to provide supporting cloud services for collaborative business engineering.

Based on atomic logistics services (such as loading, transport, customs handling, storage, packing/unpacking, quality check, labeling, commissioning, cross-docking, final delivery, additional value added activities), the composition of these basic and partial logistics services to complex, fragmented compound logistics services is necessary in order to cover the complete supply chain of preceding, hub-specific and consequent processes. The final objective consists in simulating, monitoring and management of complex logistics processes.

The supply chain process model contains different and multiple process steps (activities) which are executed by different logistics service providers. 4PL-Providers are the main target group of this use case which represents a strategic development direction of logistics clusters (e.g. the logistics cluster Leipzig-Halle). In this way logistics clusters can provide a modern communiction platform for their logistics service providers in order to enable and support cross-company cooperation and collaboration. In particular, the simulation of different combinations and variants for the implementation of complex logistics contracts is interesting to find out the most suitable variant for customers and service providers. Once a suitable chain of basic logistics services is identified and represented as a digital model of the comprehensive logistics process including the description of transport flows, freight quantities, resource volumes, times, cost and other parameters, this model can be used for the management of the process in forward (push-process) or reverse (pull-process) direction.

For the management and optimization of multimodal corridors and transports, Interporto Bologna developed cloud-based platforms and embedded applications CoSPaM and M2 TC.

One of the consequences of composing single logistics services to compound, fragmented service chains by means of collaborative business engineering will be the creation of purpose or project specific consortia of the contributing service

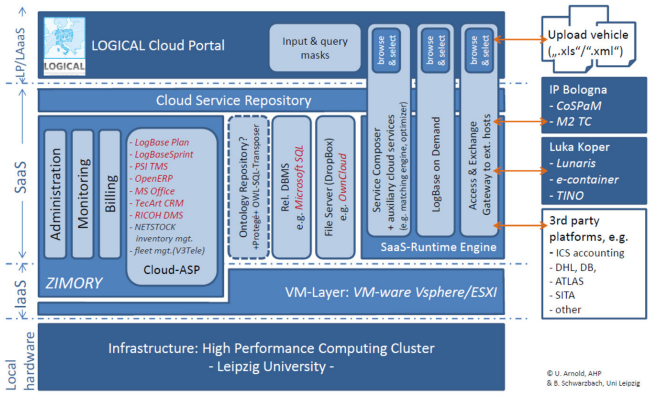


Fig. 2. LOGICAL cloud architecture, technical view

providers. Business administration of these organizations will need additional support like the technical support services of planning, optimizing and controlling the service chains as indicated before.

Thus, the management platform needs to be established with business administration services for the management of "virtual organizations". Like for a single company, the platform will have to provide access and configuration functions in order to apply the IT services of use case 1 to the virtual organization representing the cooperation consortium.

III. LOGICAL Architecture

The following sections describe the technical components and so far selected software products of LOGICAL cloud architecture at current development status in Leipzig.

Fig. 2 displays the technical view of the LOGICAL cloud as presented recently. This architecture consists of the Hardware layer and different software layers:

- IaaS-layer (Infrastructure as a Service),
- SaaS-layer (Software as a Service), with specialisation to logistics software, called "logistics applications as a service" (LAaaS)
- and BPaaS-layer (Busines Process as a Service) with specialisation to logistics processes in sense of "logistics processes as a service" (LPaaS)

So far the applications mentioned in chapter II-B1 are planned to be provided by means of ASP.

The cloud service repository (software catalog) will be used to store, administrate, select and manage the SaaS-offers of the system. The service repository needs to be developed yet. Apart from the software the repository will contain addressing data for the selection and activation of single SaaS applications.

The logistics cloud as outlined in the use case section will need additional SaaS applications to reach full functionality. Eventually required and yet to be specified or developed applications are for instance a search engine, a (fuzzy) matching engine, supply chain simulation tools, transport optimization tools, service composition tools to build complex supply chains, software services for the management of virtual organizations and interfaces to external exchange gateway services.

Logistics applications and logistics processes are specified, addressed, and activated by the cloud user on the top layer of the cloud architecture. The cloud portal serves as entry gate, orientation, service browsing, service selection and control instrument.

The LAaaS-layer and LPaaS-layer consists of the logistics-specific embedded applications and IT-services. In cases of existing web-based user-interfaces of single embedded applications and SaaS applications, these pages very likely need modifications for the integration into the cloud environment, e.g. in order to go back to the preceding cloud portal pages and to link with cloud workspace and composition service components. For the externally hosted platforms which are to be linked with the LOGICAL cloud via data exchange gateway, suitable cross-addressing and linkage tools need to be developed and implemented. The user interface is developed using ASP.NET.

IV. IMPLEMENTATION STATUS

The implementation of the LOGICAL cloud architecture startet in 2012. The implementation is separated into four main implementation project: data management, end user portal, administative portal and service management including user identification.

This section gives an overview of the current state of each of the four subprojects.

A. Model-view-controller

The whole LOGICAL project architecture is following the model-view-controller (MVC)-software architecture pattern that separates data (model), functionality (controller) and end user frontend (view). This approach increases modularity and code reusability. [16]

A model is a POCO (Plain Old CLR Object [17]) that is used to describe the data that is representing a domain specific entity that carries no business logic.

The visual representation of a model is provided by a view that contains information on how the data of the model need to be arranged. Different visual representations, e.g. a print and an onscreen version of the same entity, are generated by multiple views for the same model.

Each view has a corresponding controller that comprises all the functional logic needed by the view, i.e. there is one public method for every single action of the view.

B. Data management

The basis of the cloud is data. There are different kinds of data like data of the cloud users, data of the user's companies and data about alle the services provided by the user's. These data are provided by the data management subproject. It consists of a model of approximately 80 classes, two views for every model and for every view one controller. The first view is the end user's interface to create and edit his data entries into the LOGICAL database.

The more important view in terms of clouding is the API-view of the data. There is one REST-API-view for every model class that provides the methods to access the cloud data by other services inside and outside the cloud. This access is secured by the use of Oauth2.0. This ensures that only authorized services can access this API. The Oauth2.0-server forwards the login information to the OpenId 2 based user identification system.

If there are changes to the model, the adoption of those changes will be done to keep the full coverage of the API

C. Service management and user identification

The user identification is based on OpenId 2 [18] that provides a method to identify a end user without requiring the relying party, i.e. the cloud service, to request the end user's credentials, e.g. username and password. OpenID 2 is using a decentralized system consisting of an OpenID provider and multiple relying parties. The implementation of the OpenID 2 authentication system of the LOGICAL cloud is completed and fully functional.

The same subproject of the logical cloud is responsible for managing all the different available services. Service Providers need to specify some information about their services, i.e. name, description, URL to the logo and a class in a DLL offering state dependent information of the service. This class offers methods for:

- Pricing information returns a string that is shown to the end user to show the current pricing model of the service.
- Service state returns a value out of usable, notbooked, stopped, processing and usable that is representing the current state of the service for a particular user.
- Actions available to the end user that basically are a URL the end user is directed to and a name of the action.

This DLL is dynamically loaded into the service management engine. Since there are a multitude of possible services a webservice for all of this information would not be feasible in terms of timing.

D. Administrative portal

The third subproject is the administrative portal that provides services to the LOGICAL cloud provider to keep the cloud operating. Some of these services are:

- User management: Provides functionality to manage end users of the cloud, especially activation and suspension of an end user.
- Accounting: Provides functionality to charge the end users for their service consumption and to support the cloud service provider by monitoring payments.
- Exception handling: Provides functionality to recover misfuntional services, e.g. reset virtual machines, recovery of wrong data.

E. End user portal

The LOGICAL cloud end user portal is the entry point for end users. It provides information on all the different services that are available via the cloud. The services are assigned to different categories the end users can choose from. After selecting a category the end user receives a list that shows all the information necessary to decide which service is the best fitting for the user and offers the actions defined by the service management layer.

F. Summary

The implementation of the LOGICAL cloud architecture is almost finished. The only major part left is the accounting system which will be implemented in the next months.

Another task for the next months is to identify services that will bring a great benefit to the end users and to incooperate them into the cloud. Since there are always new services this will be an ongoing, task for the whole lifetime of the cloud.

V. SEMI-AUTOMATED INTERFACE CREATION

Compatibility of data structures and interoperability of SaaS components resp. IT-systems of collaborating partners still are prerequisites for achieving the targeted main benefit of the LOGICAL cloud: easy collaboration among different partners along heterogeneous fragmented supply chains. The survey carried out in the initial phase of the LOGICAL project revealed that more than 50% of the existing inhouse-interfaces of software applied by the interviewed LSPs are not at all functioning or insufficient.

Thus, in the beginning of the development a standardized data model concept based upon uniform ontology (InterLog-Grid ontology in OWL, transferred into SQL by means of a specific converter) was selected as a solution to the task of creating IT-interoperability. Workshops and discussions with representatives of the final user group, however, indicated that there is considerable reluctance among practitioners to adopt a standard ontology and to adapt the data models of existent data bases and proprietary software. Therefore, from a practical point of view for intermediate cloud operation an indirect path of linking existing documents and IT-systems with differing data formats and data structures was chosen: like in the Common Framework[4] customized "connectors" (here "upload vehicles", see fig. 2) are introduced for data im- and export.

Now the creation of customized "connectors" turned out to be a new bottleneck of system usability. Assuming an unlimited number of possible source-target-couples of data formats and underlying data models to be mapped, the idea of developing a sufficient repository of preconfigured "connectors" rapidly exceeds feasibility constraints. Thus, in cooperation with Leipzig-Halle cluster-member RICOH, a method of semi-automated creation of data interfaces (connectors) was developed and applied to the problem of mapping differently formatted freight documents (waybills) into each other. With respect to the as-is-situation in the field, the operative cloud concept deliberately refrains from requiring successful establishment of single-window/single-document standards as a mandatory condition. Instead, the system provides a separate tool for the on-demand creation of "connectors" (mapping procedures) which are associated to a specific pair of data

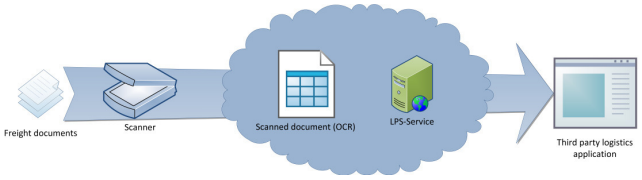


Fig. 3. Workflow of semi-automated interface creation

formats and stored in a mapping repository. Thus, the system is gradually learning during the course of being used and gradually increasing the number of already covered mapping tasks (data/document-format pairings). Once a mapping task occurs which already was tackled before, the mapping procedure does not have to be redeveloped again. Instead a pattern recognition service provides the matching mapping procedure which is applied.

Fig. 3 shows the main parts of the workflow of the semi-automated interface creation use case. Most of documents that need to be processed are paper based, therefore they need to be scanned prior they can be worked on by the software service. The scanning process can be done by existing scanners or a rented scanner that is preconfigured for sending scanned documents directly to the cloud.

Documents that are digitized will be processed directly by the cloud service. To upload such documents, and documents scanned with existing scanners as well, to the cloud a web service with a web interface is provided, that takes files of different formats, e.g. pdf, jpeg and tiff.

After the documents are stored in the cloud they are analyzed by a pattern regocnition service to determine the type of the documents. If the type of the document is known and a mapping is available in the repository the mapping of the data of the document to the LOGICAL database is done by the LPS-service.

Otherwise, a new mapping procedure is created by means of manul linkage of data fields of the original document and the entry mask of the LOGICAL data base. This procedure is supported by the interactive graphics linking features of the LSP and stored as a new mapping procedure in the mapping-repository.

The user of this service can configure a set of third party systems for every document type where these documents should be forwarded to. This process step is done by a Extract, Transform, Load job that extracts the data from the LOGICAL database and loads them to the third party software. If the ETL job requires some data that are not available due to problems with quality (paper based documents and OCR) or input field left blank in the original document, the problem is reported to the user and the user can choose whether to add the missing data or to delete the document.

If the user is operating a third pary software that is not known by the cloud, the user can request a ETL-job for his software. This new job will be available to all cloud users, once it is created.

VI. Conclusion

Cloud computing, with it's service on demand philosophy, enables even small logistics service providers to cooperate with each other. The challenges for the logistics service providers are even more complex if they want to cooperate transnational. The paper has shown the use cases of the LOGICAL cloud in general and detail, that have been developed to enable transnational cooperation. One of these use cases is the semi-automated interface creation that helps logistics service providers with converting documents of one type into another without the need of a comprehensive ontology. The whole set of use cases that are covered by the LOGICAL cloud are resulting in a multitude of new possibilities for logistics service providers to create new added value services with international partners and to be one step ahead compared to the competition.

References

[1] Parallels, "Parallels global smb cloud insights 2013: Profit from the cloud," 2013. [Online]. Available: http://bit.ly/1dXrQ4e
[2] DHL, "Logistics trend radar: Delivering insight today. creating value tomorrow!" 2013. [Online]. Available: http://bit.ly/13Hlowt
[3] OFFIS e.V., "Cloud computing - future business clouds: Europäische projekte," 2013. [Online]. Available: http://bit.ly/16dKx0u
[4] J. T. Pedersen, P. Paganelli, F. Knoors, N. Meyer-Larsen, and P. Davidsson, "One common framework for information and communication systems in transport and logistics," 2011. [Online]. Available: http://bit.ly/145qrd0
[5] WINN, "Winn - european platform driving knowledge to innovations in freight logistics," 2012. [Online]. Available: http://bit.ly/1aKPAbY
[6] ——, "Logistics innovation for a more sustainable and competitive industry," 2012. [Online]. Available: http://bit.ly/1724h6h
[7] M. Hajdul, "Virtual collaboration in the supply chains – t-scale platform case study," in *Intelligent Information and Database Systems*, ser. Lecture Notes in Computer Science, A. Selamat, N. Nguyen, and H. Haron, Eds. Springer Berlin Heidelberg, 2013, vol. 7803, pp. 449–457.
[8] R. Föhring, A. S. Kuhlmann, and S. Zelewski, "Presentation of the final software prototype for an online freight exchange," 2013. [Online]. Available: http://bit.ly/13HlqED
[9] CloudLogistic, "Cloudlogistics," 2013. [Online]. Available: http://bit.ly/166ToTY
[10] Logata GmbH, "logistics mall," 2013. [Online]. Available: http://bit.ly/12m9jyw
[11] Osthus GmbH, "Collabcloud," 2013. [Online]. Available: http://bit.ly/17FHijR
[12] Prosyst, "Cocktail," 2006. [Online]. Available: http://bit.ly/13fEF4F
[13] PSI Logistics GmbH, "Interloggrid," 2009. [Online]. Available: http://bit.ly/1blyNQD
[14] Aufbauwerk Region Leipzig GmbH, "Logical," 2013. [Online]. Available: http://bit.ly/145q42a
[15] U. Arnold, J. Oberlander, and B. Schwarzbach, "Logical - development of cloud computing platforms and tools for logistics hubs and communities," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, 2012, pp. 1083–1090.
[16] G. E. Krasner and S. T. Pope, "A cookbook for using the model-view controller user interface paradigm in smalltalk-80," *J. Object Oriented Program.*, vol. 1, no. 3, pp. 26–49, 1988.
[17] A. Kühnel, *Visual C# 2012: Das umfassende Handbuch*, 6th ed., ser. Galileo Computing. Bonn: Galileo Press, 2012.
[18] OpenID.net, "Openid authentication 2.0 - final," 2007. [Online]. Available: http://bit.ly/xDivq

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1063–1070

Integrated Model of a Social Navigation System with Self-adaptive Feedback Control Mechanism

Vangel V. Ajanovski
Faculty of Computer Science and Engineering
Saints Cyril and Methodius University, Skopje
ul. Rugjer Boshkovikj 16, 1000 Skopje, Macedonia
Email: vangel.ajanovski@finki.ukim.mk

***Abstract*—This paper presents a model of a navigation system in a public information system, that can be used to improve the structure and content of the information repository via self-organization capabilities based on social interaction. This model has the primary goal of establishing a generic and adaptive social-based self-structuring navigation system. To achieve this goal, the model integrates the concepts of social navigation, interaction and self-adaptivity in a feedback control loop. The model gives focus on self-adaptivity and includes elements of social navigation in all parts of the system which enables the implementations based on this model to get social adaptability based on user actions individually, but also as a social environment, in every possible aspect of the functioning of the system. The introduced feedback control loop gives possibility for further autonomous improvements of the organization of the information.**

I. Introduction

THIS paper proposes a model of a navigation system in a information system, that can be used in public knowledge-bases, information portals, news sites, self-support sites, online directories, etc. The model builds towards the enabling of improvement of the structure and content of the repository. We integrate the concepts of social navigation, interaction and self-adaptivity to enable semi autonomous restructuring of the repository.

The original idea was to have as generic approach as possible in order to model a system usable in a wide domain of applications, so some of the concepts are explored from the roots in the past. The concept of social navigation is used as discussed by Dourish and Chalmers in [1], and the generic model is based on the idea of recording interaction on the path that each visitor (navigator) takes, as discussed by Forsberg [2].

There are two main groups of functionalities that are expected from the new navigation system model:

- social interaction should be recorded at all points in the system, either between the system and the visitor or among visitors that are (concurrent or not) visiting the same place of interest in the system – sharing, pointing, recommending and monitoring the published resources and paths through them.
- social interaction history should be used to improve the organization of the navigation and the structure of content – autonomous self-change of the navigation elements should be possible.

In order to include such functionalities, the model first provides the ability to change all navigation point and paths, the whole structure, and then enables social interaction that is associated to the relevant versions of the elements, current to the moment that the interaction has happened. In addition to that, the model enables controlled self-adaptivity, according to results of performed analysis based on social interaction history.

The first step in building this model is the separation of the navigation structure from the content by using a separate navigation sub-system. This sub-system should use a structure that defines (on a conceptual level) and lists (on a logical level) all basic navigational elements that are interesting from a social point of view, and should also include a structure that performs the mapping of logical elements URLs on a physical level. This separation would allow independence of the physical level, which means that changes of the URLs of resources will not have impact on the logical structure and social navigation visible to users. Then, independence on the logical level is introduced, allowing the addition of new elements of social navigation or modification of existing elements, without any change to the physical level.

These are important points for the realization of the structural adaptivity of the system because the constant change of the navigation elements can be the enemy of social navigation (in terms of sharing changed URL, broken paths etc..) For the same reasons, the navigational structure must allow the storage of the entire history of changes (versions of navigation elements) and to handle the shifting of resources in a controlled way that will not disrupt the navigation.

The basic elements of the generic model developed for the case of a web-based public information system are discussed in the following sections.

II. Navigation Elements

The model is organized in levels. At a conceptual level, we identify the following basic types of social navigation elements (with possible extensions of this list), allowing compositions in more complex structures:

- An atomic resource is the basic building element (e.g. a specific resource of this type is: ”Description of the Databases course in the CS study program”, and it can also be an external resource).

978-83-60810-53-8/$25.00

- Atomic resources can not exist by themselves, they only exist as a part of resource-sets, which in turn can be:
  - unordered set of logical resources (example: ”Guidelines of undergraduate studies at FINKI” or ”Course materials for learning Databases”), where the logical resources can be either other sets or atomic elements;
  - ordered set of resources or a directed path (example: ”Installation Guide of an Oracle DBMS”), and again, the resource are other sets or atomic elements. The order of the elements is specified manually by the administrator.

At the logical level, the navigational structure of the overall system is built, individual resource sets and their connections are enumerated, and so for each type. For example, the resource ”Guidelines of undergraduate studies at FINKI” contains ”Description of the CS study program” and ”Description of the IT study program”. At the physical level, a mapping between atomic resources and physical resources or addresses, is performed. As an example, the logical resource ”Description of the CS study program” is mapped to the URL ”http://www.finki.ukim.mk/mk/studies/KNI”.

Figure 1 shows the model of the navigation structure. The class NavigationElement represents all of the navigational elements that define the logical conceptual level. The type of each element is indicated by the attributes isAtomic and isOrdered. The NavigationElementLink class defines logical links between navigation elements. The Resource class provides the physical mapping of various navigational elements to specific addresses (URL) in the WWW space. Thus, a general graph of all navigation elements and their links is formed. So, this structure can be used to model any web-site or parts of it, or even a ”forest” of web-sites.

III. Supporting Self-structuring in the Model

The presented basic conceptual model is then extended to allow self-structuring. For this purpose, two features of the system are required: changeability of the structure and resources. This means that one can change the ordering of resources, or even types of resources with other types, change links between resources (replacing one form to another, adding and deleting logical resources). Also, one should be allowed to change the content of the atomic resources (replacing the mapped physical address to another). Adding an atomic resource (a set of single element) to another atomic resource can produce either an ordered or unordered set, and this is left as a choice for the administrator.

A. Changeability of the Content and Resources

By changeability of resources we actually mean replacing the physical level. The intention is to only change the mappings of logical resources to physical addresses. Also the history of all changes to the mappings should be kept in order to enable later analysis of the behaviour of the users related to each change and pin-point the appropriate version of a resource that is of interest.

The model further elaborates version keeping. In the second row of classes in Figure 1 the versions of the elements and all changes to the physical level are presented. The mapping of physical resources is such that a navigational element is not directly associated with a single physical resource, but in fact a new version of the navigation element is created for any change in the physical mapping. Each change creates a new version object, which records that the version is associated with a change of some resource to a new address. The date of the change is also kept.

In this way, the version entries of each mapping can introduce new versions of the same physical resource – on another date, which would indicate that the navigation element references the same source, again and again, but on different dates. This can interpreted as if the content of the source is changed from time to time, at the version dates, but the URL is still the same. The idea is to have a way to address changes of the content in the model. Also, a completely new version can be created for a new physical resource with a new address. The current version is considered the one with the latest timestamp. The model allows precise monitoring of the changes of the mapping of navigation elements to physical resources, over time.

B. Changeability of the Navigation Structure

The structure of the navigation system can be changed by modifications of the set of navigation links. But, in order to effectively monitor the changes of the structure and the continuity of social navigation – i.e. enabling continuous monitoring of which were original interests of users and where they have migrated after a change, this model only allow for elementary changes and not fundamental change-set where all traces of the past would be gone. The idea is to maintain traceability of each change of the structure.

Only several basic operations are allowed that are implemented in a way that allows monitoring of changes:

- change the type of a logical resource;
- adding new resources to a logical set;
- removing resource from a logical set;
- moving a resource from one to another set.

Considering the way how the logical resources are implemented in the data model using navigational elements – then the allowed set of change operations can in fact completely rearrange navigation structure in any form, when performed as a composition. So in fact any modification can be made but it must be performed gradually, in order to keep proper records of associations of the navigation elements and their version with the social interactions.

1) Changing the type of a resource: Changing a navigation element with another kind is permitted for ordered and unordered sets. As an example, instead of an unordered set of resources, we can decide to have a directed path through the same set of resources or vice versa.

This change is easy to implement over the existing class versions of navigation elements (see Figure 1), with the added ability to record whether a new version includes change of

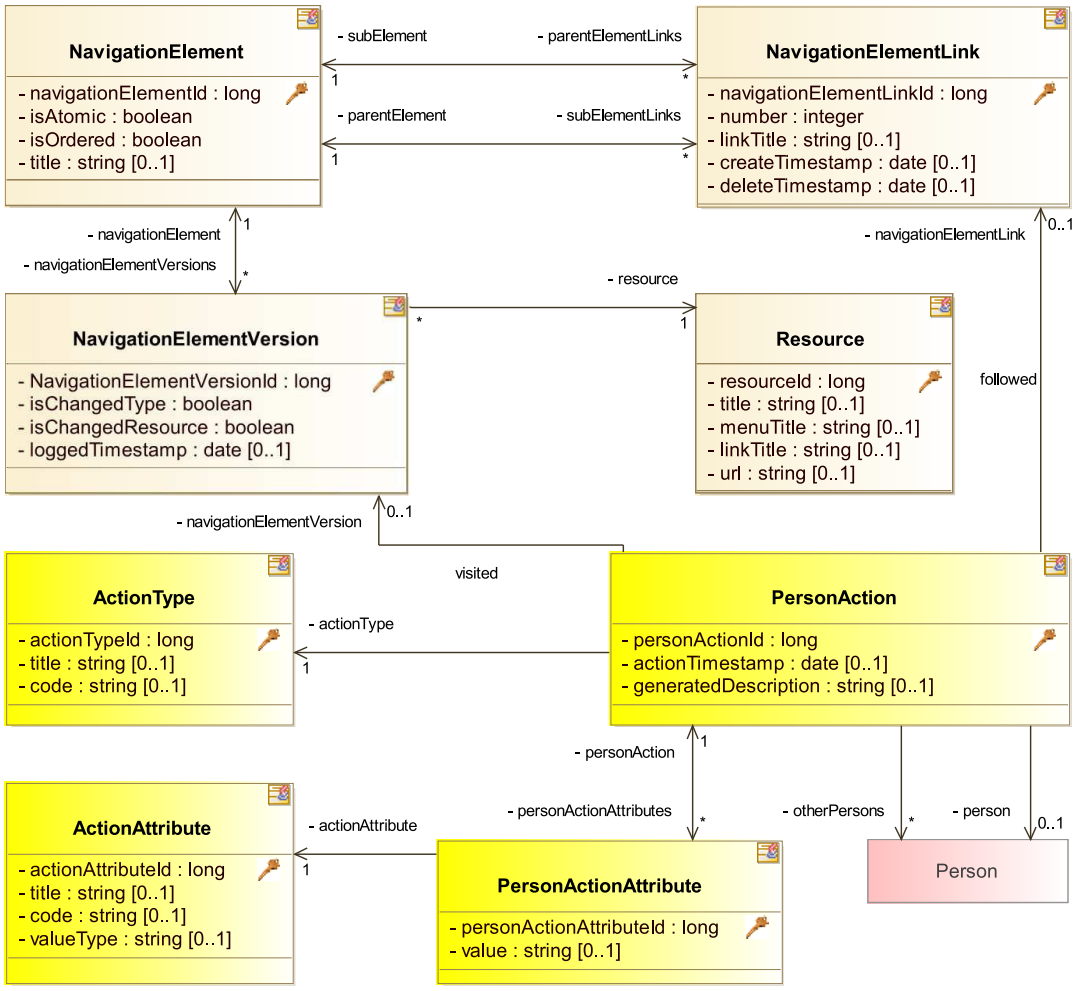


Fig. 1. Data Model of Social Navigation Features.

type. At the same moment, a change of the resource can also be indicated. It is done in this way because it can often be required to reference a new page for the new type and it would be unnatural if even for a small amount of time, first a change of type is visible, and immediately after a change to the mapping to a new content. A change of the element type that does not require a change to the physical resource will be created as a new version and relevant attributes will be denoted, but this new version will still reference the (old) resource. This is done in order to obtain better performance, since search operations are many times more frequent than structural changes.

As discussed at the beginning, an atomic resource itself does not exist, it is always part of a set – in the simplest case it is a set with only one element. For these reasons, a separate consideration of the operation of change of an atomic logical resource into a set, and vice versa, is not necessary. Specific for this case is that what a set has only one element it does not matter if it is ordered or unordered, so the isOrdered attribute is not relevant. So, in this situation the first operation is to replace the type if necessary, and then add new elements or new links where needed and in the required order (especially in the case that the order is important).

2) Adding a logical resource to a set: This operation is realized by creating a new navigation element (if it is a new resource) and then adding a link of the element to the set.

Each link keeps record of the date of creation, which allows one to find out which links were available in a certain period of time.

Note: If the set was made up of only one element, that is atomic, the type of set (isOrdered) has no meaning. So it is ambiguous what will happen once a new element is added. Therefore in this case the administrator must first decide on the new type of the set, perform the change of type operation and only then to execute the operation to add new element.

3) Removal of a logical resource: This operation is realized by setting the value of the attribute deleteTimeStamp in the link that connects the element (child item) to the resource set (parent item). Thus historical data needed for decision making and analytics will not be lost, and navigational elements themselves will continue to exist in the system in case a

new reference to them is needed in another place of the navigation structure. Elements that are not longer connected to other elements, are considered orphan elements and can be monitored in case they are needed again in the future.

Note: The re-inclusion of a navigational element that has already been connected in the past, does not change anything in the historical records, but simply creates a new link by performing the operation of adding an existing resource set.

4) Moving a resource from one to another set: This operation is the same as the composition of two operations: remove a resource from the old location set, and add the existing resource to the new location set.

IV. Social Navigation Support in the Model

Social navigation is always the result of interactions of various entities within the system. In order to enable basic support for social navigation throughout the whole of the navigation system, it is required to keep records of all interactions that have occurred, for each interaction type, with each type of navigational element, and to keep additional parameters for each interaction that occurred. This should be kept as a record in time, associated to the current version of the navigation structure. This means that interaction (viewed, read, accessed) should be recorded for the current version of a navigation element and for the link between elements that was interacted with (followed, clicked, etc.).

These requirements are realized through an extension of the previous model, presented in Figure 1. All types of actions that a person can make (indicated by the Person class) are codified according ActionType and are kept in a log of taken actions (class PersonAction), sorted by the action time stamp.

The actions that have a special meaning and are specially recorded in the journal are:

- interaction with other visitors (list of persons to whom the communication is directed);
- interaction with a navigational element (actual version of the element);
- interaction with a navigational link between two elements.

We have defined that some attributes can be of interest to monitor in association to the actions of the visitors. Such attributes and their values can be recorded. Even more, there is no strict list of attributes to be recorded by activity type, but this is left free depending on the requirements of the implementation.

The Person class is not specified in detail, but should be further elaborated in the implementation of the model according to the requirements and can be used from another system. In the case of public information systems where the majority of visitors are people who are not familiar, there are two options for addressing these visits:

- anonymous option – to use only one anonymous object: Person, toward which all actions are recorded;
- weak authentication – for Each new visitor the system creates a new object of the class Person unless there are kept data from a previous visit, such as data kept in cookies or some other technique to determine that a visitor is actually returning to the system. In such case the new actions are added to towards the found previous instance of the Person class.

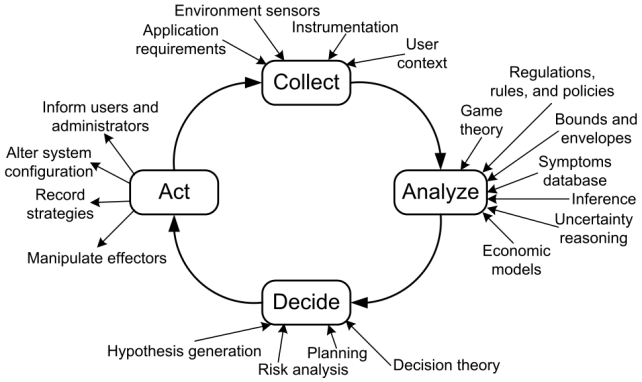


Fig. 2. Feedback control loop[4].

V. Self-adaptivity

The previous discussion identified many elements that allow a constantly changing navigational structure of the system, so it is necessary to monitor all the changes that have occurred and this is necessary for various reasons, the most important being quality control and the ability of the system to constantly adapt to the new needs.

The analysis of the literature in the field of software engineering self-adaptive systems showed different aspects that need to be addressed in such systems, most of which were not relevant for the purposes of social navigation. The most appropriate of all is the model shown in Figure 2, which represents a feedback control loop in a self-adaptive system [3]. This model was originally designed for implementation in terms of communication systems, but can be applied in terms of software engineering self-adaptive software systems [4]. The phases of this cyclic pattern are discussed in the following paragraphs and later the introduction of a control mechanism of that kind is proposed discussed in the generalized model of the navigation system.

In fact this model of a self-adaptive control can be seemed as an appropriate match to the navigation model proposed by Spence [5] (see Figure 3) and the amendments discussed by Riedl [6] for social navigation.

Table I shows the alignment of the concepts of both models. It becomes obvious that one can establish a relation between the two discussed models. The new model uses a feedback control loop to enable self-adaptivity for social navigation in the system.

In fact these are basically two different concepts, one explains the cognitive process of navigation from the perspective of the visitor, and the other describes the process of controlling the operation of a software system that changes its parameters. The idea of linking the two concepts lies in the need of the

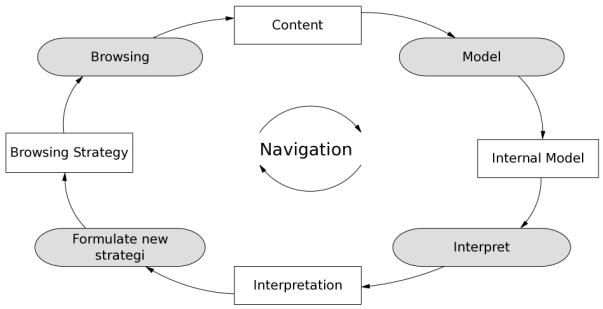


Fig. 3. Navigation Model by Spence[5].

TABLE I
CYCLES OF SOCIAL NAVIGATION AND SELF-ADAPTIVITY.

Social Navigation	Self-adaptivity
review	collecting
modelling	analysis
interpreting	deciding
formulating strategy	action

self-adaptive feedback control loop cycle in order to improve social navigation over the entire set of visitors, but without making direct identification and synchronization of the two processes. The overlap of these two cycles must be understood through an analogy – the mutual coupling of two orthogonally placed gears with varying cogs – the mechanism of social navigation makes many rotations, while the mechanism of self-adaptivity makes one rotation. Through a cycle of the self-adaptive control loop, a number of visitors pass and their (inter)actions are reviewed. The mapping of the cycles in this context would mean:

- viewing information through the system by users generates a multitude of navigational data that some may characterize the overall behaviour of the system, that one needs to collect and analyse;
- symptoms defined by the self-adaptive cycle should apply especially to the navigation system, as they can affect the formation of wrong cognitive model of the visitor or misinterpretation that could lead the visitor leaving the site;
- visitors form their own cognitive model of the system based on how the system is shown and based on personal experience, which can not be directly affected – especially immediately for each user, but one can perform analysis of the behaviour of users, so that in a next cycle assumptions are made about how the visitors perceive the system and what could be changed to improve the navigation
- the interpretation and final formulation of a new strategy for the next search can be influenced by appropriate and timely decisions and actions on behalf of the system, even at low levels - without a major reorganization and just setting some social indicators on well-defended positions, the key in determining whether such interventions help with the navigation process is monitoring the changes made by subsequent self-adaptive cycles in the navigation of the users – what has happened before and after structure changes.

All these features are implementation dependent and are not feasible to be realized in a generic model Therefore it is proposed to plan, define, analyse, design and implement the process of implementation of this generic model with specific issues, depending on the requirements and expectations. What can be generalized is mere conduct the required cycles and their mapping into a generic data model, which will provide the necessary analysis regardless of the implementation at hand. This model is presented in the following text.

A. Structure for Monitoring Control Cycles

Because in social navigation systems it is necessary to monitor all activities of visitors and their usage of resource, in order to establishing rules of behaviour of users and help future visitors, and in addition to that the application system needs to have a certain measure of self-adaptivity – which means that that the system has to adapt itself, and not only at the request and parametrization of visitors, there it is necessary to collect enough data for full system operation and association of these two conceptual models. This requires a detailed data model with the following features:

- tracking all actions that occur in a cycle of the self-adaptive control loop;
- tracking the evolution of all system and process parameters across all cycles of the system.

A basic data model is proposed first, for the control cycle and it is shown in Figure 4. This model enable the monitoring of the system through all the cycles. After that this model is linked to the data model for social navigation.

The part of the model that presents the feedback control loop, records are kept on:

- Sets of defined values of the measured systemic and procedural parameters, normal range of values, increased boundaries and critical limits – in order to analyse the status of processes and objects in the system and
- Journal of measured values of parameters, identified symptoms, made decisions and actions that were taken.

The journal is particularly important element because it tracks the success of the control cycle and can be used to find out if the correct symptom was identified, whether requested action at a time was the one really needed, the results can be seen and the history of decisions checked and to conclusions drawn on the correctness of the decision. Of course, some testing would be done manually by the analytical team and some analysis should be programmed for automated monitoring to work.

The phases of the cycle contain elements which seem directly and solely related and ask why it is necessary separation in four phases. The answer is that this form increases modularity – the system is flexible in terms of combining elements. As an example – setting different symptoms for

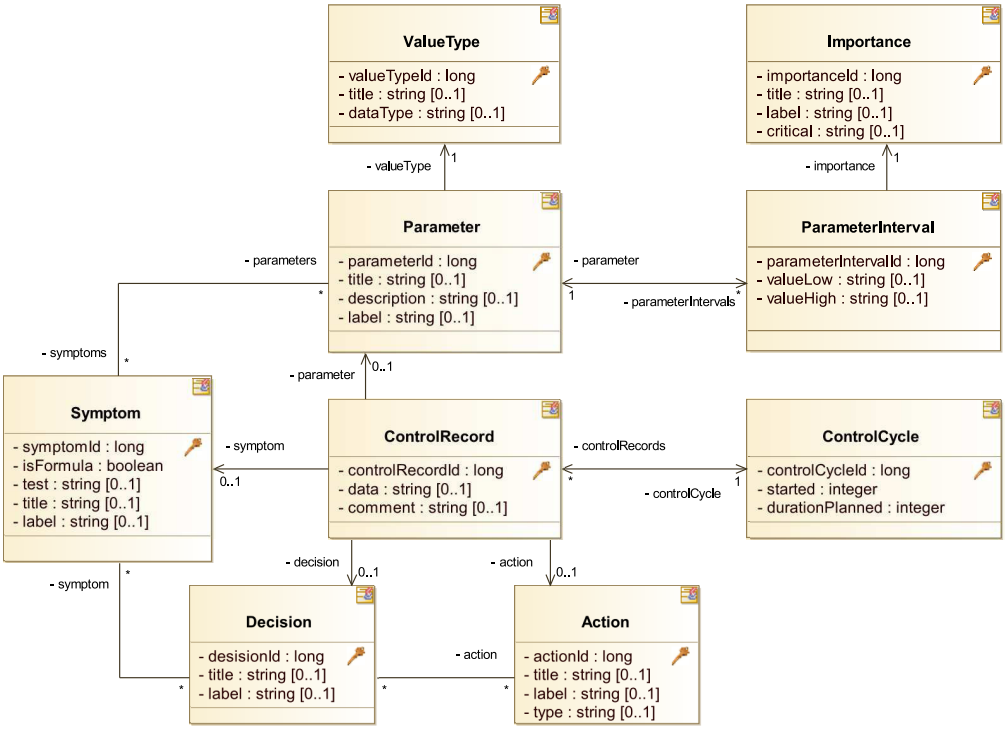


Fig. 4. Data model of the control loop.

the same parameter is allowed, depending on the general context, combination of parameters per symptom, combination of decisions per symptom and combination of actions per decision.

VI. Integration of the Two Models

In order to follow the partial change of the structure of the system and the impact on the behaviour of users, and the impact of changes over the general parameters of the system and boundaries that are allowed, a the system records the changes that have occurred in each self-adaptive cycle.

The model that describes these changes is shown in Figure 5. It should be noted that all associations are optional, which means that the model allows for the incoherent behaviour of the two segments and selecting those components that are required depending on the requirements of the implementation. The RecordedAction class is used to store the information that a certain PersonAction took place within a ControlCycle and is associated with a ControlRecord. Similar to that, the class RecordedModification is used to store the fact that a modification of the structure of content took place and is related to the ControlRecord.

VII. Implementation of the Integrated Model

The steps to implement this generalized model in a production system include final consideration of the requirements of the target system, the amendment of the structures necessary attributes - especially time stamps, recording the users who make administrative changes in the structure and the like.

If the system evolves from the beginning, it is recommended to use this model in all stages as an initial model for defining and navigating through the system.

If the system is implemented as a social and self-adaptive update over the existing system, then you should consider the question of the willingness of the existing system to replace the functionality associated with navigation. If this is not possible, the only solution is the implementation of the navigation model as a separate navigation system, and then mapping and forwarding copies of the shares of users from one system to the other in order to sync the content.

This integrated model can be applied in many different areas, but primarily set in information systems aimed at presenting knowledge and processes related to the management of knowledge and example implementations can be:

- Information portals
- Directories and databases of knowledge systems using
- E-learning
- Social networks oriented learning

VIII. Prototype Implementation and Related Work

The presented model was used to implement a social-navigation based virtual academic adviser[7] that uses recommendations based on social interaction to guide students in the process of course selection, creation of personal study plan until the end of the studies on a longer time scale[8], but also as a self-adaptive control mechanism to create best possible class schedule in term of less conflicts between classes[9]. In

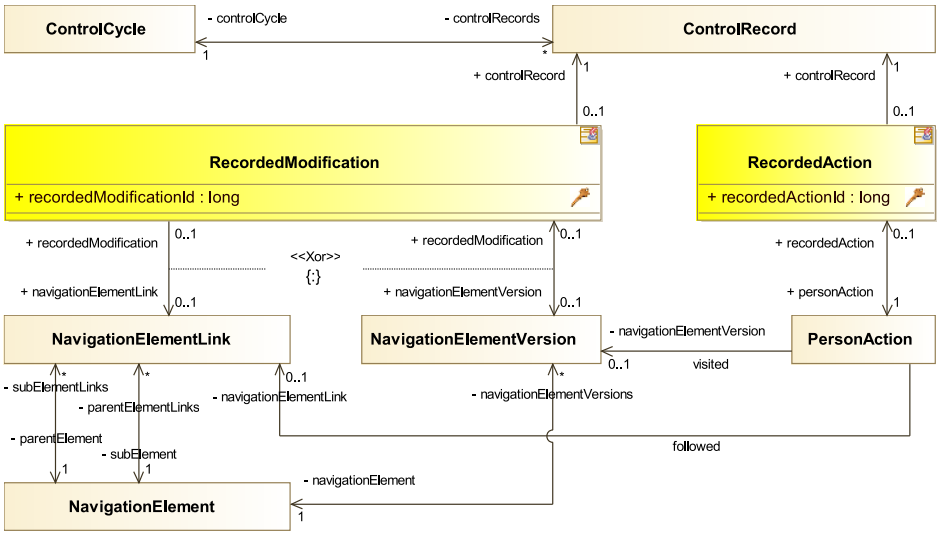


Fig. 5. Integration of the models of social navigation and self-adaptivity.

this implementation, the repository is the database of curricula and study plans for all study programs at a university level.

The virtual academic adviser component gives the student more personal guidance in the process of technical rearrangements of the personal study plan in order to experiment future scenarios with various choices before discussing the final scenario with the real adviser and submitting the application for term enrolment. The first version of the virtual academic adviser enabled the student to try a what if experiment with the various choices on offer such as number of credits per year, choose another study program and specialization profile, rearrange the order of enrolment of courses per future terms etc. After these experiments the student would decide on some preferred scenario and enrol the tern according to the profile and the official rules. In case there were any issues with some scenario, the student could discuss them with the real adviser. It should be noted that the role of the adviser is to give advice and not decide on behalf of the student, and once the choice is legitimate, the adviser can not prevent the student from enrolment, but can only suggest better options.

With the implementation of the new model, the second version of the virtual academic adviser is developed with two additional features: giving the students manual and automated course recommendations and introduction of a new integrated process for term enrolment, class scheduling and construction of timetables. Within this evolution mechanisms for mutual dynamic self-regulation of the processes of term enrolment and class scheduling are investigated. The latest virtual academic adviser is presented in Figure 6.

Each row represents a semester, and each box in the row is a course enrolled in that semester. The semesters are ordered in such a way that the last or active one is on the top, and downwards follow earlier enrollments. Each box shows: the name of the course, whether the lecturer has certified the student was present at majority of lecture hours and is allowed to take exams, the final grade of the student, how many ECTS credits is the course worth if successfully finished. The boxes are colour coded in order to be easily distinguished by the student:

- green boxes represent active courses in current semester
- orange boxes represent courses where the lectures have finished but the student did not have a chance to pass yet
- white boxes represent courses that the student passed
- red boxes represent courses that the student failed.

The system takes into account all interdependencies and course prerequisites and will propose a *realistic plan*. Whether this plan will succeed depends only on the ability of the student to follow and stay through exactly to the new plan and pass all the necessary courses. The system also visually indicates courses that are critical in the future courses where there was a significant amount of failing grades and where students had to re-enroll a course after failing.

The control loop was implemented as a component within the existing information system and used to orchestrate the process of enrollment, phase by phase, student by student or group by group whatever relevant. In the following few paragraphs the four steps of the control loop will are explained with all the included mechanisms.

In the Collect step, the system is monitored and information needed from the students and other system components will be gathered:

- critical time-slots that are almost full
- length of student waiting lists
- numbers of students per status of the enrolment
- numbers of students per year, per group, per status
- number of issues reported by students per category
- number of students without grades for past enrolments

In the Analyse step, the operational status of the system is analysed according to the gathered parameters, historical data and boundary values for several symptoms that are identified

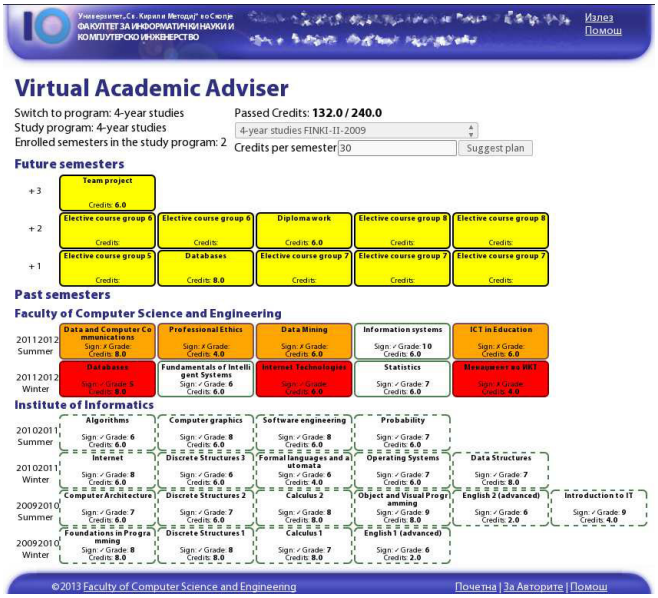


Fig. 6. Integration of the models of social navigation and self-adaptivity.

in the symptom database:

- new groups will be needed soon on a course
- new teachers will be needed soon on a course
- course resources are exhausted
- courses will not be activated due to lack of students
- students ask for courses that are not on offer
- student grades are not input on time
- increasing number of students have complaints

In the Decide step, the system makes decisions on the actions that are to be performed, depending on the severity of the symptomps encountered and how critical are the values of the monitored parameters. In this case mainly the decisions should be made on when and how to act:

- send only information and status via e-mails
- invoke critical alarms to administration staff
- boundary limits can be changed because number of students expected will not exceed significantly
- ask advice on action from administration staff

In the Act step, all actions are orchestrated based on the types of decisions that were made and which simptomps were triggered. The actions that are present are:

- status information is sent to all users
- administration staff is informed that there are issues at hand together with the symptom
- analysis, respective numbers and possible decisions for the decision database
- critical alarms are activated per symptom
- boundary limit are modified and the action is logged
- decisions are logged
- symptoms are logged
- measured parameter values are logged

In such process the course enrolment and time-table creation can be monitored, analysed and acted upon automatically or manually via the proposed control loop framework. In this case it can be argued that the process is successful if finishes on time before the official start of the semester. If that is not the case, the framework gives possibility to monitor the percentage of finished cases of student term and course enrolments and gradually increase the severity of symptoms and frequency of issued critical notifications to the administration staff and to students that have not been active.

IX. Conclusion

The presented model defines a system that is able to change its structure, with traceability of all the modifications of the structure. At the same time records are kept for the interaction of the visitors among themselves and with the system, in association with the exact moment and context regarding how the structure of the system has changed.

This gives possibility to introduce a self-adaptive feedback control loop, that the system will use to monitor itself, identify problems as symptoms and take actions in the form of slight modifications of the structure. The modification are performed in control loop cycles that gives the ability to monitor the causality of the change of behaviour of the visitors and link this change back to a modification in the system, and so investigate if the structural change was an improvement or failure. In such way, the system can undo its steps and take counter measures of bad decisions.

References

[1] P. Dourish and M. Chalmers, "Running out of space: models of information navigation," in *Short paper presented at HCI*, vol. 94, 1994, p. 2326. [Online]. Available: http://fields.eca.ac.uk/deaua/wp-content/uploads/2008/10/hci94-navigation.pdf

[2] J. Forsberg, "Social navigation: an extended definition," *Verfgbar ber: www. nada. kth. se/ forsberg/Documents/(letzter Aufruf am 12.08. 2007)*, 1998. [Online]. Available: http://www.nada.kth.se/~forsberg/Documents/SocNav.pdf

[3] S. Dobson, S. Denazis, A. Fernndez, D. Gati, E. Gelenbe, F. Massacci, P. Nixon, F. Saffre, N. Schmidt, and F. Zambonelli, "A survey of autonomic communications," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 1, no. 2, p. 223259, 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1186782

[4] R. Lemos *et al.*, "Software engineering for self-adaptive systems: A second research roadmap (draft version of may 20, 2011)," Tech. Rep.(October 2010), Tech. Rep., 2011.

[5] R. Spence, "A framework for navigation," *International Journal of Human-Computer Studies*, vol. 51, no. 5, pp. 919–945, Nov. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581999902653

[6] M. O. Riedl, "A computational model and classification framework for social navigation," in *Proceedings of the 6th international conference on Intelligent user interfaces*, ser. IUI '01. New York, NY, USA: ACM, 2001, p. 137144. [Online]. Available: http://doi.acm.org/10.1145/359784.360320

[7] V. Ajanovski, "Towards a virtual academic adviser," in *Proceedings of the 8th International Conference for Informatics and Information Technology (CIIT 2011)*. Molika, Bitola, Macedonia: Institute of Informatics, 2011, p. 146149. [Online]. Available: http://www.getcited.org/pub/103508579

[8] ——, "Personalized adaptive system for term enrollments based on curriculum recommendations and student achievement," in *Proceedings of the International Conference Information Systems 2013*. Lisbon, Portugal: IADIS International Association for the Development of the Information Society Press, Mar. 2013, pp. 342–346.

[9] ——, "Integration of a course enrolment and class timetable scheduling in a student information system," *International Journal of Database Management Systems*, vol. 5, no. 1, pp. 85–95, Feb. 2013. [Online]. Available: http://www.airccse.org/journal/ijdms/papers/5113ijdms07.pdf

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1071–1077

Concept of Platform for Hybrid Composition, Grounding and Execution of Web Services

Lev Belava
AGH University of Science and Technology, Mickiewicza 30, Cracow, Poland
Email: Lev.Belava@gmail.com

***Abstract*—This paper presents a concept of a software platform and a method of hybrid composition of web services and hybrid grounding of abstract composition plans. The paper also describes the architecture of the implemented platform and its modules.**

I. Introduction

RECENT scientific and industrial progress in the field of information technology clearly shows a tendency to switch data processing and computation from traditional models to network architectures. There is also another trend of switching over from large, complex monolithic software systems to groups of smaller, well-defined and interoperable applications. Due to this reason Service Oriented Architecture paradigm and its specific realization – Web Services – naturally fit these general trends. The practice of SOA services composition is a promising approach to developing new software systems with highly refined functionality that is achieved by using a combination of different services chosen from those available in a computer network. Very important parts of such software systems include service composition and grounding. Service composition focuses on creating complex plans from available services, while grounding is a process that transforms abstract composition plans into execution plans so that every abstract service used in an abstract composition plan is effectively associated with a real-world service instance and thus can be called during the execution process.

Hybrid service composition is a method that allows its users to combine different service composition techniques. It offers more flexibility and control of the composition process itself due to the ability to choose different composition techniques for different parts of the composed plan. Hybrid grounding is also a method that allows similar flexibility and control of a grounding process. It allows to mix and match different grounding techniques for different parts of an abstract composition plan.

A. Service Composition and Grounding in SOA

SOA is a software engineering paradigm which generally describes various aspects of software systems that utilize specific entities called services. On the other hand, SOA does not define services strictly – they just have to be relatively independent from each other and offer various functionalities. Such an engineering approach offers different advantages like easier integration of legacy software into new business processes, safer and more reliable upgrades of software components, etc. Web Services are possibly the most common SOA implementation nowadays.

Service composition is a concept of combining different services for data processing purposes. It enables engineers to create complex processes by combining various functionalities offered by available services. So far numerous service composition techniques have been developed.

Manual and semi-automatic service composition methods e.g. [1] and [2] are relatively popular in the scientific community. Such kinds of approaches are fairly easy to understand and implement. When creating service compositions all decisions are made by the user who is provided with some kind of advice or narrowing choice options at most. However, such methods do not offer service composition process automation.

Different automatic service composition approaches have been proposed. Variations of forward and backward chaining methods as in [3], [4], etc. were presented. Hierarchical Task Networks methods were proposed in such works as [5] and [6]. Ontological descriptions of services can be used by reasoners for composition creation [7]. Petri nets were used for service modeling and composition in [8].

Composition methods can, but do not have to, assume that service instances are available and reachable somewhere in a network. So, if a method is not concerned with the availability of service instances, it will produce abstract composition plans. On the other hand, grounding is a process of enriching composition plans with vital information that allows necessary service instances to be used during the execution process. Therefore, abstract composition plans have to be grounded prior to being ready for execution. Several service composition grounding methods have been proposed, some of them are based on brokers [9] while some others are matching-based [10], heuristic [11], agent-based [12] or even ontology-based [13] and [14]. Each one of these approaches uses different perspectives on the grounding process thus allowing their users to fit their needs in a very varied and not always interoperable ways.

B. Problem Statement

There are numerous methods for creating and grounding service compositions. However, every particular approach cannot be an ideal solution from all points of view. Imagine a situation when a user of an SOA software system wants to use some predefined service composition parts and combine

978-83-60810-53-8/$25.00

them with an output of an automatic composition method. The concept of hybrid service composition was specifically proposed in order to solve such kind of problems [15]. This concept allows to use multiple service composition methods during the creation of a service composition plan.

A similar problem is observed in the grounding of abstract composition plans because grounding methods may vary a lot in terms of their work principles as well as optimization targets (QoS, cost, etc.). The concept of hybrid grounding tries to address this problem by utilizing different grounding methods during the grounding of one particular abstract service composition plan.

So far, several service composition platforms have been presented in scientific literature. The most important ones include SWORD [16], METEOR-S [17], MAESTRO [18], SPICE [19]. However, none of them tries to solve the problem of more flexible composition or picking and using a grounding method. SWORD uses first order logic, METEOR-S adopts the Constraint Satisfaction Problem engine for producing a composition, MAESTRO is based on a particular graph method with backward chaining and SPICE uses backward chaining with branching for optimization purposes.

A variety of concepts and methods for service composition and grounding methods, platforms and approaches has been proposed. However, none of them is perfect from every point of view, e.g. such perspectives as composition plan languages or optimization targets for grounding. In order to solve this issue a concept of a hybrid composition, grounding and execution platform was developed. It adopts approaches that enable users to have more flexibility during composition and grounding processes by allowing to use different composition and grounding methods together.

II. Proposed Platform Concept

The architecture of the hybrid composition, grounding and execution platform consists of five key modules that are cooperating together. The concept also incorporates external elements – web services. These services are used by various modules to produce and execute composition plans. Fig. 1 shows the architecture of platform and data flows between different modules.

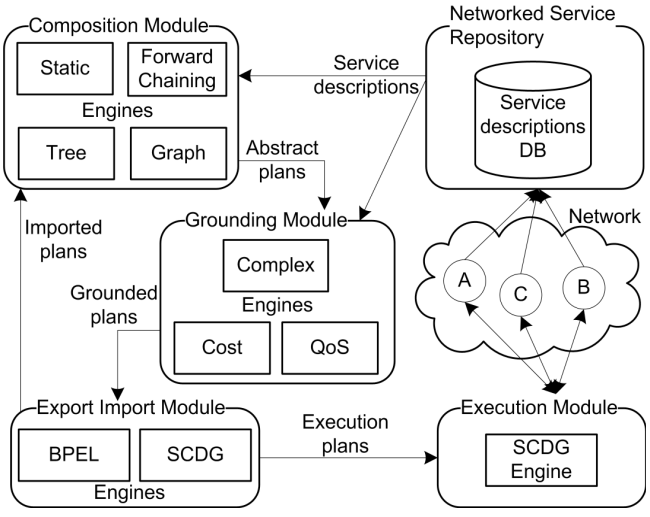


Fig 1. Architecture of Hybrid Composition, Grounding and Execution Platform

The Composition Module is a platform component that actually performs all hybrid service composition tasks and produces abstract composition plans. This module interacts closely with the Networked Service Repository from which it gets web service descriptions. Moreover, it might use the Export Import Module from which it obtains plans for the static composition engine. Abstract composition plans that are produced by this module are forwarded to the Grounding Module for further processing. The module consists of four different service composition engines which can work together to produce composition plans.

The static service composition engine provides necessary functionality to combine different pieces of the composition plan that can be imported or generated by other composition engines. The static engine uses two main operations to work on composition plans: DELETE and INSERT. The "Delete" operation cuts out a specified part of a plan and "Insert" pastes one plan into another. The INSERT (1, 2, plan1, plan2, 4) operation scheme is presented on Fig. 2. Plan1 and plan2 represent two input plans for the operation. Plan3 is the result of inserting plan2 from the first non-root node to "Service 4" node into plan1 between "Service1" and "Ser-

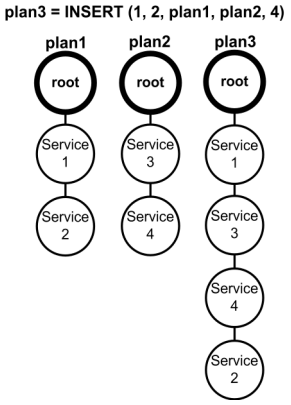


Fig 2. INSERT operation scheme

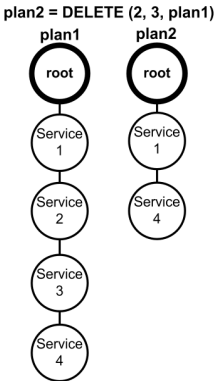


Fig 3. DELETE operation scheme

vice 2" nodes. The DELETE (2, 3, plan1) operation scheme is presented on Fig. 3. Plan2 is the result of cutting a chain of services from plan1 starting at "Service 2" and finishing at "Service 3".

To proceed further we need to provide a definition for a service input and output type. Input or output service types in the proposed approach consist of two parts: the first – a formal description of the data format that the service accepts as input or returns as output, the second – semantic information that describes the meaning of that data.

A forward chaining service composition engine creates service composition plans by using a simple chaining algorithm similar to the one proposed in [4]. Its simplified scheme of action is to successively add new elements to the end of the plan if their input types are consistent with the previous element's output type. The general idea is to create such a chain of elements that its last element will have the desired output type.

A tree-based service composition engine creates service compositions by using a method that creates not just a chain of elements, but a tree. This method is relatively similar to forward chaining but it allows to search the produced trees and because of that the results of its work are more optimal than the results of simple chaining techniques.

A graph-based service composition engine uses a composition method that is similar to the one proposed in [20]. Basically, at the beginning the composition algorithm produces a complete services dependency graph. This directed graph is created by treating abstract services as nodes in a graph and then connecting the nodes with directed arcs if one service's output type is identical to other service's input type. Then such a graph could be processed by Dijkstra or some other pathfinding algorithms. Fig. 4 presents a sample complete services dependency graph. Each node in that graph is described by its input type ("IN") and output type ("OUT").

The Grounding Module allows abstract composition plans, that were produced by the composition module, to be grounded. It cooperates closely with the services repository from which it gets full information profiles about service instances that are available on the network. Such a profile consists not only of the service address and input/output types but also includes additional parameters such as QoS and cost. The three grounding engines in the Grounding Module include QoS, cost and complex. QoS and cost grounding methods were chosen as sample approaches that can be successfully combined in a complex engine. There is a possibility to use and combine other grounding methods as well.

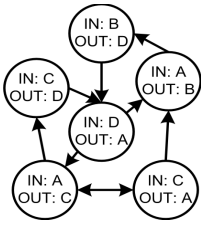


Fig 4. Example of a complete service dependency graph

The goal of the QoS optimization engine is optimizing QoS parameters of composition plans or their parts. For example, one can request that QoS parameters for some part of the abstract composition plan have to reside between some desired maximum and minimum values. In such case the QoS engine will look for service instances that fit the provided values best.

The cost optimization engine works similarly to the QoS engine, but it has a task to optimize the cost of composition plans or their respective parts.

The complex optimization engine allows to create a hierarchical structure of grounding preferences which let the user apply additional optimizations in cases where the engine on a higher level of hierarchy will find several equally fitted service instances. For example, we can imagine a situation in which the cost parameter is the most important target of the composition optimization, but we would like to choose a service with the best QoS in case there are several service candidates with the same cost value.

The Export Import Module provides functionality that allows the abstract and grounded composition plan to be imported or exported from or to files. There are two export-import engines that were implemented for the proposed platform – BPEL and SCDG.

The BPEL engine is able to import [22] and export [21] composition plans that are written in a BPEL language. Not all the BPEL functionality is currently implemented, but core elements like conditionals, loops and the parallel execution of services are fully supported.

The SCDG engine allows to work with composition plans that are presented as Service Composition Directed Graphs. The SCDG is a graph-based model of service composition representation that was proposed in [22].

An Execution Module executes grounded service composition plans. To-date only the SCDG execution engine has been implemented, although there is a possibility to include other engines. To do that one might also need to develop an appropriate import-export engine first.

A Networked Service Repository Module is a web service that on the one hand allows web services to be registered in it and on the other hand provides information about these services for composition and grounding modules. This module also employs a standalone database for service descriptions to be stored in it. A database engine could be either external or internal in relation to the Networked Service Repository. External database engines, however, are much faster and reliable with large data sets and thus more preferable.

III. Use Case Scenario

We can imagine an on-line trading system which allows its users to search for products, place orders and ultimately buy goods by entering financial and personal data into the system. There are all sorts of government regulations and industry standards for personal and financial data because of its sensitive nature. Therefore, we can be sure that some parts of the composition plans in such kind of software platforms will be predefined specifically to obey all sorts of reg-

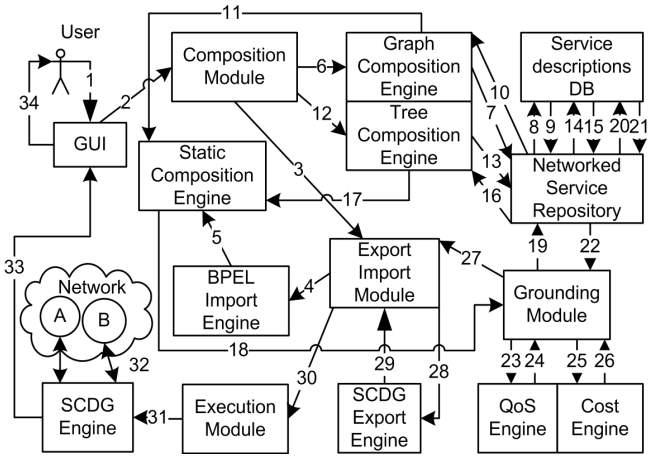


Fig 5. Service composition, grounding and execution diagram

ulations and standards. On the other hand, such systems may benefit from automatic or semi-automatic service composition techniques after all.

Hybrid service composition was proposed to solve exactly such kinds of problems by providing the necessary interoperability between different service composition methods.

A. System's Internal Operation - From Composition to Execution

Fig. 5 presents a diagram with an example of how a service composition plan is made, grounded and executed in a system which implements the platform concept proposed in this article.

1. The user provides necessary personal and financial data and the parameters of the desired products.

2. This data is delivered to a Composition Module.

3. The Composition Module sends a request to an Export Import Module to make an import of a standard-required part of the composition plan which will handle personal and financial data.

4. The Export Import Module transfers the request to a BPEL Import Engine which will actually perform the task of importing.

5. The BPEL Import Engine sends a part of the imported composition plan to a Static Composition Engine, so it can be merged with automatically composed parts later.

6. The Composition Module initiates a Graph Composition Engine and transfers composition parameters to it.

7. The Graph Composition Engine makes a request to a Networked Service Repository and asks for a list of available services types.

8. The Networked Service Repository makes an appropriate query in a Service Descriptions Database.

9. The Service Descriptions Database processes the query and sends back the results.

10. The Networked Service Repository provides the Graph Composition Engine with a list of all available services types (not instances).

11. The Graph Composition Engine sends the prepared part of the future service composition plan to the Static Composition Engine.

Steps 12..17 are similar to steps 6..11.

18. The Static Composition Engine merges all parts of the composition plan into one abstract service composition plan and delivers it to a Grounding Module for grounding.

19. The Grounding Module makes a request to the Networked Service Repository and asks it to provide a list of real-world service instances whose inputs and outputs correspond to the inputs and outputs of the services in the abstract composition plan.

Steps 20 and 21 are similar to steps 8 and 9.

22. The Networked Service Repository provides the Grounding Module with a list of required real-world service instances.

23. The Grounding Module initiates a QoS Engine and delivers the appropriate part of the plan plus the lists of service instances to it.

24. The QoS Engine grounds a part of the greater plan and sends it back to the Grounding Module.

Steps 25 and 26 are similar to steps 23 and 24.

27. The grounded composition plan is delivered to the Export Import Module.

28 The Export Import Module initiates a SCDG Export Engine and provides it with a grounded composition plan.

29. An exported composition plan is delivered back to the Export Import Module.

30. The Export Import Module sends the exported composition plan to the Execution Module for plan execution to be made.

31 The Execution Module initiates a SCDG Execution Engine and provides it with a composition plan.

32. The SCDG Execution Engine executes the composition plan.

33. Plan execution results are delivered to the interface.

34. The interface renders the acquired results and presents them to the user.

B. A Closer Look at Composition and Grounding

Fig. 6 presents a visualization of an abstract composition plan which deals with sensitive personal and financial data provided by the user. There are several service calls in it: “AssignUniqueID” – assigns a unique ID number to a user-provided data set, “Encrypt” – encrypts the data set, “Archive” – archives the previously encrypted data, “ValidateData” – makes appropriate validations of the user-provided data.

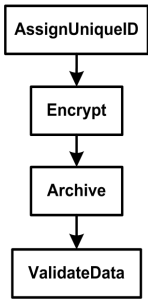


Fig 6. Abstract composition plan in a BPEL language with predefined service calls

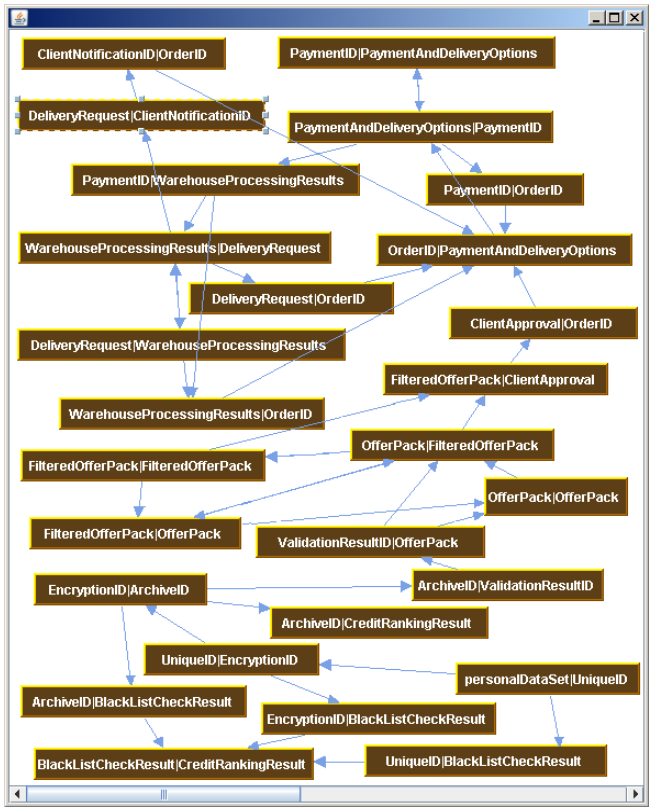


Fig 7. Complete service dependency graph for Service Repository

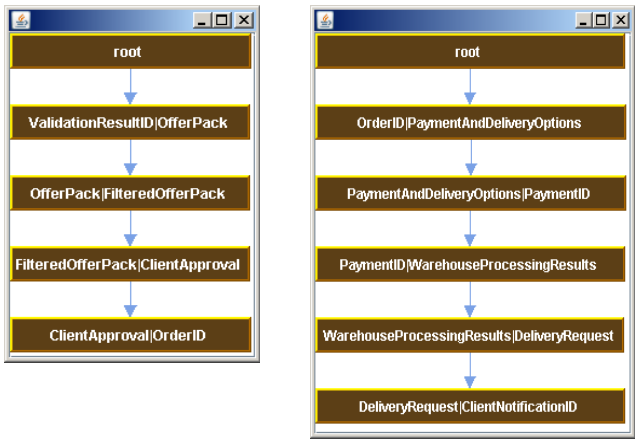


Fig 8. Graph (left) and Tree (right) Composition Engines work results

The automatic generation of the second and third parts of a composition plan was done by graph and tree composition engines. The graph-based engine composed the part of the plan responsible for product finding, selecting and placing an order in a system. The tree-based engine composed the part responsible for the processing of the order that had been placed earlier.

Fig. 7 presents a visualization of a complete services dependency graph of all registered types of web services that were registered in the Networked Service Repository. That exact graph was generated by a Graph Composition Engine during the composition process itself and visualized by the visualization functionality of the software platform. All service type IDs were automatically generated by the Networked Services Repository. Each of these IDs consisted of service input type name, "|" character and service's output type name.

The left part of Fig. 8. presents a visualization of an abstract plan that was generated by a Graph Composition Engine. "ValidationResultID" is an output type of the last service call in a predefined part of the service composition which was imported from a BPEL file, so it was passed to the Graph Composition Engine as a desired input type. "OrderID" type is a type which corresponds to the output type of the order creation service, so it was passed to the composition engine as a desired output type of the composition.

The subsequent steps of a plan generated by the Graph Composition Engine are as follows:

1. "ValidationResultID|OfferPack" represents an automatic wide search of possible products on the client's request.
2. "OfferPack|FilteredOfferPack" represents automatic filtering of the previously found products.
3. "FilteredOfferPack|ClientApproval" represents the client's acceptance of a product offer.
4. "ClientApproval|OrderID" represents generating an order for the offer that had already been accepted.

The right part of Fig. 8. presents a visualization of an abstract plan generated by a Tree Composition Engine. "OrderID" type was passed to the composition engine as a desired input type because it has to be the same as the output type of a Graph Composition Engine's work result. "ClientNotificationID" represents the result of client notification which always happens after an order is processed, so it was passed to the Tree Composition Engine as a desired output type.

The subsequent steps of a plan generated by the Tree Composition Engine are as follows:

1. "OrderID|PaymentAndDeliveryOptions" represents a user's process of choosing payment and delivery options for a created order.
2. "PaymentAndDeliveryOptions|PaymentID" represents the act of payment for the delivery by a client.
3. "PaymentID|WarehouseProcessingResults" represents all background warehousing processing such as searching for the warehouse nearest to the client, scheduling product pickup from the shelf, packing etc.
4. "WarehouseProcessingResults|DeliveryRequest" represents creating a delivery request to a logistic company which will actually deliver the products to the customer.
5. "DeliveryRequest|ClientNotificationID" represents the client notification process during which the client receives information about the delivery and other order related things.

All three parts of a complete abstract service composition plan were merged after they were created or imported by the corresponding engines. After that the complete plan was divided into three grounding areas and grounded in a hybrid mode.

The first grounding area consisted only of the steps from the first part of the plan which had been imported from the

Fig 9. Grounded composition plan

BPEL. Because this part is very important and regulated by government and industry standards, it was grounded only by a QoS Engine which was tuned to select the best available service instance no matter the cost.

The second grounding area was defined as a steps chain from "ValidationResultID|OfferPack" service till "OrderID|PaymentAndDeliveryOptions". The main grounding engine for that area was the Cost Engine and the second one was the QoS engine. The Cost Engine, however, was configured to choose not the absolutely best service from a variety of the available ones, but a range of acceptable services within a provided distance from the best one. The additional grounding engine for the second area was the QoS Engine which was able to choose the service instance with the best cost from the range of the previously selected ones by the Cost Engine.

The third grounding area was defined as a steps chain from "PaymentAndDeliveryOptions|PaymentID" till "DeliveryRequest|ClientNotificationID". It was grounded similarly to the second part but the difference was that the main grounding engine was the QoS Engine and the second one was the Cost Engine.

Fig. 9. presents a visualization of a grounded composition plan. The only difference between the visualizations of the abstract and grounded composition plans are URL addresses of the WSDL files in every step of the composition. These addresses unequivocally correspond to real-world service instances due to that fact that the data in each WSDL file describes a concrete service instance.

The execution of a service composition plan was made with the use of an Execution Module which was making service calls to appropriate instances by their URLs.

C. Implementation Details

The described platform was implemented in Java 6 programming language. Apache Tomcat 7 was used as the servlet container which hosted all the services and the Networked Service Repository as well. Spring Web Services 2 framework was used for the creation of the all web services including the Networked Service Repository. All the communication between the services, repository and Execution Module was carried out using a SOAP protocol over HTTP. Hibernate 4 was used as an object relation mapping framework for storing all service descriptions data in an in-memory H2 version 1.3 database. Such kind of database was used instead of a standalone one because it is easier to use and maintain in projects of the prototype nature. The SCDG was implemented upon JgraphT 0.8 library abstractions which also provided valuable algorithms for the Graph Composition Engine. Jdom 1.3 library for XML was used to handle all XML operations across all platform modules. JGraph 5 library was used to draw the visualizations of service composition plans in the SCDG format such as Fig. 7, 8 and 9.

IV. Conclusion and Future Work

This work presents an approach to creating a software platform that allows its users to combine different composition and grounding methods. Such features enable software users to control composition and grounding processes in a different and more powerful way thus allowing them to create better suited abstract and grounded composition plans.

The proposed approach was also verified during the implementation and execution tests of the described platform. The verification revealed that hybrid composition and hybrid grounding approaches are viable tools that can be used to create a better suited abstract and executable service composition plans.

The main implication of the presented work is the fact that users of software platforms that implement the proposed approach will have more flexibility and control over service composition and grounding processes. Many composition and grounding methods have been proposed, yet each of them is different and may not suit all the needs of the end-point customer. Furthermore, to satisfy all the upcoming and even hitherto unknown customer needs software systems must allow changes to be introduced in them. The ability to choose and combine different service composition and grounding methods addresses these problems by enabling users to select and merge optimal methods for their needs.

There are also three main directions of the upcoming work for the proposed concept. The first one is studying the desired properties of the service universe and service granularity in order to achieve high automation rates. The second one is a hybrid execution of grounded plans. The combined usage of different execution engines might bring some additional features since these engines might employ different approaches and thus be valuable from different points of view. The last direction of the studies has to be made in the field of dynamic composition, grounding and execution methods. Such methods can be very desirable e.g. in soft-

ware platforms where the fault-tolerance level of services is low or the environment itself may constantly be changing.

REFERENCES

[1] E. Sirin, J. Hendler, B. Parsia, "Semi-automatic composition of Web Services using semantic descriptions", in *Proc. Web Services: Modeling, Architecture and Infrastructure workshop in ICEIS2003*, Apr. 2003, pp. 17–24.

[2] E. Sirin, J. Hendler, B. Parsia, "Filtering and selecting semantic Web Services with interactive composition techniques", *IEEE Intelligent Systems*, vol. 19, pp. 42–49, 2004.

[3] S. Thakkar, C. Knoblock, J. Ambite, C. Shahabi, "Dynamically composing Web Services from on-line sources", in *Proc. AAAI-2002 Workshop on Intelligent Service Integration*, July 2002.

[4] M. Sheshagiri, M. Desjardins, T. Finin, "A planner for composing services described in DAML-S", in *Proc. AAMAS Workshop on Web Services and Agent-based Engineering*, July 2003.

[5] E. Sirin, B. Parsia, D. Wu, J. Hendler, D. Nau, "HTN planning for Web Service composition using SHOP2", *Web Semantics: Science, Services and Agents Journal*, vol. 4, pp. 377–396, 2004.

[6] S. Sohrabi, J. Baier, S. McIlraith, "HTN planning with preferences", *Web Semantics: Science, Services and Agents Journal*, vol. 4, 2004, pp. 377–384.

[7] A. Ankolekar, M. Burstein, J. Hobbs, O. Lassila, D. Martin, "DAML-S:Web Service description for the Semantic Web", in *Proc. International Semantic Web Conference (ISWC) 2002*, June 2002, pp. 348–363.

[8] R. Hamadi, B. Benatallah, "Petri Net-based model for Web Service composition", in *Proc. 14th Australasian database conference on Database technologies*, 2003, pp. 191–200.

[9] D. Chakraborty, Y. Yesha, A. Joshi, "A distributed service composition protocol for pervasive environments", in *Proc. 2004 IEEE Wireless Communications and Networking Conference*, Mar. 2004, pp. 2575–2581.

[10] S. Sun, X. Tang, X. Yan, D. Chen, "A symmetric matchmaking engine for Web Service composition", in *Proc. 15th International Conference on Parallel and Distributed Systems*, Dec. 2009, pp. 810–814.

[11] D. Liu, Z. Shao, C. Yu, D. Chen, G. Fan, "A heuristic QoS-aware service selection approach to Web Service composition", in *Proc. 8th IEEE/ACIS International Conference on Computer and Information Science*, June 2009, pp. 1184–1189.

[12] J. Tang, X. Xu, "An adaptive model of service composition based on policy driven and multi-agent negotiation", in *Proc. 5th International Conference on Machine Learning and Cybernetics*, Aug. 2006, pp. 113–118.

[13] H. Yan, W. Zhijian, L. Guiming, "A novel Semantic Web Service composition algorithm based on QoS ontology", in *Proc. 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, June 2010, pp. 166–168.

[14] S. Bleul, T. Weise, "An ontology for quality-aware service discovery", in *Proc. First International Workshop on Engineering Service Compositions*, Dec. 2005, pp. 35–42.

[15] L. Belava, "Concept of hybrid service composition in SOA environment", *Automatyka*, vol. 13/2, pp. 189–197, 2009.

[16] S. Ponnekanti, A. Fox, "SWORD: A developer toolkit for Web Service composition", in *Proc. 11th International WWW Conference*, May 2002.

[17] R. Aggarwal, "Constraint driven Web Service composition in METEOR-S", in *Proc. IEEE International Conference on Services Computing*, Sep. 2004, pp. 22–30.

[18] V. Chifu, I. Salomie, A. Riger, V. Radoi, "A graph based backward chaining method for Web Service composition", in *Proc. IEEE 5th International Conference on Intelligent Computer Communication and Processing*, Aug. 2009, pp. 237–244.

[19] E. Silva, L.F. Pires, M. Sinderen, "An algorithm for automatic service composition", in *Proc. 1st International Workshop on Architectures, Concepts and Technologies for Service Oriented Computing*, July 2007, pp. 65–74.

[20] Y. Wang, H. Wang, X. Xu, "Web Services selection and composition based on the routing algorithm", in *Proc. 10th IEEE International Enterprise Distributed Object Computing Conference Workshops*, pp. 69–73, Oct. 2006.

[21] L. Belava, "Algorithm for the conversion of service composition directed graph into BPEL service composition plans", *Automatyka*, vol 15/2, pp. 71–80, 2011.

[22] L. Belava, "Transforming BPEL service composition into a service composition directed graph for better composition plan management", in *Proc. 25th European Conference on Modelling and Simulation*, June 2011, pp. 424–429.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1079–1086

Analysis of the importance of business process management depending on the organization structure and culture

Witold Chmielarz
University of Warsaw Faculty of Management ul. Szturmowa 1/3, 02-678 Warszawa, Poland
Email: witek@mail.wz.uw.edu.pl

Marek Zborowski
University of Warsaw Faculty of Management ul. Szturmowa 1/3, 02-678 Warszawa, Poland
Email: mzborowski@mail.wz.uw.edu.pl

Aneta Biernikowicz
BOC Information Technologies Consulting al. Jerozolimskie 109/26, 02-011 Warszawa, Poland
Email: aneta.biernikowicz@boc-pl.com

***Abstract*–the present survey mainly aims at analysing determinants of possibilities of improving processes in an organization. The early fragments of the study are devoted to a theoretical analysis of determinants of the process management and its connection with the project management. Then the assumptions of the survey on the impact of the organizational structure and culture on possibilities of applying business process management were presented. The verification of theoretical deliberations and survey assumptions is included in the last part of the article presenting the initial results of the obtained survey and the resulting conclusions.**

Key words–business process management, organization structure, organization culture.

I. INTRODUCTION

THE basic objective of the present article is an attempt to define the meaning of the organization structure and culture for the purposes of streamlining the business process management within it. Numerous Polish and foreign publications [1], [2], [3], [4], [5], [6], [7] define process management in the wide and narrow scopes. The wide scope shows it as a discipline comprising activities identifying, evaluating and analysing the existing processes performed in an organization and their fit for accomplishment of strategic objectives of the organization, It is the base for improvement, optimization, modification or designing new processes (within projects). In the narrow scope – it is a formalized sequence of systematic, measurable steps concerning management of individual business processes in the organization by means of: intuition, explicit and tacit knowledge, inborn and acquired skills, internal (e.g. employees) and external (e.g. customers) stakeholders; theoretical – methodical solutions in the scope of management (change, quality, time, scope, budget management) and related social sciences (economy, sociology, psychology) etc., tools for analysis and tools for process improvement as well as implementation techniques together with process innovations and projects introducing change on the enterprise level; information technologies – supporting the processes, modelling and designing the organization and allowing design and implementation of IT systems using the process management solutions in the management practice of the organization set in a specific economic environment; oriented and changeably (dynamically) conditioned by: the organizational structure (bidirectional relation structure - processes – more efficiently implemented in a proper organizational structure allowing to monitor, analyze and improve the processes; a specific strategy of the organization (relation structure - processes – more efficiently implemented than through competition on a given market), where proper relations result from combining the results of the processes with Key Performance Indicators (KPI); the organizational structure – with the possibility of questioning the inviolability and optimum of the present state, which serves a basis of a possibility of improving the organization, including also transferring and distributing tacit knowledge through a social component of corporate portals and sharing it with other employees of the organization.

From the practical point of view the relations between processes and projects are also essential. At present determinism, explicitness and statics in defining features and results of projects move towards probability calculus, indeterminacy and dynamism. In theory the span between two basic kinds of activities recognized in the contemporary organization: projects and processes should increase. After all projects were defined – as unique, one-time undertakings requiring proper preparation – while processes are repeatable and may be subject to automation or become routine activities. The main difference is the fact that processes are performed permanently and by nature are repeatable, although they can proceed in an unpredictable and changeable way depending on impulses coming from their environment, and projects are performed when new needs occur, and each of them is totally different. But relations between process management and project management have bilateral dimension. On one side process management is treated just as a technique of streamlining project performance. But on the other hand – in a sense – projects are subsets of processes – they are all processes that we could define as non-routine (change-oriented), innovative, pragmatic, burdened with a big risk and unique. This results from peculiar similarities – both kinds of activity are performed by marked out teams of people, determined by specific and limited in time resources, following the rule of planning, steering, supervising and controlling particular acts. This in turn makes the changes within process management have a direct impact on project management. Projects are performed in order to improve the existing processes, create totally new processes and solve specific problems connected with the necessity to change processes (isn't it a component of process management?). In each organization there are both process and project activities. Contrary to its classic definition, projects basically do not end. Each end of one project is the beginning of another, in essence they sometimes create a never-ending cycle of

978-83-60810-53-8/$25.00

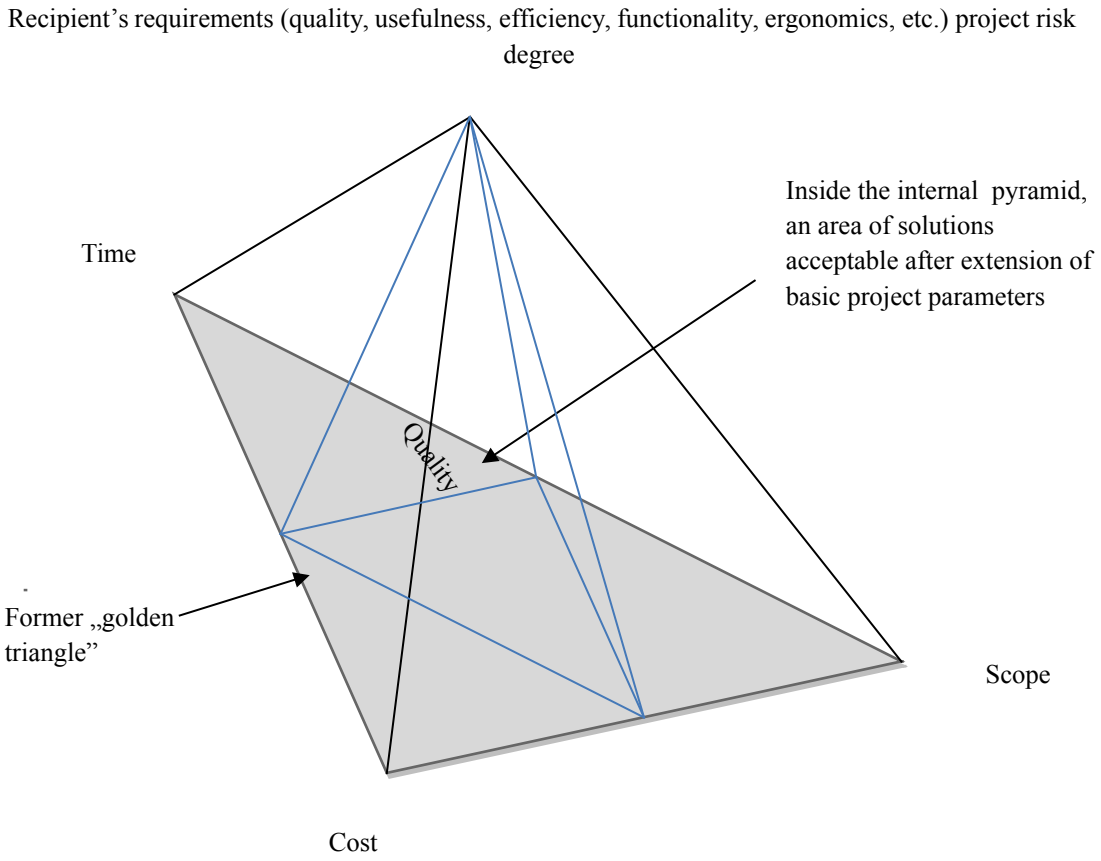


Fig. 1. The area of acceptable combinations of basic project parameters and its extension
Source: own study

projects, which cannot be even named as subprojects because we never know – if only due to uncertainty and high risk – in which direction end users' requirements will develop. But the most symptomatic for project development is the fact that essentially methodologies of project management were in their classic version created, generalized, „toughened", standardized so as to the best possible extent normalize processes occurring in the project. So the paradox – as it shows – consisted in the fact that they got closer and closer to methodologies of process management, as they strived for operating standard rules of solving non-standard problems, which tried to standardize them (that is change into processes) through far reaching formalization.

Due to the above reasons, the notion of a project success at present evolves towards an evaluation exceeding the classic, narrow triangle of balance between costs, time and scope [8]. The point of view of a user – project recipient and his/her way of seeing the project is taken, both in internal projects (in which both persons performing the project and its recipients are employees of the same organization, in which the final product of the project remains), and external projects (products performed for stakeholders from outside of the organization, e.g. customers and may be a source of income of the organization producing them). The extension of the „golden triangle" itself by the fourth parameter – requirements – characteristic for process management – causes also extension of possibilities of making decisions in the scope of its implementation (each decision is described by four sorted out parameters (time, scope, costs, requirements), not by three parameters as before). And the relations between those parameters are becoming – as it seems – non-equivalent – recipients' requirements are superior to other parameters. If we additionally introduce a fifth parameter, which is the quality (one of the component of user requirements) – the number of allowable solutions will again narrow, which will affect multidimensionality of a project and the close connection of a user's requirements with a specific quality level. Not all solutions acceptable within a project and conforming to the user's requirements may meet the assumed quality

standards, and thus the process management (cf. Fig. 1). So changes of relation between project management and process management are affected by their surroundings (environment). The environment, in which projects are implemented, splits into [9]: economic (prices, customs duties, taxes, exchange rates, interest rates, economic policy, markets, economic development degree), legal (legal system, its adjustment to the conditions of implementation, licenses),

technological (technological development, technological state in an organization, quality standards), organizational (organizational structures, management style, managerial staff and employees' skills and knowledge, functionality of the organization, project management method), psychological, (culture, opposition to changes, innovation degree, performance and execution safety) and political (geo-political factors, developmental tendencies, alliances, trends). And here another important issue emerges. The success of a project in the classic perspective and the success of a project in the contemporary perspective (and its management) resulting from practise significantly differ. In the classic perspective (treated this way by many studies) the success is not to exceed costs (and the best thing - execution of the costs), full conformity of the schedule with performance dates and conformity of the performed scope of work with the one specified in the project. Adding the end user's (recipient, customer) point of view means adding to the success evaluation criteria the issue of customer satisfaction with the obtained product or service. Adding a dynamic environment – decrease of a risk of failure, efficiency, effectiveness, flexibility, adaptivity, functionality, etc. And these are evaluations very close to an evaluation of a success of a proper process management in an organization. And very strongly influences them. In streamlining the processes the fact that individual organizations may be at different levels of progress in the scope of process management should also be taken into account. To evaluate this level most often the CMM model (Capability Maturity Model) is applied, which recognizes five basic stages of maturity to process management: first (initial – where processes are not defined at all), second (repeatable – processes were identified in selected departments of an organization and are performed); third (defining) – processes are known in the whole organization and are performed, fourth (managed – conscious use of process management by managers, manifested by collection of data on efficiency of stages of the process and the process as a whole), fifth (optimization – managers and employees monitor on a continuous basis efficiency of processes and introduce necessary modifications). An attempt to introduce process management to an organization, which has not been properly prepared – lacking suitable organizational resources and competence – may result in a failure. Going through each of the organization process maturity levels in this model is an undertaking requiring both extensive knowledge in the scope of process management and using tools dedicated to this purpose and an established and strong internal support centre combining those two elements, for the whole organization, which is e.g. the so-called Process Competence Centre.

The analysis of the results obtained by the Standish Group [10] indicates very practical determinants of the success of project management, and thus of related process management: customer's commitment to the project implementation, project's managerial staff (sponsor) support, clear business objective of the project (specified requirements in the light of existing limits), optimized project scope (adjusted to performance capability), methodology of flexible planning (agile) instead of traditional one, an experienced and competent project manager, proper management of project budget, educated human resources, formal methodology of running the project and standard programming tools and infrastructure.

Among the success factors the „soft" and procedural factors predominate. So it seems that the survey concerning the human and cultural factors as well as the organizational structure as possibly having the most essential impact on organizational improvements, in this situation may be to the fullest extent legitimate.

II. Survey Assumptions

The survey is divided into three parts:

– defining the significance of business process management (BPM) in an organization,

– identifying roles and significance of organizational units dedicated to BPM, for the needs of the survey named Process Competence Centre (PCC),

– defining cultural aspects of process approach implementation.

In the first area attempts were made to define how business process management is perceived by an organization, how this notion is understood and what are strengths and weaknesses of its implementation. Respondents referred to such specific issues as:

– meaning and understanding of business process management (BPM) in an organization (important strategic initiative promoted by managerial staff, essential support for many important projects on a scale of the whole company, support on operating level for medium and small process projects, necessity before implementation of an IT system, studying new possibilities),

– objectives to be achieved by an organization thanks to BPM (create a foundation for development of the whole organization (allow comparison with competition, develop a new organizational structure, improve co-ordination of activities in the company, improve measures and KPI, control risk, increase business process effectiveness, ensure timely delivery of products and services, improve relations with contractors, lower process costs, standardize processes, implement new software, meet employees' requirements for information, try a new approach, introduce new knowledge),

– indication of key processes that require streamlining within a context of income growth (from generic processes list),

– methodologies, techniques and approaches used by an organization: (strategic BPM, Rummel- Brache approach, BPTrends approach, Six Sigma method, Lean - Six Sigma method, Lean method, modelling in BPMN, methods required by ISO, organization's own methods, other),

– subjective perception of organization's process maturity (according to the maturity model based on CMM),

– the extent of implementation of process governance, especially developing (or not) the enterprise process architecture,

– establishing (or not) an organisational unit supporting business process management.

The second area addressed the role and place of organizational structures responsible for construction of the process management system and streamlining business processes, in the scope of the following characteristics: duration of its functioning and its place in the organization's structure; main tasks and services delivered by this organizational unit; resources assigned to an entity dealing with processes; employment of external staff, who deals with the issue of business process management; BPM competence areas required by the organization.

The third area concerned social and psychological factors forming the culture of an organization, such as e.g. defining managerial staff's support for the entity dealing with processes; impact of the entity dealing with processes on the organization's operation; weak points of the entity dealing with processes; strong points of the organizational culture in a given situation.

In each of the defined survey areas a choice was made through marking some proposals from among the ones initially defined by the research team. Moreover, the closed part of the questionnaire ended with open questions, such as: what features should characterize a leader of the team dealing with business processes in an organization?, what features should characterize an employee of the team dealing with business processes in an organization?, what are obstacles in the operation of the Process Competence Centre (PCC)?, what are key support areas for PCC operation?

The survey is at present being implemented by collection of data from the questionnaires filled in via the website www.bpmwpolsce.pl and individual interviews. So far (May 2013) in total more than 50 responses have been collected but the survey is still pending and almost every day new questionnaires are received. The conclusions from selected partial results are presented in chapter three of the present article. The data from the questionnaires were processed by means of IBM SPSS Statistics software. In case of open questions a context interpretation was performed, and then the above-mentioned application was used. A part of the questions in the questionnaire was formulated in a way allowing to compare responses with the results of studies conducted by BPTrends – an American publishing company, whose founders and columnists are experienced practitioners and opinion leaders in the area of business process management, who regularly publish articles and studies presenting the best BPM practices [11]. References concern a part of questions defining understanding and significance of business process management concept and practices in organizations and questions connected with functioning in organizations of special structures dedicated to such activities.

III. Initial Results Of The Conducted Survey

A. *Survey Participants*

Almost a half of the survey participants are representatives of the top and middle management (48%). Another major group of respondents (22%) are experts, who – although usually do not manage teams directly, are employees with high expertise and often create standards of operations for the entire organizations. On the basis of the above results we can state that 70% of respondents are persons with high level of knowledge of the organization and significant impact on its functioning.

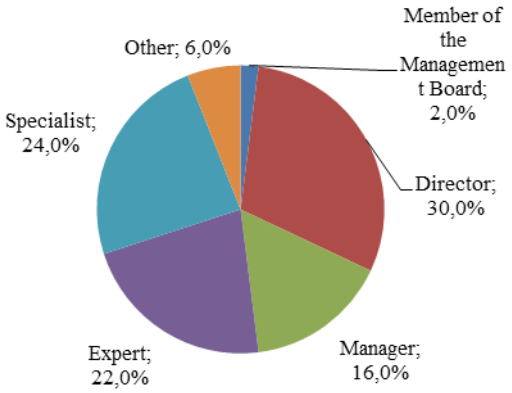


Fig. 2. Survey participants – positions held in an organization
Source: own study

Straight majority of organizations (68%) are big and very big organizations with more than 500 employees, including 26% with more than 5000 employees.

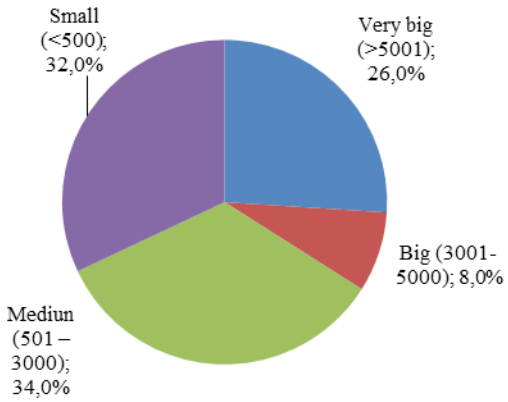


Fig. 3. Survey participants – size of organizations
Source: own study

B. *Significance Of Business Process Management In An Organization*

In one of the questions. t respondents were asked to tick one of the presented below phrases – the one that best describes their understanding of the very concept of business process management. The following results were obtained: 32% of the respondents understand business process management as an approach to manage the entire organization on the strategic level; 30% - as an approach for individual process analysis and improvement; 12% - as a cost reduction and efficiency increasing initiative and for 22% this is just a set of information technologies that help manage and automate processes.

We also wanted to know the significance (importance) of the business process management initiatives for organizations: 30% of the respondents defined it as an

important strategic initiative promoted by management, 24% - stated that the process approach was treated as an essential support for many important cross departmental projects; 18% - sees its significance as a support at the operating level for medium and smaller improvement projects; 16% - as the necessity before IT system implementation; for 12% is only testing of a new approach.

Regardless of the above-mentioned way of understanding BPM and its importance, organizations have some specific expectations connected with investing resources in the area of process management. We asked what goals and objectives were set for the process initiatives in an organization? The respondents could mark all expectations and objectives known to them. We obtained the following answers (see Fig.4).

It is noticeable that the answers selected most often contained generally defined strategic level objectives connected with entire organization systems coordination (foundation for development of the entire organisation, activities coordination). The second most often selected were objectives at the business process level (increase of efficiency, standardization of processes) and only later the detailed, specific, implementation level objectives were selected.

In connection with the above it is logical to ask: do organizations know and use tools (methods, approaches, techniques), which will help to achieve the previously indicated goals and will help to bring expected results. It turned out that the best known process related method is BPMN (52% of the respondents indicated that they know and use it)

This is a popular notation used to describe processes at a very detailed level but it is not useful ether at the process level or at the strategic level as it misses many business elements and symbols. As little as 2% of the respondents indicated specific names of BPM methodologies useful for strategic management, further 32% declared using various unnamed methods of strategic process management. (however they did not specify them later in the item „Other, specify") Awareness and usage of other methods spread out in the following way: Lean (6%), Lean 6 Sigma (16%), 6 Sigma (6%), methods required by ISO (32%), organization's own methods (40%).

So what is the general process maturity level of the surveyed organizations? The participants of the survey were asked for their subjective evaluation of the maturity level in a scale from 1 to 5, corresponding to the levels of the frequently used CMM model applied for process management. This model was selected as there was a solid reference material inter alia coming from regular BPTtrends surveys conducted since 2006. The obtained answers are presented in the first column of the table below and compared with the results of surveys published by BPTrends [11], [12].

In Poland we notice a significantly higher percentage of enterprises, which admit that their organization is at the first stage of maturity, where processes are chaotic and problems are solved ad hoc. We also see a significantly higher percentage of organizations, which place themselves at level 3, where it is expected that organizations have defined their process architecture and key elements of the process management system are implemented. This an interesting result as we take also into account significantly lower (by half) than in the world percentage of organizations, which evaluate their maturity at the preceding level, that is level 2

TABLE 1.
SUBJECTIVE EVALUATION OF THE PROCESS MATURITY LEVEL

Survey result in Poland 2013	BPTrends world survey results 2012	Definition of process maturity level
36%	22%	Processes function thanks to efforts and creativity of employees. Problems are solved ad hoc and not systemically. There are few initiatives referring to processes.
20%	48%	Processes are being improved, but usually within departments or other organizational units. The most important processes have already been described and improved..
32%	22%	Majority of core and enabling processes are identified, published and improved at the organization level. Process architecture is defined. Processes are measured and monitored systematically.
6%	2%	Process owners are appointed, decisions are based on process measures. Processes are managed in the whole organization.
6%	5%	Processes are systematically improved, process governance is executed.

Source: own study and [11], [12]

C. The Role And Significance Of The Process Competence Centre In Organizations

On the basis of collected and presented below data, which concern:

- duration of functioning of a process competence center: less than one year -28,6%; 1 -2 years -14,3%; 3 – 5 years - 25% and above 5 years -32%,
- its position in the organisational structure: at the management board – the manager of the process office reports to the management board - 53,6%; in a division – the manager of the process unit reports to the director of the division 39,3%; in a department – the manager of the process unit reports to the director of the department 7,1%,

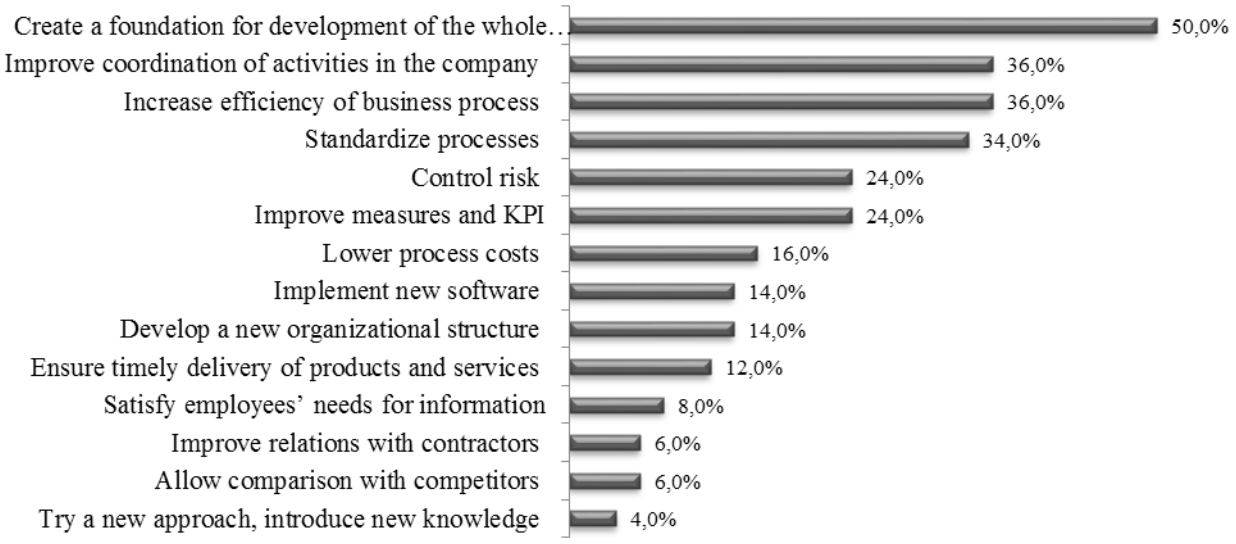


Fig. 4. Objectives to be achieved by an organization thanks to application of methods of process management
Source: own study

– number of employees in this unit: less than 5 employees - 60,7%; 5 – 10 employees - 25%; and 11 – 25 employees -14,3%,

– number of employees in the area of business process management outside this unit: 0 employees -10,7%, 0 – 10 employees 32,1%, 11 – 50 employees 39,3%, 51 – 100 employees -14,3%, more than 10 employees -3,6%,

an image of a typical process competence centre emerges. A statistical process competence centre: functions for less than a year, reports to the management board, employs less than 5 employees but co-operates with a dozen or several dozens (11-50) of persons scattered throughout the whole organization.

Employees of the process competence centre mainly deal with implementation of the following tasks specified in Table 2.

The choices made in the questionnaire by the respondents clearly show their focus on systemic and educational activities. In comparison with tasks implemented by this kind of process competence centre in the world, illustrated by Fig. 5, some fundamental differences can be observed:

1. the Polish competence centres to a significantly greater extent deal with developing rules for the management system, which should not be a surprise as the highest percentage of them is located at the management board and they are expected to build foundations for the whole organization's functioning,
2. the process competence centres in Poland conduct on their own much more training than in other countries,
3. the process competence centres in Poland much less often manage the company's process repository.

The reasons of these visible differences are not subject to an analysis within the present survey. On the basis of our own observations and experience we can suppose that the Polish organizations much less often use IT tools with process repositories but use simpler drawing tools instead and they perform more tasks on their own instead of buying external services (e.g. training).

TABLE 2.
TASKS OF THE PROCESS COMPETENCE CENTRE

Tasks	Participation in realization
Developing rules for the process management system	71%
Conducting training	67%
Maintaining process architecture	57%
Designing new processes	57%
Process modelling	53%
Developing measures and benchmarking	46%
Management of process repository	42%
Popularization of knowledge	42%
Management of project portfolio	14%

Source: own study

Fig. 5. Comparison of tasks performed by the process competence centre in Poland and in the world.
Source: own study

Another question we asked concerned the impact of the process competence centre on an organization. The answers are presented in Table 3. On the basis of the answers we can state that 32% of the respondents have no impact on the organization or performs fake activities – which is even more depressing evaluation of the situation. This is a surprising result in comparison with just 3% of similar answers obtained in the world surveys. At the same time in Poland merely 7% answered that the process competence centre was very important for the success and co-ordination of activities, which – in comparison with exactly such expectation expressed in answers to preceding questions – shows the competence centres mostly still do not meet the expectations and have not yet worked out a position that could help them succeed in an organization.

TABLE 3.
IMPACT OF THE PROCESS COMPETENCE CENTRE ON AN ORGANIZATION

Choice	Our survey – Poland	BPTrends survey – world
Is very important for the success and co-ordination of activities in the whole organization	7%	20%
Has big impact	18%	23%
Just starts to have the impact	32%	53%
Has no serious impact	25%	3%
Performs fake activities	7%	-

Source: own study and [11], [12]

Another question under the study is management support for the process competence centres. Our survey showed that the straight majority (85%) of the respondents said that the management declared support for PCC, but unfortunately only 46% supports them indeed. For the remaining 39% of the management this is just a declaration not supported by actual activities. Other 7% of management is indifferent, and the following 7% even questions the significance of process competence centres.

The survey ended with open questions, from among which we would like to discuss here the following most essential two:

1. What features should characterize a manager of the process competence centre?

Most often the necessity of a very strong and even charismatic leadership was emphasised. The manager holding this position should be able to think strategically, long term and in a holistic way, understand system dependencies, and at the same time should be able to speak in an operating language on details of processes, in order to be credible for line employees. The need of perseverance and consequence was emphasized many times as a necessary characteristics of the PCC manager.

2. What to the greatest extent would help to achieve results by the process competence centre?

The most required is knowledge, resulting also from experience, concerning the way the process organization function. Especially a thorough knowledge is needed, as well as the skill to persuade the advantages of accepting the process approach. The respondents equally often mentioned the need for official support and personal commitment of the organization's management (which in turn confirms the Standish Group's survey).

These and other answers collected during our survey show an image typical for a situation of introducing cultural changes in an organization, where employees' knowledge, belief and attitude play the key role and decide about the success or the failure of a new undertaking.

IV. CONCLUSIONS

Business process improvement initiatives are frequently key projects within an organization – they are managed using project management methods and principles. However the nature of a business process is best recognised and captured by business process management methods. Project management and process management complement each other How does the two approaches interact?

We claim the conjunction point is the newly defined "requirement definition" point at the top of the project management pyramid, which adds the new perspective on traditional project management triangle (scope, cost and time) as described in the previous sections of this paper.

Process management methods are the most suitable for defining the requirements for process improvement projects. The more matured process management the better requirements definition in the project. Therefore it is important to be aware of process maturity of the enterprise, how well it understands its process, how systematic is its view on a business processes how does its culture supports process approach. By combining the process and project methods organisations increase their chances for project success and avoid the situation when improving one process has an adverse affect on other processes

From answers to our survey, it could seem that the strategic understanding of BPM as a holistic system for management of the entire organization predominates. However, if we take a closer look on the results, and refer them to three levels of organization's efficiency described by G. A. Rummler and It A. P. Brache [13], [14] who distinguish the following levels: strategic level, process level and job level – then we find that understanding BPM as an initiative of cost reduction (12%) and as implementation of information technologies (as much as 22%) refer usually to a local management on the job level or even are limited to the area of IT tools. We then state that the interpretation of the concept of business process management spread out more or less evenly among three levels of organization management. There is an equal probability that a person asked by us will classify BPM as an element of strategic management or operating management at the process level or narrowly understood management of the implementation level.

Again it seems that the biggest percentage of an organization (32%) initiates process projects as strategic initiatives at the top management level, thus having

fundamental significance for the whole company's functioning. Not until we set the declared values again at the above referred [13], [14] three level of efficiency, we will be able to for almost a half of surveyed organizations (for 46%) BPM projects have local significance (18%) or are a necessity before implementation of an IT system (16%) or are an experiment and survey of new possibilities (12%).

As research shows organisations expect o lot from the process approach and those expectation refer rather to effective planning of the organization wide systems than to solving specific problems. Taking into account ambitious and system expectations of BPM projects, weak recognition of methods, which could be used for achievement of own objectives, is noteworthy. In connection with it is reasonable to doubt whether organizations will be able to achieve these objectives by means of recognizable and presently applied tools. Probably in most cases – unfortunately not.

The world results show gradual building of process maturity according to the scheme of inverted funnel, that is lower and lower percentage of companies at the following higher and higher levels. In Poland we can observe a leap to the third level. We may propose a thesis that the reason behind this is superficial understanding of system elements characteristic for level three and thus the easy placing home organizations at it. Though also in the world results we can observe a similar, yet smaller, „optimistic leap" but it concerns the fifth level. It is possible that here again we deal with too superficial and unrealistic evaluation of process maturity. As this is a question about a subjective evaluation of the situation, we can suspect that the „leap" phenomenon is more connected to the respondent him-/herself than the very situation of the organization. The issue whether it is connected with the respondents' aspirations or the superficial understanding of the levels too distant from the majority, or whether it results from other reasons – is to be studied in further surveys.

At the same time in Poland merely 7% answered that the process competence centre was very important for the success and co-ordination of activities, which – in comparison with exactly such expectation expressed in answers to preceding questions – shows the competence centres mostly still do not meet the expectations and have not yet worked out a position that could help them succeed in an organization

Summing it up. How well are Polish organisations prepared to derive improvement project requirements from their expertise on business process management?

Probably not too well yet. Business process management has become a popular topic and objectives and expectations towards these initiatives are set high, however the awareness of methods and real managerial support is not there yet. We will need more education and awareness building before business process management methods will be used to define requirements for improvement projects so that project management skills and techniques are used to their full potential to bring process improvements to organisations.

REFERENCES

[1] *Zarządzanie. Tradycja i nowoczesność*, red. J. Bogdanienko, W. Piotrowski, PWE, Warsaw, 2013.

[2] Hensel P.: *Transfer wzorców zarządzania. Studium organizacji sektora publicznego*, Dom Wydawniczy Elipsa, Warsaw, 2008.

[3] *Podejście procesowe w organizacjach*, red. St. Nowosielski, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 169. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu. Wrocław 2011.

[4] *Podejście procesowe w organizacjach*. red. St. Nowosielski Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 52. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 2009.

[5] Nosowski A. *Zarządzanie procesami w instytucjach finansowych*, Wydawnictwo C.H.Beck, Warsaw, 2010.

[6] *Nowoczesne zarządzanie projektami*, red. M. Trocki, PWE, 2012.

[7] *Handbook on Business Process Management. Introduction, Methods, and Information Systems*, red. J. vom Brocke, M. Rosemann, Springer-Verlag, Heidelberg, 2010.

[8] Chmielarz W.: *Zarządzanie projektami @ rozwój systemów informatycznych zarządzania*, Wydawnictwo Naukowe Wydziału Zarządzania UW, Warsaw, 2013 (in print).

[9] Stępień P.: *Wprowadzenie do zarządzania projektami*, part I, http://www.skutecznyproject.pl/ artykul.htm?AID=65, February 2012.

[10] The Standish Group, *Chaos Summary,* West Yarmoth, Massachusetts 2009, str. 1, The Standish Group International, Incorporated, CHAOS Report 2009, http:/blog.standishgroup.com/news, http://www.controlchaos.com/storage/S3D%20First%20Chapter.pdf, November 2012.

[11] BPTrends Surveys: a. *Business Process Centers of Excellence* – 2012 www.bptrends.com; b.*The State of Business Process Management* – 2012 www.bptrends.com.

[12] http:// www.bpmwpolsce.pl, April-May, 2013,

[13] Burlton R.: *Business Process Management, Profiting from Process,* Sams Publishing 2001.

[14] Harmon P: *Business Process Change. A guide for Business Managers and BPM and Six Sigma Professionals,Morgan Kaufmann Publishers, 2007.*

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1087–1094

Process-based evaluation and comparison of OTS software alternatives

Maria Jesus Faundes, Hernan Astudillo and Bernhard Hitpass
Universidad Técnica Federico Santa María, Departamento de Informática, Avenida Vicuña Mackena 3939,
San Joaquín, Santiago, Chile
Email: mariajesus.faundes@usm.cl, hernan@inf.utfsm.cl, bernhard.hitpass@usm.cl

***Abstract*—Many Off-The-Shelf Software (OTSS) assessment techniques have been proposed, most of them using criteria related to standard quality models. However, these techniques are not as useful to evaluate and compare alternative OTSS as solutions to specific process-driven organizational changes. This article proposes PBEC-OTSS (Process-Based Evaluation and Comparison of OTSS), a technique for evaluating and comparing OTSS regarding impact in the organization, based on process models, and using fuzzy decision making systems. The technique was compared with an Ad-Hoc approach (systemized from the literature) in an experimental study with IT professionals, some new to BPM and some experts; the experts obtained similarly good results with either approach, but the novice professionals obtained better results with PBEC-OTSS than with Ad-Hoc. These results suggest that organizations can improve their Business/IT alignment with this technique even if no process experts are available.**

I. Introduction

THE software selection process is a subjective and uncertain decision process [23], where meeting the specific needs of the organization is complex and requires time [8], considering that a company may lose its competitive advantage by investing in wrong alternative technologies or by investing too much time in selecting the right one [18].

In practice, there are numerous organizations that lack a rigorous selection process [11], which is often made under pressure by evaluators who may not have time or the expertise to plan it [8], or that select according to their experience or intuition [20], [21].

A systematic and repeatable selection methodology for *Off-The-Shelf Software* (OTSS), is a crucial need to minimize uncertainty and risk in companies [19]; so, choosing a suitable OTSS, is a non-trivial task for organizations and requires a thorough assessment process. Therefore, the key question seems to remain: how to identify the most appropriate OTSS to meet organizational needs? The recommendation to managers is to choose an appropriate IT infrastructure to facilitate the alignment between strategy and organizational structure, achieving higher performance levels [4]. However, the benefits of an organization's systems are generally known only after some period of use [2].

In the literature, there are studies that emphasize the important relationship between business processes and IT, stating: IT will be used if, and only if, the functions available for the user, support, or fit to their activities [5]. The selection and implementation of proper IT applications, is an important precondition for the efficient execution of business processes [17]. IT will only have a positive effect in the organizational performance, if this fits with the business processes [9].

Therefore, based on the close relationship between business processes and IT, is that this study proposes a new approach to evaluate and compare OTSS, based on processes models and fuzzy decision making systems, which allows:

- Generate alternative reconfigurations of processes models, which serve as input data for processes improvement and processes standardization.
- Identify and measure the potential contribution of the OTSS to the organization, through impact indicators (coverage, automation, and implementation).
- Generate OTSS traceability regarding the processes and activities of the organization.
- Identify collaboration between OTSS, and implemented systems in the organization.

All of which improves decision making, and promotes the Business/IT alignment.

The remainder of this paper is structured as follows: section II presents the Related Work; in Section III the proposed approach, along with an illustrative example; in section IV the Case Study, and finally Section V summarizes and concludes.

II. Related Work

In literature, the software assessment has been a subject of interest of suppliers, and topic of study for academics and professionals, developing and proposing:

- Preliminary proposals, such as: general recommendations for evaluation of commercial software [3], generation of Domain-specific quality model to assess software [7], and approach for determining the software selection strategy [25].
- Proposals for the evaluation of specific types of software, such as: ERP [6], data warehouse system [13], and workflow type software [14], among others.
- Frameworks, methods, and tools of assessment and/or selection of OTSS (see Table I). However, in most of these proposals, criteria are repeated or are similar, and related to standard quality models. Unfortunately, criteria associated to software quality evaluate the product and its interaction with the user, but ignore the importance

978-83-60810-53-8/$25.00

of the contribution that assessed system can make to the organization [1]. Therefore, choosing between a proposal and the other, apparently, depends on the value that is given to: application requirements (time, effort, and tools), evaluation criteria, and the complexity of each.

- Methods of Commercial Off-the-shelf (COTS) Selection, such as [12]: OTSO (1995), PORE (1998), STACE (1999), PECA (2002), and CARE (2004), among others. However, these are specific for components, were created for the software development, and do not seem to be adaptable to the different domains and projects [16].
- Assessment services and online software comparison: TEC[1] and Capterra[2]. Although a comprehensive review is allowed, the evaluation criteria are similar to those of Table I, or are wielded in a generical and superficial manner, as is the case of processes.

Therefore, although there are many evaluation proposals, it can be appreciated that there is no evaluation technique based on process models able to allow assessing and comparing OTSS, which could lead to identify their potential effect in the organization.

III. Process-Based Evaluation and Comparison of OTSS

This section presents the PBEC-OTSS (Process-Based Evaluation and Comparison of OTSS) technique, which consists of five stages, and considers six evaluation criteria (see Figure 1), defined from the process perspective.

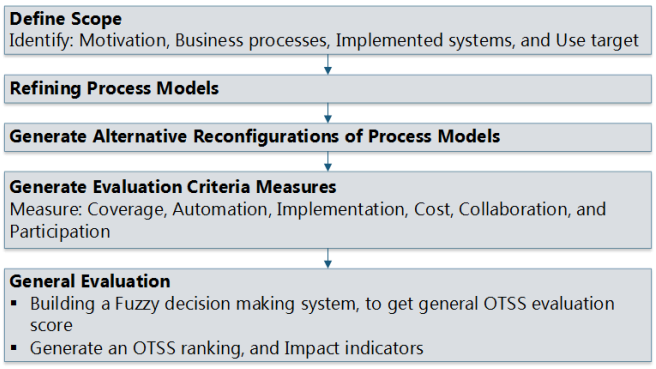


Fig. 1. Process-Based Evaluation and Comparison of OTSS

The details of the stages of the proposal are hereunder described, with an illustrative example, which is based on a real scenario of a Chilean public sector institution, *Solidarity and Social Investment Fund* (FOSIS[3]). The results correspond to evaluation performed by an BPM and IT expert, who was also Case Study expert (see section IV).

A. Define Scope

To establish the assessment dimension, identify: (i)*Motivation*: problems or situations that are intended to be solved or improved, (ii) *Business Processes*, (iii) *Implemented systems*, that participate in the identified processes, and are involved with motivation, and (iv) *Use Target* of each implemented system; i.e., according to the motivation identified, and as a complementary measure, to establish whether to decrease, maintain, or increase current usage level of each implemented system.

Example. (i) *Motivation*: Problems between different areas of the organization with HR, during recruitment. Current Situation: Extensive process cycle, exceeding by 75% the established time limit. Low process standardization, varying its implementation in different regions of the country. High effort, and null automation. High loss information, lack of centralized repository for generated documentation. Low standardization of generated documentation. No versioning of created documents; high dependence of general purpose tools. Therefore, the organization wants to buy a new tool to help them with the problems identified; (ii) *Business process*: Recruitment; (iii) *Implemented systems*: Microsoft Office (Word, Excel, and Access), and e-mail, and (iv) *Use target*: decrease the use of both systems.

B. Refining Process Models

Review process models identified in previous stage. If implemented systems are not specified in process model, or documented as a black box, process model should be refined, specifying for each actor: user's activities, automatic activities, and manual activities (not performed on any system).

Example. Figure. 2 illustrates refined Recruitment process model.

C. Generate Alternative Reconfigurations of Process Models

Review documentation, and identify OTSS functionalities. Subsequently, for each identified process model in previous stage, generate a new process model including OTSS. Each process activities that can cover the OTSS, will exclusively depend on the features identified in the revised documentation. Activities cannot be modified (name and quantity), and automatic activities must be a explicit OTSS feature.

Example. OTSS assessed are PeopleNet Recruitment[4] and Email2DB[5]. Reconfigurations generated are displayed in Figure 3 and 4.

D. Generate Evaluation Criteria Measures

To generate evaluation criteria measures, all models and reconfigurations generated in the previous stages should be considered, and calculate the different types of activities, as specified in Table II.

Example. Table III presents evaluation criteria results, considering Figure 2, 3 and 4.

TA: Total of activities of refined models, **TUA:** Total of user's activities of refined models, **TUAr:**Total of user's activities of alternative reconfigurations, **TAA:** Total of automatical

[1]http://www.technologyevaluation.com
[2]http://www.capterra.com
[3]http://www.fosis.gob.cl/
[4]http://www.meta4.com/solutions/105/personnel-selection-recruitment.html
[5]http://www.email2db.com

TABLE I
OTSS FRAMEWORKS, APPROACHES AND ASSESMENT TOOLS

Approach	Input	Assessment Criteria	Output
Expert System for Software Evaluation (ESSE) [23]	(i) Set of OTSS (ii) OTSS documentation	(i) Hardware, (ii) Software, (iii) Legacy software porting,(iv) Network management software, (v) Training, (vi) Maintenance, (vii) Company profile, (viii) Miscellaneous.	(i) Best subset of OTSS, or (ii) Subset of OTSS grouped according to classification, (iii) Ranking of OTSS, or (iv) Formal description of each OTSS (without ranking).
Five-phase COTS selection model [21]	(i) Set of OTSS (ii) OTSS documentation	(i) Cost, (ii) Supplier's support, (iii) Technological risk, (iv) Closeness of fit to the company's business, (v) Easy of implementation, (vi)Flexibility to easy change as the company's business changes, (vii) System integration.	Ranking and qualification of OTSS
Enterprise COTS software analyzer [10], [11]	(i) Set of OTSS (ii) OTSS documentation	(i) Functionality, (ii) Reliability, (iii) Cost, (iv) Ease of customization, (v) Ease of use.	Qualification of each assessed OTSS
Enterprise Software Selection Method (ESSM) [19]	(i) Set of OTSS (ii) OTSS documentation	(i) Functional requeriments, (ii) Non-functional requeriments, related to: Quality characteristics, Technology factors, and Socio-economic factors, (iii) Total cost, (iv) Implementation time.	Selected software
Decision making framework for software selection [8]	(i) Set of OTSS (ii) OTSS documentation	(i) Functional, (ii) Technical, (iii) Quality, (iv) Vendor, (v) Output, (vi) Cost and benefit, (vii) Opinion.	Ranking and qualification of OTSS

TABLE II
EVALUATION CRITERIA MEASURES

Criteria	Measure
Coverage: OTSS user's activities rate	$\frac{(TUA_r)}{TA} x100$
Automation: Variation of Automatic Activities	$\frac{(TAA_r - TAA)}{TA} x100$
Implementation: Variation of Manual Activities	$\frac{(TMA - TMA_r)}{TA} x100$
Cost: Cost Qualification	0 (very low) - 1 (very high)
Collaboration: Average of compliance of implemented systems (equal to 0% if implemented systems do not meet the use target)	$\frac{\sum_{i=1}^{n} \left\lvert \left[\frac{TUA_r - TUA}{TA} x100 \right]_{IS_i} \right\rvert}{n}$
Participation: Average of non-compliance regarding the use of implemented systems (equal to 0% if implemented systems meet the use target).	$\frac{\sum_{i=1}^{n} \left\lvert \left[\frac{TUA_r - (TUA+G)}{TA} x100 \right]_{IS_i} \right\rvert}{n}$

TABLE III
EVALUATION CRITERIA RESULTS

Criteria	Measure	
	PeopleNet	Email2DB
Coverage	61%	39%
Automation	39%	30%
Implementation	13%	0%
Cost	0.6	0.6
Collaboration	44%	35%
Participation	0%	0%

activities of refined models, **TAAr**: Total of automatical activities of alternative reconfigurations; **TMA:** Total of manual activities of refined models, **TMAr:** Total of manual activities of alternative reconfigurations; **IS:**for each implemented system (of an n total), **G:** Use target (if it is to decrease = -1; if it is to maintain = 0; if it is to increase = 1)

E. Introduction to Fuzzy Logic and Fuzzy Decision Making System

For a better understanding of last stage of the proposal, it is here presented an introduction to the basic concepts of Fuzzy logic [26] and Fuzzy decision making system [15].

Definition 1. *Fuzzy sets*: The fuzzy set theory was proposed in 1965 by Zadeh, where he defines that a fuzzy set is characterized by a membership function that maps the elements of a universe of X discourse to a unit interval [0,1].

Definition 2. *Linguistic variable*: A linguistic variable is a fuzzy variable whose values are categories represented by fuzzy sets. The value of a linguistic variable is a compound term $T(x) = \{L_1, L_2, \ldots, L_n\}$.

Definition 3. *Membership function*: Any function of the form μ_A:X→[0,1] describes a membership function associated with a fuzzy set A. The shape of the membership function depends on the concept to represent and the context in which it is used.

Definition 4. *Fuzzy If-then rules*: They have the structure *If* (x is A_i) and (y is a B_j) *then* (z is C_k)

Definition 5. *Fuzzy decision making system*: Corresponds to a system that is responsible for mapping an input space to a determined output space using fuzzy logic. It comprises four components (see Fig. 5), hereunder detailed.

Fuzzification interface. Inputs are identified, and through membership function, it can be established the degree of membership of each input to the relevant fuzzy set.

Knowledge base. A database that consists of the expert

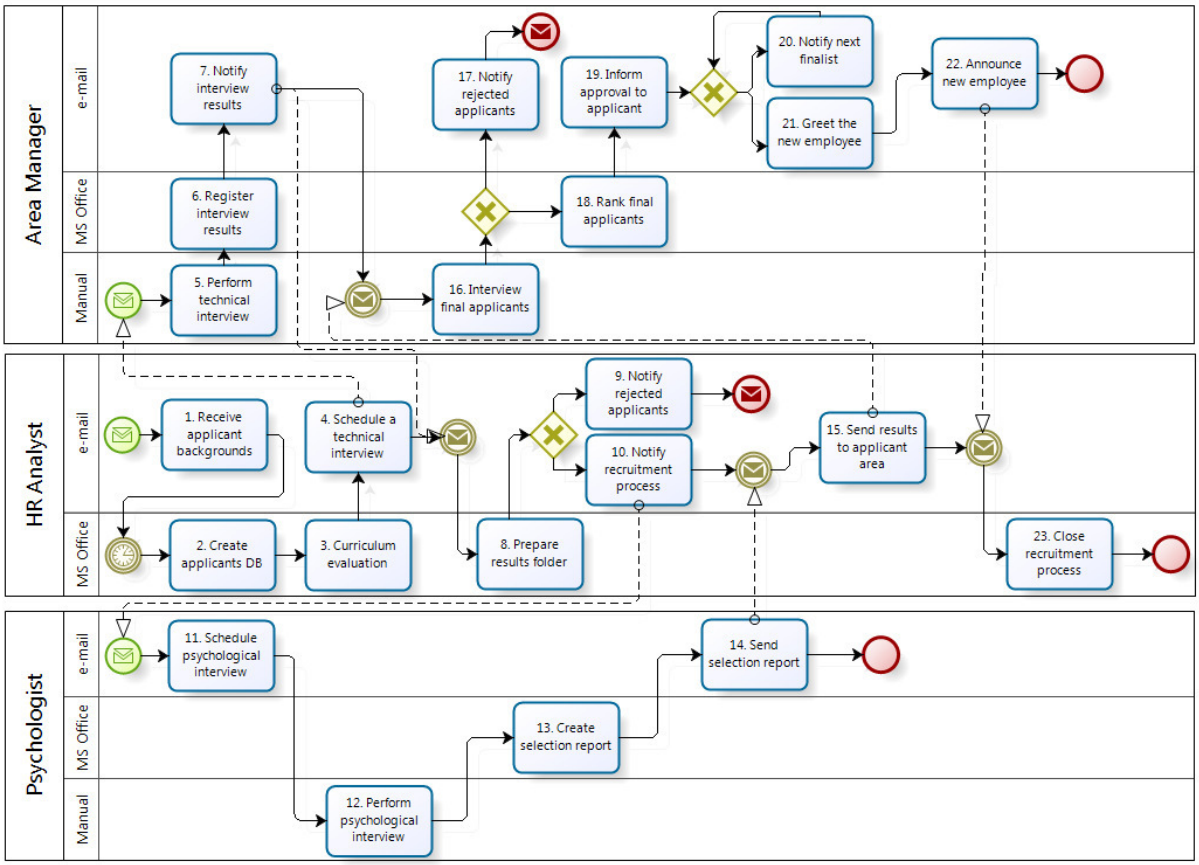


Fig. 2. Refined Recruitment process model

knowledge of the domain in question. In this database are defined the parameters of the membership function, linguistic categories of input and output variables, and set of rules.

Decision making unit. Simulates human decision making, by performing inference using the *fuzzy If-then rules*. For this purpose, first are identified the antecedents of each rule and their relationship; i.e., the various values that every input can take are identified, and related to each other, through logical operators. Subsequently, each antecedent is related to a particular consequent (fuzzy set represented by membership function), by implication operators. Finally, the fuzzy sets representing the output of each rule are grouped in a single fuzzy set.

Defuzzification interface. To the fuzzy set resulting from the previous stage, it is applied a defuzzification method to obtain a single value, which corresponds to the final result.

F. General Evaluation

Building a Fuzzy decision making system, to get general OTSS evaluation score. Following is part of used structure.

Fuzzification interface and Knowledge base. The inputs correspond to evaluation criteria results obtained in the previous stage. Based on the knowledge and experience of two BPM experts, it was defined for each input, linguistic terms, their membership functions, and domain (see Table IV).

Decision making unit. Antecedents of fuzzy rules are: Coverage and Automation and Cost; Coverage and Automation and Implementation; Implementation and Cost; Collaboration and Participation. Consequent for all rules is Evaluation.

Defuzzification interface. Defuzzication method used is centroid, and output corresponds to OTSS evaluation result. Finally, according to OTSS evaluation result, generate an OTSS ranking, indicating percentages obtained by Coverage, Automation, and Implementation criteria, which together correspond to *Impact Indicators*.

Example. Table V presents OTSS evaluation result, delivered by fuzzy decision making system, considering as input the evaluation criteria results of Table III.

G. PBEC-OTSS Benefits

The main benefits of PBEC-OTSS are: (i) favors Business/IT alignment, (ii) improves decision making, i.e., decision making is well informed, and with less uncertainty, (iii) provides new information, which can serve as an antecedent for reducing time and efforts.

IV. Case Study

The purpose of the case study was to evaluate quality of results, duration, and use satisfaction, of PBEC-OTSS. In this regard, it was compared with an Ad-Hoc approach,

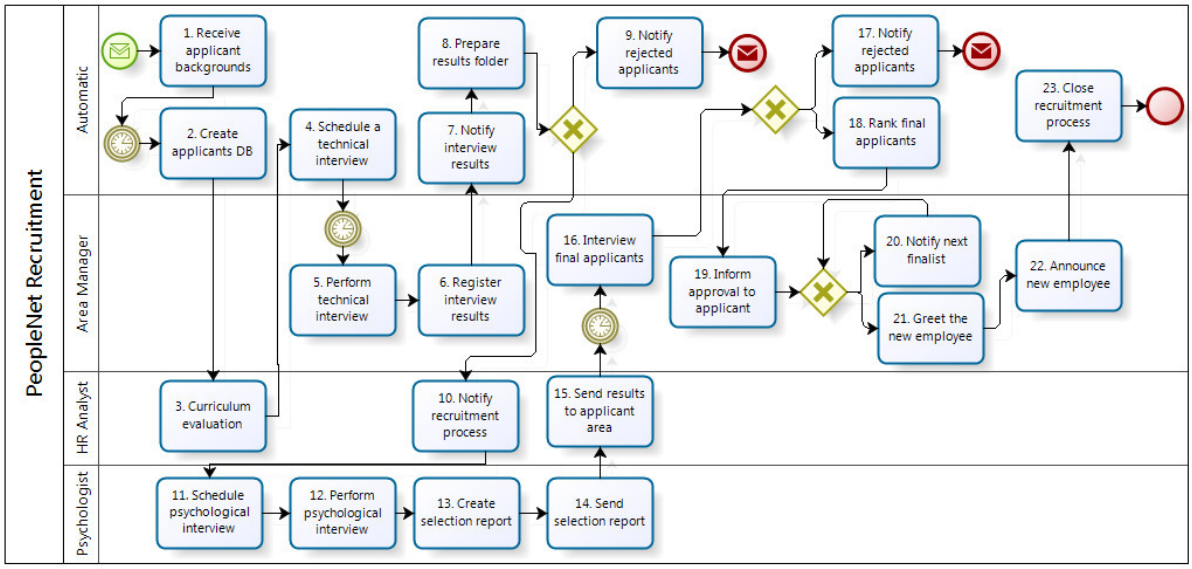


Fig. 3. Reconfiguration of Recruitment process model including PeopleNet Recruitment

TABLE IV
KNOWLEDGE BASE

Input	Linguistic term	Membership function	Domain
Coverage	Very low	Trapezoidal	0% - 21%
	Low	Trapezoidal	19% - 41%
	Medium	Trapezoidal	39% - 61%
	High	Trapezoidal	59% - 81%
	Very high	Trapezoidal	79% - 100%
Automation	Very high decrease	Trapezoidal	-100% - -29%
	High decrease	Trapezoidal	-31% - -19%
	Medium decrease	Trapezoidal	-21% - -9%
	Low decrease	Trapezoidal	-11% - 0%
	Equivalent	Triangular	-2% - 2%
	Low increase	Trapezoidal	0% - 11%
	Medium increase	Trapezoidal	9% - 21%
	High increase	Trapezoidal	19% - 31%
	Very high increase	Trapezoidal	29% - 100%
Implementation	High decrease	Trapezoidal	-100% - -29%
	Medium decrease	Trapezoidal	-31% - -14%
	Low decrease	Trapezoidal	-16% - 0%
	Equivalent	Triangular	-2% - 2%
	Low increase	Trapezoidal	0% - 16%
	Medium increase	Trapezoidal	14% - 31%
	High increase	Trapezoidal	29% - 100%
Cost	Very low	Trapezoidal	0 - 0.21
	Low	Trapezoidal	0.19 - 0.41
	Medium	Trapezoidal	0.39 - 0.61
	High	Trapezoidal	0.59 - 0.81
	Very high	Trapezoidal	0.79 - 1
Collaboration Participation	Null	Triangular	0% - 1%
	Low	Trapezoidal	1% - 10%
	Medium	Trapezoidal	11% - 40%
	High	Trapezoidal	41% - 100%

through an experimental study, where to each approach was simultaneously applied [24].

TABLE V
OTSS EVALUATION RESULTS

No.	OTSS	Impact Indicators
1	PeopleNet **82.5**	**61%** Coverage, **39%** Automation, **13%** Implementation
2	Email2DB **73.5**	**39%** Coverage, **30%** Automation, **0%** Implementation

A. Approaches to Experimental Study

PBEC-OTSS. The aim of this approach is to evaluate and compare OTSS using processes models, and fuzzy decision making systems. The result is OTSS ranking, and impact indicators.

Ad-Hoc Approach. The objective of this [10], [11] is to evaluate OTSS through five criteria (Functionality, Reliability, Cost, Ease of customization, and Ease of use). The result is a score for each OTSS analyzed.

B. Experimental Study Design

Experimental study was designed to be done by professionals who finished a *Diploma in Process Management and IT* (196 hrs. postgraduate program taught by a Chilean university). Of all participants, two groups were formed. Simultaneously, one group applied PBEC-OTSS and other applied Ad-hoc approach. The assignment to each group was random, but mark and professional experience of each participant were considered. Thus, they sought the greatest homogeneity between groups, and prevented the influence on the results of factors such as experience or knowledge.

C. Instrumentation

Materials used by participants of both groups were the same: Instructive, Personal Data Questionnaire, Problem Definition,

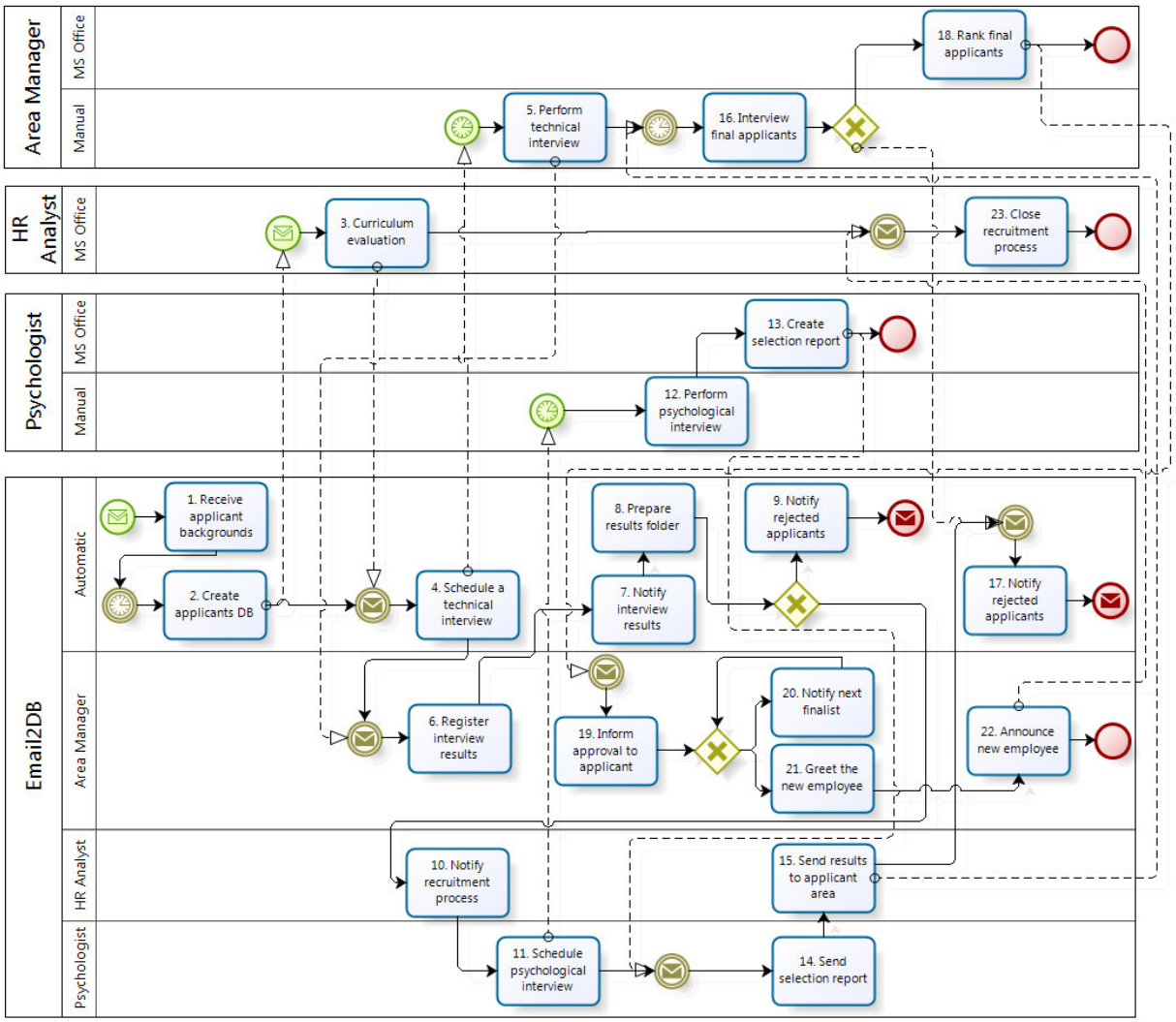


Fig. 4. Reconfiguration of Recruitment process model including Email2DB

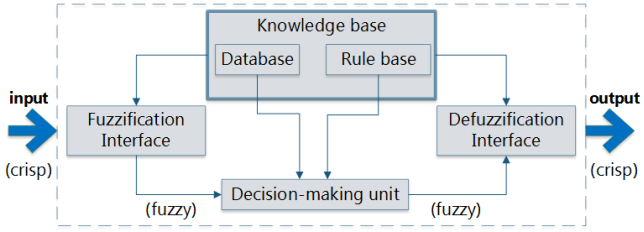


Fig. 5. Fuzzy decision making system [22]

Recruitment Process Description and Model, OTSS Documentation, How-to approach, Questionnaire on applied assessment approach. For both approaches, the problems, process, and OTSS to evaluate were the same as those presented in Section III example, which was adapted to be resolved in a maximum of two hours (time calibrated by previous execution of a pilot experiment). In addition, the OTSS real name was omitted, to avoid influence on the results, due by possible foreknowledge of these.

D. Experimental Study Execution

The participants were 13 Chilean IT professionals, who held positions as Software Engineer, Project Manager, Process Engineer, or Consultant, and Area Assistant Manager. Six of these used PBEC-OTSS, and seven Ad-hoc approach. None of them had previous experience in any of two approaches. Table VI presents group and expert characteristics (in average years).

Each group was trained in its respective approach. The training was conducted in parallel, same day and time, but in different rooms and with different instructors. The results of each approach were compared with results obtained by an BPM and IT expert, who performed the same experiment (see results in Section III example).

TABLE VI
GROUP AND EXPERT CHARACTERISTICS

Characteristic	PBEC-OTSS	Ad-Hoc	Expert
Age	32.3	30.6	56
Professional experience (total)	6.8	6.6	26
Process modeling experience	2.2	2.6	20
Experience about IT decision	1.3	1.4	15

E. Experimental Results and Analysis

Table VII presents participants and expert results. Table VIII shows average duration and average rating of the application of each approach (where 1 is very low and 5 is very high).

TABLE VII
PARTICIPANTS AND EXPERT EXPERIMENTAL RESULTS

	PeopleNet Recruitment		Email2DB	
	PBEC-OTSS	**Ad-Hoc**	**PBEC-OTSS**	**Ad-Hoc**
Subject 1	82.5	72.9	82.5	57.2
Subject 2	82.5	78.4	79.5	69.1
Subject 3	85.5	78.7	73.5	45.2
Subject 4	82.5	75.5	79.5	42.4
Subject 5	73.5	80.3	67.5	58.3
Subject 6	82.5	78.7	82.5	34.7
Subject 7	-	79.2	-	66.8
Average	81.5	77.7	77.5	53.4
Standard Deviation	4.1	2.6	5.9	12.9
Expert	**82.5**	**79.4**	**73.5**	**71.0**

TABLE VIII
DURATION AND QUALIFICATION OF APPROACHES

	PBEC-OTSS	Ad-Hoc
Duration	56.8 minutes	45.7 minutes
Difficulty	3.67 points	3.71 points
Satisfaction	3.67 points	3.57 points
Recommendation	100% yes	86% yes, 14% no

The results in Table VII and VIII show that:

- For PeopleNet Recruitment, both approaches generated similar results with respect to expert, with an acceptable variation of 1% for PBEC-OTSS, and variation of 2% for Ad-Hoc.
- For Email2DB, approaches differ, PBEC-OTSS presents better results than Ad-hoc approach. Respect to expert, PBEC-OTSS has 5% variation, while Ad-hoc approach has 25% variation.
- Regarding duration, PBEC-OTSS takes 11 minutes more than Ad-Hoc, this could be due to greater degree of PBEC-OTSS difficulty. Despite this, average satisfaction qualification, and recommendation rate is higher for PBEC-OTSS.
- Comparing both approaches, expert results are similar, this could be due both techniques have a expert knowledge base. PBEC-OTSS uses fuzzy decision making systems built with parameters identified by two BPM specialists (different to case study expert), and Ad-Hoc approach was based on a survey of MIS managers.

Therefore, for complex cases (Email2DB), PBEC-OTSS reduces the risk of an inadequate evaluation, by presenting reasonable results and close to an expert. While for simple cases (PeopleNet Recruitment), both approaches are similar.

V. CONCLUSION

It was proposed PBEC-OTSS technique, that allows evaluate and compare OTSS, measuring its possible contribution to an organization. This proposal considers six criteria (Coverage, Automation, Implementation, Cost, Collaboration, and Participation), based on process models and fuzzy decision making systems. The main benefits of PBEC-OTSS are: favors the Business/IT alignment, and improves decision making.

PBEC-OTSS was compared with an Ad-Hoc approach, in an experimental study, based on real scenario of a Chilean public sector institution. 13 IT professionals participated, and a BPM and IT expert. The professionals results showed that for simple cases PBEC-OTSS is similar to Ad-hoc approach, while for complex cases, the techniques differ, being PBEC-OTSS the best. Additionally, it was observed that when applying both approaches (PBEC-OTSS and Ad-Hoc) by an expert, the results were very similar, which could infer that PBEC-OTSS results, appear to be comparable with a technique from the literature. Although experimental study was conducted with two specific software, PBEC-OTSS is applicable to other types of software, and in any type of organization, because PBEC-OTSS focuses on Business/IT alignment.

Finally, based on the study and analysis fulfilled, we identified the following topics of interest to perform as future work: adaptation of the technique proposed so that the assessment to include different levels of activities importance, inclusion of new experts to redefine Fuzzy decision making system configuration parameters, and new experimental study with a greater number of professionals, experts, and other real scenario of an public or private sector institution.

ACKNOWLEDGMENT

Marina Pilar, Esteban Romero, and Cristobal Castillo.

This work has been partly supported by projects ADAPTE (Fondef D08i1155), CCTVal (Conicyt Basal FB0821), DGIP 241250, and BPM Center (UTFSM).

REFERENCES

[1] G. Boloix and P. Robillard. A software system evaluation framework. *Computer*, 28(12):17–26, 1995.
[2] E. Brynjolfsson. The productivity paradox of information technology. *Commun. ACM*, 36(12):66–77, Dec. 1993.
[3] D. J. Carney and K. C. Wallnau. A basis for evaluation of commercial software. *Information and Software Technology*, 40(14):851 – 860, 1998.
[4] P. D. Chatzoglou, A. D. Diamantidis, E. Vraimaki, S. K. Vranakis, and D. A. Kourtidis. Aligning IT, strategic orientation and organizational structure. *Business Process Management Journal*, 17(4):663–687, 2011.
[5] M. T. Dishaw and D. M. Strong. Extending the technology acceptance model with tasktechnology fit constructs . *Information & Management*, 36(1):9 – 21, 1999.
[6] B. E. and K. S. ERP selection process in midsize and large organizations. *Business Process Management Journal*, 7(3):251–257, 2001.
[7] X. Franch and J. Carvallo. Using quality models in software package selection. *Software, IEEE*, 20(1):34–41, 2003.

[8] A. S. Jadhav and R. M. Sonar. Framework for evaluation and selection of the software packages: A hybrid knowledge based system approach. *Journal of Systems and Software*, 84(8):1394 – 1407, 2011.

[9] J. Karim, T. Somers, and A. Bhattacherjee. The impact of erp implementation on business process outcomes: A factor-based study. *J. Manage. Inf. Syst.*, 24(1):101–134, July 2007.

[10] M. Keil and A. Tiwana. Beyond cost: the drivers of COTS application value. *Software, IEEE*, 22(3):64–69, 2005.

[11] M. Keil and A. Tiwana. Relative importance of evaluation criteria for enterprise systems: a conjoint study. *Inf. Syst. J.*, 16(3):237–262, 2006.

[12] R. Land, L. Blankers, M. Chaudron, and I. Crnković. COTS Selection Best Practices in Literature and in Industry. In *Proceedings of the 10th international conference on Software Reuse: High Confidence Software Reuse in Large Systems*, ICSR '08, pages 100–111, Berlin, Heidelberg, 2008. Springer-Verlag.

[13] H.-Y. Lin, P.-Y. Hsu, and G.-J. Sheen. A fuzzy-based decision-making procedure for data warehouse system selection. *Expert Syst. Appl.*, 32(3):939–953, Apr. 2007.

[14] P. M. and R. T. Evaluation of Workflow-type software products: a case study. *Information and Software Technology*, 42(7):489–503, 2000.

[15] E. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1 – 13, 1975.

[16] A. Mohamed, G. Ruhe, and A. Eberlein. COTS Selection: Past, Present, and Future. In *Engineering of Computer-Based Systems, 2007. ECBS '07. 14th Annual IEEE International Conference and Workshops on the*, pages 103–114, 2007.

[17] T. Neubauer. An empirical study about the status of business process management. *Business Process Management Journal*, 15(2):166–183, 2009.

[18] R. F. Saen. A decision model for selecting slightly non-homogeneous technologies. *Applied Mathematics and Computation*, 177(1):149 – 158, 2006.

[19] C. G. Sen, H. Baracli, S. Sen, and H. Basligil. An integrated decision support system dealing with qualitative and quantitative objectives for enterprise software selection. *Expert Syst. Appl.*, 36(3):5272–5283, Apr. 2009.

[20] N. Shehabuddeen, D. Probert, and R. Phaal. From theory to practice: challenges in operationalising a technology selection framework. *Technovation*, 26(3):324 – 335, 2006.

[21] H.-J. Shyur. COTS evaluation using modified TOPSIS and ANP. *Applied Mathematics and Computation*, 177(1):251–259, 2006.

[22] S. Sivanandam, S. Sumathi, and S. N. Deepa. Introduction to Fuzzy Logic using MATLAB. 2007.

[23] I. Vlahavas, I. Stamelos, I. Refanidis, and A. Tsoukis. ESSE: an expert system for software evaluation. *Knowledge-Based Systems*, 12(4):183 – 197, 1999.

[24] C. Wholin, P. Runeson, M. Host, M. Ohlsson, B. Regnell, and A. Wesslen. Experimentation in Software Engineering. 2000.

[25] M. Wybo, J. Robert, and P.-M. Léger. Using search theory to determine an applications selection strategy. *Inf. Manage.*, 46(5):285–293, June 2009.

[26] L. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1095–1102

Multi-attribute Reverse Auctions and Negotiations with Verifiable and Not-verifiable Offers

Gregory (Grzegorz) E. Kersten
InterNeg Research Centre
Concordia University
Montreal, Canada
gregory@jmsb.concordia.ca

Tomasz Wachowicz
Department of Operations Research
University of Economics
Katowice, Poland
tomasz.wachowicz@ue.katowice.pl

Margaret Kersten
SLaLS, Carleton University
Ottawa, Canada
margaret.kersten@carleton.ca

***Abstract*—Comparative studies of auction and negotiation exchange mechanisms have typically compared the outcomes obtained from the two mechanisms. Their result are inconclusive. The question which this paper aims to address is the viability of outcome-based comparisons. Such comparisons assume that both mechanisms produce the same types of outcomes but their values differ. An argument can be made that this is not necessarily the case. Based on several experiments of multi-attribute auctions and two formats of multi-bilateral negotiations the paper argues that the two mechanisms produce some outcomes which are comparable and other outcomes which are qualitatively different. A surprising finding of our experiments is that the outcomes of the non-verifiable negotiations were more similar to the outcomes of the reverse auctions than to the verifiable negotiations, despite the fact that the latter employ rules taken from the auction mechanism.**

I. Introduction

AUCTIONS and negotiations are exchange mechanisms used by individuals and institutions including, among others businesses and governments. There are many kinds of auctions and negotiations and their underlying rules and regulations are well established, leaving however, some space for adjustment in individual cases. The opportunities created by the internet technologies and the web engaged researchers in captivating discussions about the viability of conducting business transactions over internet. Early on, arguments were made that on line auctions will replace negotiations, that there will be a paradigm shift where market forces will replace the more subtle domain of negotiation skills [1-3]. In 2000, a group of researchers gathered in Montreal to discuss the issue. They concluded that there are limits to auctions and consequently not all electronic transactions lend themselves to auctions [4]. One of the outcomes of the Montreal workshop was a framework for designing e-negotiations [5].

In order to gain some clarity about the strengths and weaknesses of these two types of transaction mechanisms it is necessary to compare them. Broadly speaking, there are two types of negotiations, i.e. bilateral and multilateral (each type can be either simultaneous or sequential). There are different types of auctions, including price-only and multi-attribute forward and reverse auctions. The discussion here is restricted to multi-bilateral reverse auctions, henceforth called auctions.

Bilateral negotiations appear to be comparable to auctions. Experimental studies [6, 7] and field studies [8-10] comparing auctions with bilateral negotiations indicated that auctions are used in different situations than negotiations, namely, auctions are used when: (1) the exchanged goods (services) have only one attribute – price; (2) there are several (possibly many) suppliers of the good (service); and (3) there is no need to communicate (exchange information). Negotiations, on the other hand, are used when: (1) one or more of the above conditions do not apply; and (2) there is a strong likelihood of future interaction.

It is difficult to compare auctions and negotiations because they are very different mechanisms—on the general level the assumptions underlying each mechanism differ significantly, on the specific level there are differences in participants' knowledge and behavior. Auctions involve multiple bidders who compete for the same good or service; it is assumed that that bidders follow a strict, fixed protocol and that they know the buyer's price (valuation). Other than submitting bids, there is no other form of communication. In contrast, negotiations rely on significantly weaker assumptions; it is assumed that the parties negotiate in good faith and that each party has preferences so that they may compare alternatives. No assumption about the sellers' knowledge of the buyer's valuation is made.

Another reason why it is difficult to compare auctions and negotiations are the differences in settings and protocols. Studies mentioned above compared auctions with *N* bidders with bilateral negotiations (i.e., 1:1). In this situation the competition among sellers, which is a key characteristics of reverse auctions, disappears in negotiations with only one seller. In order to maintain competition among sellers in negotiations, bilateral negotiations need to be replaced with

This work has been supported by the grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), Carleton University, and Concordia University.

978-83-60810-53-8/$25.00

multi-bilateral negotiations, in which a single buyer negotiates with many sellers. Furthermore, these negotiations should be simultaneous rather than sequential so that the organization of both auction and negotiation processes are similar.

Thomas and Wilson [11-13] compared auctions and multi-bilateral negotiations. They set up several experiments making sure that for the sake of comparison the mechanisms were structurally similar, namely, there were N participants in auctions and N:1 participants in multi-bilateral negotiations. The experiments lasted 4 minutes and involved a single attribute–price. Despite the particular setting, the auctions' outcomes were not significantly better than the negotiations' outcomes.

Comparative studies of auctions and negotiations are not only difficult, they are also inconclusive. Bulow and Klemperer [14] showed that simple English auction with N+1 bidders (buyers) always yields higher revenue than a scheme they call "negotiation with N participants". Manelli and Vincent [15] demonstrated that the outcomes of auctions and negotiations depend on situations; they noted that in order to judge the effects of the two exchange mechanisms it is necessary to consider the overall context, including the goods, participants, market, and so on. They concluded that auction mechanisms are often inefficient in a procurement environment.

The above examples indicate that typically the process substantive outcomes, that is the values of the attributes, were used to compare auctions and negotiations. This paper also starts from this position. Excluding the discussion on the differences in organizing auctions and negotiations, it is important, from the pragmatic point of view, to identify the potential differences in results achieved if the same contract can be negotiated or established by means of an auction. Therefore, the first question is: *Is it worth spending time and money for often difficult negotiation rather than setting up an auction which reduces the bid-taker involvement in the process of contract designing to the minimum?*

In order to address this question, several experiments were conducted which compared multi-attribute reverse auctions and multi-bilateral negotiations. The auctions and the negotiations experiments comprise the first study discussed here.

In this first study experiments showed that the buyers received higher profits in auctions than in negotiations. Auctions were also more efficient. The possible explanation is that the negotiation protocol followed a typical rule and did not allow the sellers to obtain independent information about the best offer that the buyer received from one of the sellers. The auctions followed the rule that the winning (best) offer is displayed to all bid-makers (sellers). This makes auctions more transparent mechanisms than negotiations and could have placed auctions at an advantage over negotiations.

Thomas and Wilson [13] compared auctions and verifiable negotiations in which the best offer was displayed to all participants. Their experiments were stylized, meaning that there was no context, the sole issue was price, and the time allotted was four minutes. They used the sealed bid auction protocol which removed the dynamics of iterative multiple bids auctions and made it dissimilar to negotiations. Consequently, their results showed that for buyers, auctions produce better outcomes than both negotiation and that verifiable negotiations are better than non-verifiable. This latter result is surprising because it states "that providing sellers with more information about their rivals' price setting behavior unexpectedly leads to higher rather than lower prices" [13, p. 1030].

In our experiments the participants represent firms, their exchange problem is described in detail, there are three issues, and the allotted time is ten days. Following Thomas and Wilson's results we sought to address the following question: *Do multi-attribute verifiable negotiations produce worse results for sellers that the non-verifiable ones in a business context?*

In the second study we conducted exploratory experiment with the two types of negotiations. The result of the second study confirmed results obtained by Thomas and Wilson [13], i.e. the sellers were worse off in the negotiations with verifiable offers than negotiations with no verifiable offers. We were not able, however, to determine whether these results were statistically significant.

The third study builds on the previous two and addresses both questions formulated above. In this study we included auctions in order to compare the three mechanisms. The results of the third study confirmed the results obtained in the earlier two studies. This means that less transparent negotiations produce results which are better for the buyers (and worse for the sellers) and which are closer to the auction results than the results of more transparent negotiations.

The third study led to an observation that the buyers could run verifiable negotiations following the same protocol as the iterative auction protocol, which was not possible in Thomas and Wilson's [13] experiments. Consequently, in verifiable negotiations the buyers could achieve profits similar to the profits achieved by buyers in auctions. A possible reason for this outcome is that negotiation participants seek outcomes which are not possible to achieve in auctions. We conclude this paper with a discussion on the viability of outcome-based comparisons. When comparing outcomes, an assumption is made that both mechanisms produce the same types of outcomes and only their values differ. This is in contradiction with the social exchange theory.

II. Two Studies

Several experiments in which the participants used auction and negotiation web-based systems were conducted. These experiments and their results are briefly.

A. The case and two systems

In the Milika case a producer of perishable goods (the buyer) wants to sign a contract with one of the several logistics providers who offer their services. The minimum quantity of goods to be transported is a fixed part of the contract but three attributes must be negotiated, i.e. standard rate of transportation, rush rate for unexpected delivery, and penalty for the non-delivery or delivery of spoiled goods. Each attribute has a discrete number of options (fifteen per attribute) resulting in the total of 3375 possible agreements. All issues are specified and cannot be changed during the experiment.

The system relies on a single criterion used to compare alternative bids and offers, for example, utility, production, cost and profit functions. In the Milika case the selected function is quasi-linear and it describes profits of the buyer and the sellers. The profit function is different for different participants and its values (normalized between 0 and 100) are not disclosed to anyone.

Also the systems give the sellers breakeven points. Anything below these points means losses for the sellers' companies. This implies that the sellers should be careful not to cross these levels as their objective in both the auction and the negotiation is to obtain a contract that maximizes profit.

Two systems were used in the experiments—both implemented on the Invite e-negotiation system platform [16] -- (1) Imaras (InterNeg multi attribute reverse auction system); and (2) Imbins (InterNeg multi-bilateral negotiation system).

B. Study 1

There were six different lab and online auction and negotiation experiments. The results of these experiments are hardly comparable because of differences in: (1) the controlled variables, e.g., number of sellers (from two to six), number of alternatives (360 vs. 3375), and the participation of software agents (in one experiment); and (2) the process design, (e.g., fixed and flexible rounds, introduction of video, tests, and handouts). The results showed that, with the exception of one experiment, in auctions the sellers achieved very low and the buyers' very high profit. Table 1 illustrates two selected experiments.

TABLE 1.
STUDY 1: AUCTION AND NEGOTIATION EXPERIMENTS.

	Experiment 1		Experiment 2	
	Auction	Negotiation	Auction	Negotiation
No. of instances	17	40	27	23
No. of sellers	74	151	95	89
No. of offers (avg.)	4.4	3.0	**5.6**[*]	3.1
Agreement (%)	—	95	—	96
Seller's profit	3.9	19.9	**-7.4**[*]	23.4
Buyer's profit	66.9	52.6	**75.7**[*]	47.1
Dominating alt. (%)	6.4	1.9	4.0	4.0

[*] Significance compared to negotiations, $p < 0.01$

Sellers in the auctions made more offers than sellers in the negotiations. Their average profit was low, 3.9 in Experiment 1 and -7.4 in Experiment 2. In the latter experiment, the sellers' winning bid was (on average) a little below their breakeven value. In comparison, successful negotiators achieved a profit of 19.9 and 23.4, respectively in Experiment 1 and 2. In Table 1 we show that buyers achieved higher profit in auctions than they did in negotiations.

Table 1 also shows that the efficiency of the two mechanisms' is measured by the percent of alternatives which dominate the agreements. These results are not conclusive. In Experiment 1, auctions were less efficient than negotiations (6.4% of alternatives dominated the winning bids vs. and 1.9% of alternatives dominated agreements), while in Experiment 2 the two mechanisms were equaly efficient.

C. Subjective and objective concessions

An analysis of the results in Experiment 2 led to verification of the concession-making model in the auctions and the negotiations in which subjective and objective concessions were proposed [17]. The difference between these two types of concessions is the basis of comparison. A subjective concession is determined by two consecutive offers, i.e., made at t_1 and t_3 as shown in Fig. 1, both made by the same concession-maker. An objective concession is determined by two offers, the best offer on the table (market), which the concession-taker received at time t_2 from any concession-maker and the offer made at time t_3.

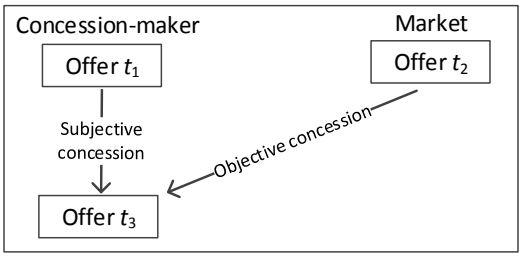


Fig. 1 Objective and subjective concessions

Subjective concessions occur in bilateral negotiations, in which both the concession-maker and the concession-taker can compare offers made by the same concession-maker. In multi-bilateral negotiations, in which one side is represented by many and the other side by a single negotiator (the case presented in Section II, A), objective concessions are possible. Their use requires significant transparency of the process and a fixed protocol, which typically are not employed. We know of only one negotiation study—done by Thomas and Wilson [13], in which objective concessions were made possible. In their study both the process and the systems were highly stylized and devoid of context.

Objective concessions are typical for these auctions in which the winning bid is shown to the bidders. Every bidder either submits a bid that is better (for the bid-taker) than the winning bid or drops out from the auction. The difference between the winning offer (on the market) and the submitted bid is the objective concession.

The sellers' profits given in Table 1 are the results of the concessions they made; the sellers made significantly greater concessions in the auctions than in the negotiations. The reason could be attributed to transparency: in the auctions the sellers knew the best bid, however, not in the negotiations. In the negotiations, even if the buyer sent information about the best offer she had received, this offer could not have been verified, hence the sellers may have considered it as a ploy. Realization of the above led us to a design of a negotiation experiment in which a version of the Imbins system displays the best offer on the table in the same way the Imaras does.

D. Study 2

Table 2 shows the results of the second study (Experiment

3). The column "Non-verifiable" results shows instances, when the Imbins system did not display the best offer; that is, the buyer could have shown the best offer but it could not have been verified by the sellers. The "Verifiable" column shows the results of the multi-bilateral negotiations, in which the system displayed the best offer.

TABLE 2.
STUDY 2: NEGOTIATION EXPERIMENT WITH VERIFIABLE AND NON-VERIFIABLE OFFERS.

	Experiment 3	
Best offers:	Verifiable	Non-verifiable
No. of instances	12	13
Sellers		
No. of sellers	33	35
No. of offers (avg.)	4.0	4.4
No. of offers w/out message (avg.)	1.0	1.2
No. of messages w/out offers (avg.)	1.3	0.6
Agreement (%)	100	92
Seller's profit	19.1	22.3
Dominating alt. (%)	4.0	4.0
Buyers		
No. of offers (avg.)	7.4	7.2
No. of offers w/out message (avg.)	1.0	1.5
No. of messages w/out offers (avg.)	2.9	1.4
Buyer's profit	53.3	48.0

Contrary to our expectations there were no significant differences between the two types of negotiations. We thought that the negotiations with verifiable offers would result in a significantly higher profit for the sellers than the negotiations with non-verifiable offers.

E. Discussion

The restrictions imposed on the negotiation protocol were severe but necessary. A fluid and evolving negotiation process with issues coming and going and preferences changing, cannot be compared with fixed protocol auctions. Verifiable offer negotiations have the same degree of transparency as auctions but they differ in the following three aspects:

1. The negotiating sellers are not forced to make positive objective concessions, i.e., make offers which are better for the buyer than the best offer on the table;
2. The negotiators can exchange messages with and without accompanying offers; and
3. The buyer can make offers.

The impact of the first difference needs to be further studied, but it does not appear to have potential for changing the process because both sides know about the best offer. Hence, sellers who (would) submit a worse offer than the best offer (make negative objective concessions) would do it knowing that the buyer has a better offer on the table. There may be, however, a good reason for these seller to do so, for example, if they offer some additional benefits for the buyer in the message that accompanies the offer.

The free-text communication with the buyer and the buyer's interaction with the seller are the remaining two key differences between auctions and negotiations (with fixed issues and options). Table 2 shows that in both the verifiable and the non-verifiable negotiations the sellers sent messages to the buyers (there were as many buyers as instances). About 75% of offers were accompanied by messages. In addition, every seller sent, on average, 0.6 messages in the non-verifiable negotiations and 1.3 messages to which no offer was attached.

The buyers used their ability to communicate with the sellers, as shown in Table 2. In the negotiation with non-verifiable best offer they made 7.2 offers, of which, on average, only 1.5 were without a message attached. They also sent 1.4 messages without an offer. The results are similar in the verifiable negotiation, with the exception of messages sent with no offer attached—2.9 on average, i.e., over twice as many as in the non-verifiable negotiation. This difference is attributed to two sellers who sent about four times more messages than other sellers. If we remove these two sellers from the dataset, then the averages are similar for both types of negotiations.

The number of offers made by the buyers is much greater than the number of offers made by the sellers because buyers made offers to three sellers, per instance (the number of sellers shown in Table 2 is smaller because inactive sellers were removed from the analysis). The buyers could make an offer and send a message to any subset of sellers (one, two or three), but they often addressed their communique to a single seller.

III. STUDY 3

Experimental comparison of the verifiable and non-verifiable negotiations done in Study 2 did not result in statistically significant results. Although there were some notable differences (e.g., in the buyers' and the sellers' profits), the number of instances was small and the distribution too large to obtain significant results. While there is an indication that, in terms of profits, verifiable negotiations can be positioned in-between auctions and non-verifiable negotiations, we were not able to test this result. Therefore, we conducted a third study that looked at auctions and the two types of negotiations.

A. Auctions and two negotiations experiments' settings

The experiment was conducted in spring 2013 and there were 583 students who participated in it as sellers and 83 students who were buyers. Students came from four universities (located in Canada, the Netherlands, Poland and Taiwan). Because of the differences in the student groups, the requirement that students from one group could participate either in the auction or negotiation but not both, and that students from the same group could not be buyers and sellers, the instances were formed with four and five sellers.

The experiment was conducted online and, as it is often the case, a number of participants were no-shows, dropped out or did not undertake any activity; in this experiment 21% of the students, who played the role of sellers had to be removed from the analysis.

The buyers' average profit values in Experiment 4 were lower than in the earlier experiments because in this experiment the breakeven value for buyers was increased from 16 to 48. The purpose of this change was to place buyers and

sellers on a relatively equal level; both the buyers and the sellers could achieve similar profit values. In negotiations, profit values were greater than the 50-60 units on a [0; 100] scale, which means that the buyers may have been led to make concessions which they would not have otherwise made. This is because of their expectations and the perceived fairness.

B. Results

The results of the third study (Experiment 4) are shown in Table 3. As before, the column "Verifiable" shows the results of the multi-bilateral negotiations, in which the system displayed the best offer made by a seller and the "Non-verifiable" column—where these offers were not displayed. In addition the column "Auctions" refers to the results from the multi-attribute auction experiment.

TABLE 3.
STUDY 3: AUCTIONS AND NEGOTIATIONS WITH VERIFIABLE AND NON-VERIFIABLE OFFERS.

	Experiment 4		
	Auctions	Verifiable	Non-verifiable
No. of instances	38	42	39
No. of sellers	173	147	141
Duration (days, avg.)	**3.35**	**4.63**	**6.20**
Agreement (%)	100	100	100
- Buyer's offer accepted (%)	—	30 (71)	24 (61)
- Seller's offer accepted (%)	38 (100)	12 (29)	15 (39)
Total profit	38.7	39.6	39.7
- Buyers' profit (avg.)	**45.9**	**20.8**	**27.8**
- Sellers' profit (avg.)	**-7.2**	**18.8**	**11.9**
Efficiency			
- Allocative	0.4	0.4	0.4
- Pareto	**3.5**	**81.5**	**38.1**
	Sellers		
No. of offers (avg.)	**9.6**	**5.5**	**4.9**
No. of messages (avg.)	—	**5.9**	**4.7**
Messages' length (words, avg.)	—	**219**	**182**
	Buyers		
No. of offers (avg.)	—	**10.6**	**12.2**
No. of messages (avg.)	—	**12.3**	**15.3**
Messages' length (words, avg.)	—	193	188

The data shows that the auctions took the least time to conclude, followed by the verifiable negotiations, then by non-verifiable negotiation. These differences are significant: for auctions and verifiable negotiation $p < 0.004$; for auctions and non-verifiable negotiation $p < 0.001$; and for verifiable and non-verifiable negotiations $p < 0.001$. This result places verifiable negotiation in-between auction and non-verifiable negotiation in terms of process efficiency.

In addition to the process time we used the number of offers to compare the three mechanisms and the number of messages (with and without offers) to compare the two types of negotiations.

The sellers participating in auctions made significantly more offers (9.6 on average) than the sellers participating in verifiable (5.5, $p < 0.001$) and non-verifiable negotiations (4.9, $p < 0.001$). The difference in the average number of offers made by sellers in the two types of negotiations is not significant.

The average number of messages sent by the sellers was not significantly different in the verifiable and non-verifiable negotiations (5.9 vs. 4.7). However, the average total length of messages (measured in words) was significantly ($p < 0.025$) different in the verifiable and non-verifiable negotiations (219 vs. 182 words).

The buyers participating in the verifiable negotiations made fewer offers (10.6, on average) than the buyers participating in the non-verifiable negotiations (12.2, on average). This difference is not significant ($p = 0.145$).

The average number of messages sent by the buyers is significantly different ($p < 0.05$) in the verifiable and non-verifiable negotiations (12.3 vs. 15.3). However, the average total length of messages (measured in words) is not significantly different ($p = 0.440$) in the verifiable and non-verifiable negotiations (193 vs. 188 words).

The comparison of the three mechanisms based on profits shows a different picture. Both types of negotiations resulted in a very similar total profit; auctions yielded a smaller profit than negotiations. These differences are, however, not significant. This result is interesting because the total profit (social welfare or value allocation) has been frequently used as an indicator of mechanism efficiency [see, e.g., 18, 19, 20]. Our results, however, imply that auctions are no more efficient than the negotiations.

Profit distribution is, however, very different. The auctions were best for the buyers and worst for the sellers, who incurred losses (on average). The non-verifiable negotiations were in-between—they were worse for the buyers and better for the sellers than auctions but better for the buyers and worse for the sellers than the verifiable negotiations (Table 3). These results confirm the negotiation results obtained in Study 2 (Table 2).

The differences between the three mechanisms in terms of the achieved profit are significant. The buyers' profit significance is: for auctions and verifiable negotiations $p < 0.001$; for auctions and non-verifiable negotiations $p < 0.001$; and for verifiable and non-verifiable negotiations $p < 0.01$. The buyers' profit significance is: for auctions and verifiable negotiations $p < 0.001$; for auctions and non-verifiable negotiations $p < 0.001$; and for verifiable and non-verifiable negotiations $p < 0.02$.

Markets are evaluated based on the efficiency of their mechanisms. We used two efficiency measures (see Table 3): allocative efficiency and Pareto efficiency.

Allocative efficiency is the ratio of the average total profit achieved and the maximum total profit that is possible to achieve by the winner. In this work, we used allocative efficiency in a somewhat different way than typically used in economic literature. Rather than using the absolutely maximum total profit, which is the highest possible profit for the theoretical winner (i.e., across all bidders), we used the maximum profit available to the winning bidder. This shows the difference between what the winner achieved and what she could achieve.

The reason for using winner-dependent allocative efficiency is that it can be compared with Pareto efficiency, which is the average number of alternatives dominating the

winning offer. The dominating alternatives need to be selected from the winner's set of feasible alternatives, otherwise the Pareto efficiency is not be comparable across different instances. This is because the value depends on both the winner's and the theoretical winner's feasible sets so that an efficient winning offer can become an inefficient one.

Allocative efficiency is the same for all three mechanisms: they are 40% efficient, which is quite low. Pareto efficiency is significantly different across these mechanisms. On average, there are only 3.5 alternatives, which dominate the winning offer in auctions, but there are 81.5 and 38.1 dominating alternatives in, respectively, verifiable and non-verifiable negotiations. We find these results surprising for two reasons. The first reason is that allocative efficiency is the same but there are significantly more dominating alternatives in negotiations than in auctions. This suggests that the negotiation efficiency can be improved (particularly verifiable negotiation) but there is very limited possibility to improve auctions' efficiency.

The second reason is that auctions produce results, which are close to Pareto frontier (only 3.5 offers dominate the winning offer) but their allocative efficiency is low (0.4). This seems to contradict auction theory which posits that Pareto efficient winning offers are also allocative efficient [18, 21, 22]. This result is related to the exchange problem used in our experiments which did not have a quasi-linear evaluation function, which, in the business case, is profit.

C. Negotiations like auctions

We mentioned above that while the verifiable offer negotiations have the same degree of transparency as auctions, they also differ. The involvement of the buyer (who can present her offers and engage in discussion with individual sellers on any topic they wish) is one of the key differences between auctions and negotiations. This difference, however, need not occur in any given negotiation; because the sellers are shown winning offers the buyers may decide to be inactive. In other words, the buyers may change the negotiation process to auctions without giving any information to sellers a priori.

An analysis of the verifiable negotiations transcripts (Experiment 3) showed that some buyers behaved similarly to the buyers in auctions and did not engage in negotiation activities. This means that this type can be divided into two sub-types: (1) negotiation-like-auctions; and (2) multi-bilateral negotiations.

We investigated the negotiations in which the buyer was inactive for some period of time, analyzing the sellers' actions when they received neither messages nor offers from the buyer. There were 10 instances with 31 sellers who faced the problem of inactive buyer in Experiment 3. These sellers decided to submit new offers, even though they did not receive any answer from the buyer regarding the offers they had sent earlier. In eight instances fourteen sellers sent on average 2.7 offers before their counterpart (buyer) replied. The buyer did not respond until the negotiation deadline in the remaining two instances with nine sellers. In these instances the sellers sent 3.9 offers, on average.

In general, in the inactive-buyer negotiations the sellers submitted on average 2.9 offers, which were not replied to with any counteroffer of the buyer. While reviewing the sellers' negotiation transcripts and their assignment reports we could not find their motivation for doing this.

The sellers could sent offers in order to get the buyer's attention and to induce them to start messaging during which they could convince them to accept the sellers' own offers or they could get involved in the bidding game with other sellers, hoping to eliminate them by sending at this stage of the negotiation process the offers more beneficial for the buyers than the ones submitted by their competitors.

The similarity of the verifiable negotiations and auctions was also noticed by the sellers in their post-negotiation feedback. The participants of auctions described their activities and behaviour in terms of the bidding process (e.g., "I bid", "the other bidders", and "the auction rounds"). In contrast, the participants of non-verifiable negotiations used terms such as: "the counterpart", "I submitted an offer", "tried to achieve a compromise".

The participants of non-verifiable negotiations, however, did not employ negotiation terminology uniformly. In 9 out of 13 feedback messages (69%) they described the negotiation process as a bidding process with bidding rounds and bids submitted by the parties, which is typical for auctions rather than negotiations. This suggests that some participants viewed verifiable negotiations as negotiation-like auctions. Taking into account the fact that some sellers in buyer-inactive instances behaved in a way typical to bidders in auctions (they did not wait for the buyer's response before making a new offer), we may conjecture that verifiable negotiations may be seen as a mechanism in-between auctions and traditional (i.e., non-verifiable) negotiations.

D. Discussion

The purpose of Study 3 was to explore the participants' behavior and the outcomes they achieved in the two types of negotiations and in auctions. The results of this study partially confirm the results of Study 2. Some of the differences may be due to the larger sample in Study 3 and a small??? revision of the assignment which was administered. The revision concerned additional clarification of: (1) the relationship between breakeven values and profits and losses; and the requirement that students achieve profits if they can and avoid losses, for which they are penalized (bonus points are not given); and (3) the allocation of bonus points if student obtained contracts (but not at a loss) as well as if students did not achieve contracts only because making an additional offer would push them into losses. Another difference in this versus the earlier experiments was the change of the breakeven value for the buyers (from 16 to 48), so that the buyers and the sellers could achieve similar profit values.

In Study 3 we were able to determine strong relationship between a number of variables describing auctions and negotiations (Table 3). The purpose of the study was experimental comparison of the three mechanisms and we obtained interesting yet surprising results. We introduced a new negotiation

mechanism which shares one rule with auctions, namely winning offer verifiable disclosure. This rule is critical for auctions (particularly multi-attribute) because it provides guidance for the bidders.

The additional rule provided sellers in both the auctions and the verifiable negotiations with information about the winning offer (i.e., best offer on the table). In these negotiations the buyers could communicate with the sellers, while in the auctions the sellers were given information about admissible bidding sets [23]. These sets comprise alternatives, which are better for the buyer than the winning bid.

The results show that the verifiable offers improved the sellers' position, however at the process costs measured by the number and length of offers. Because the buyers were worse off than in the auctions and the non-verifiable negotiations, they were not likely to introduce the verifiable offer negotiation mechanism.

IV. Conclusions and Future Work

Auctions are *economic processes* in the sense that nothing except for the attribute values can be submitted. Auction outcomes are thus defined solely by the attributes defined by the bid-takers. In negotiation literature this type of outcomes is called "substantive"; their values are discussed over the course of the process and they constitute the agreement [24].

In negotiation literature, substantive outcomes have been contrasted with relational outcomes; the roots of this distinction are attributed to an effort to contrast the economic perspective with the psychological perspective [25, 26]. The argument which we posit here is that negotiations among market participants and businesses are *socio-economic processes* and that neither the "social" nor the "economic" aspects can be ignored.

The social exchange theory is concerned with the formulation and evolution of the relationship between parties engaged in giving and getting "something", and with the rules which govern exchanges between the parties [27, 28]. There are two main types of rules [29]: (1) negotiated rules; and (2) reciprocity rules. The negotiated rules are explicit and simple, they deal with bargaining in which reciprocity is not required. The reciprocity rules are implicit and govern different forms of relationships, which emerge during interactions among people (e.g., trust, empathy, and reputation).

Despite of its recognition of the reciprocity rules, the social exchange theory reduces negotiated exchanges to haggling or double auctions, noting that in negotiations "reciprocity is a trivial byproduct of a bilateral trade, and the same actions that reduce the risk of loss also increase gain." [30]. However, even this narrow perspective on negotiation identifies reciprocity as an important device used by negotiators. An action by one party calls for some kind of a response by the counterpart, it creates an obligation. If it is clear that the party makes an effort, provides explanation, proposes a significant concession, and is genuinely interested in getting the contract, then it is only natural for the counterpart to reciprocate. This is one reason why buyers accept less (lower profit) in the multi-bilateral negotiations, than in auctions.

The participants in our negotiation experiments play roles of buyers and sellers; they perform and interact with others. They may also discuss other issues (e.g., their interests, weather, and universities). The negotiations are anonymous at the outset, but the participants can exchange any information they wish to exchange. The participants' discussions may have a subjective value for them.

A person may not know her counterpart but during the ten-day long interaction may develop some affinity with him, which can lead her to make a bigger concession than she would have made if she felt animosity. This particular motivation for concession-making can be related to the experimental settings, however, in real-life situations we also observe parties trading off some substantive values in an effort to achieve higher relational values. In some job markets, for example, employers engage in multi-bilateral negotiations with several potential candidates in order to determine their trustworthiness, fit to the position and the team, as well as professional skills. If they need to determine skills only, then auction often is the preferred mechanism [31]. This implies that reciprocity need not be a "trivial byproduct" but a set of complex rules which are invoked when the negotiators realize the potential of achieving important relational outcomes.

Relational outcomes are inherently social and they can be achieved in negotiations. However, they cannot be achieved in auctions in which bid-makers do not interact with one another. This shortcoming of auctions has been recognized and led to augmentation of auction protocols, e.g., with post-auction negotiation in buyer-determined auctions [32].

While non-augmented auctions cannot produce relational outcomes, they can produce game-like outcomes, for example, excitement [33]. Auctions produce winners and losers, the outcome is a win or a loss, while negotiations result in agreement or disagreement achieved through negotiation.

Our results confirm the theory that auctions produce better substantive outcomes for the bid-takers who decide on what exchange mechanism to use. The assumption is, however, that the bid-takers are not interested in any other outcomes, relational in particular. If so, the answer to the first question formulated in Section1, is negative: For the buyers, it is not worth spending time and money for negotiation, because they achieve better results from auctions. The results also point to the necessity to study communication between negotiators. Messages affect offers; if they are ignored then the changes in offers (concessions) cannot be explained.

Notwithstanding the results obtained from Study 3 about the verifiable and non-verifiable negotiations, which confirm the results from Study 2, we consider these results as tentative and more work is required to validate them. The reason is due to the participants' different behaviors in each type of the negotiations. We mentioned above that in the verifiable negotiations there were inactive buyers during the first few days of the process; there were also a few inactive buyers during the entire process. While these negotiations concluded with an

agreement, that is, at some point the buyer accepted an offer, they were structurally different from the negotiations in which the buyer made offers and sent messages.

The data obtained from the verifiable and non-verifiable negotiation experiments (Study 2) was inconclusive; the differences in the buyers' and the sellers' profit values were not significant. However, this difference was observable and therefore it suggested that transparency could be better for buyers but not necessarily for sellers (Table 2). The results of Study 3 show that the differences are significant but not in the expected direction.

Contrary to our expectations, the verifiable negotiations did not produce better results for the buyers and worse for the seller that the non-verifiable negotiations (Table 3).

Because transparency has been found to have positive effect on trust and other relational outcomes [34], in some situations verifiable-offer negotiations may be preferred over both auctions and non-verifiable negotiations. It is possible to further augment this type of negotiation by providing information about admissible bidding (offer) sets. In multi-attribute auctions bidders can only submit a bid that is an element of one of these sets. In negotiations they could submit other bids (i.e., worse for the buyers) but this information would give them a better understanding regarding objective positive concessions.

ACKNOWLEDGMENT

We thank Norma Paradis, Dmitri Gimon, and Shikui Wu for their contribution to the system design and experiment organization.

REFERENCES

[1] C. Beam, A. Segev, M. Bichler, and R. Krishnan, "On Negotiations and Deal Making in Electronic Markets," *Information Systems Frontiers,* vol. 1, pp. 241-258, 1999.

[2] M. Kumar and S. I. Feldman, "Business Negotiation on the Internet," IBM Institute for Advanced Commerce, Yorktown Heights, NYMarch 11, 1998 1998.

[3] M. Ströbel, "On Auctions as the Negotiation Paradigm of Electronic Markets," *Electronic Markets,* vol. 10, pp. 39-44, 2000.

[4] G. E. Kersten, S. J. Noronha, and J. Teich, "Are All E-Commerce Negotiations Auctions?," presented at the 4th International Conference on the Design of Cooperative Systems, Sophia Antipolis, 2000.

[5] M. Bichler, G. E. Kersten, and S. Strecker, "Towards a Structured Design of Electronic Negotiation Media," *Group Decision and Negotiation,* vol. 12, pp. 311-335, 2003.

[6] T. F. Gattiker, X. Huang, and J. L. Schwarz, "Negotiation, Email, and Internet Reverse Auctions: How Sourcing Mechanisms Deployed by Buyers Affect Suppliers' Trust," *Journal of Operations Management,* vol. 25, pp. 184-202, 2007.

[7] Z. Neeman and N. Vulkan, "Markets versus negotiations: The predominance of centralized markets," *The BE Journal of Theoretical Economics,* vol. 10, 2010.

[8] P. Bajari, R. McMillan, and S. Tadelis, "Auctions versus Negotiations in Procurement: An Empirical Analysis," *Journal of Law, Economics, and Organization,* vol. 25, pp. 372-399, 2009.

[9] A. Bonaccorsi, T. Lyon, F. Pammolli, and G. Turchetti, "Auctions vs. Bargaining: An Empirical Analysis of Medical Device Procurement," Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa2000.

[10] P. Bajari, "Econometrics of Sealed-Bid Auctions," in Business and Economic Statistics Section of the American Statistical Association, 1998, pp. 41-49.

[11] C. J. Thomas, "An Alternating-Offers Model of Multilateral Negotiations," 2012.

[12] C. J. Thomas and B. J. Wilson, "A Comparison of Auctions and Multilateral Negotiations," *The RAND Journal of Economics,* vol. 33, pp. 140-155, 2002.

[13] C. J. Thomas and B. J. Wilson, "Verifiable Offers and the Relationship Between Auctions and Multilateral Negotiations," *Economic Journal,* vol. 115, pp. 1016-1031, 2005.

[14] J. Bulow and P. Klemperer, "Auctions versus Negotiations," *American Economic Review,* vol. 86, pp. 80-194, 1996.

[15] A. Manelli and D. Vincent, "Optimal Procurement Mechanisms," *Econometrica,* vol. 63, pp. 591-620, 1995.

[16] S. Strecker, G.E. Kersten, J. Kim, and K.P. Law, "Electronic Negotiation Systems: The Invite Prototype," in *Proceedings of the Collaborative Business*, Potsdam, Germany, 2006, pp. 315-331.

[17] G. E. Kersten, R. Vahidov, and D. Gimon, "Concession Patterns in Multi-attribute Auctions and Multi-bilateral Negotiations: Theory and Experiments," *Electronic Commerce Research and Applications,* p. (in print), 2013.

[18] M. Bichler and J. Kalagnanam, "Configurable Offers and Winner Determination in Multi-attribute Auctions," *European Journal of Operational Research,* vol. 160, pp. 380-394, 2005.

[19] S. Strecker, "Information Revelation in Multiattribute English Auctions: A Laboratory Study," *Decision Support Systems,* vol. 49, pp. 272-280, 2010.

[20] P. Milgrom, "An Economist's Vision of the B-to-B Marketplace," 2000.

[21] Y. K. Che, "Design Competition through Multidimensional Auctions," *The RAND Journal of Economics,* vol. 24, pp. 668-680, 1993.

[22] D. C. Parkes and J. Kalagnanam, "Models for Iterative Multiattribute Procurement Auctions," *Management Science,* vol. 51, pp. 435–451, 2005.

[23] G. E. Kersten, P. Pontrandolfo, and S. Wu, "A Multiattribute Auction Procedure and Its Implementation," in *HICSS 45*, Hawaii, 2012, pp. 600-609.

[24] L. Thompson, "Negotiation Behavior and Outcomes: Empirical Evidence and Theoretical Issues," *Psychological Bulletin,* vol. 108, pp. 515-532, 1990.

[25] M. H. Bazerman, J. R. Curhan, and D. A. Moore, "The Death and Rebirth of the Social Psychology of Negotiation," *Blackwell Handbook of Social Psychology,* pp. 196-228, 2001.

[26] M. J. Gelfand, V. S. Major, J. L. Raver, L. H. Nishii, and K. O'Brien, "Negotiating Relationally: The Dynamics of the Relational Self in Negotiations," *Academy of Management Review,* 31, pp. 427-451, 2006.

[27] W. P. Bottom, J. Holloway, G. J. Miller, A. Mislin, and A. Whitford, "Building a pathway to cooperation: Negotiation and social exchange between principal and agent," *Administrative Science Quarterly,* vol. 51, pp. 29-58, 2006.

[28] R. Cropanzano and M. S. Mitchell, "Social Exchange Theory: An Interdisciplinary Review," *Journal of Management,* vol. 31, pp. 874-900, 2005.

[29] L. D. Molm, "The Structure of Reciprocity," *Social Psychology Quarterly,* vol. 73, pp. 119-131, 2010.

[30] L. D. Molm, "Theoretical Comparisons of Forms of Exchange," *Sociological Theory,* vol. 21, pp. 1-17, 2003.

[31] A. Schram, J. Brandts, and K. Gërxhani, "Information, Bilateral Negotiations, and Worker Recruitment," *European Economic Review,* vol. 54, pp. 1035-1058, 2010.

[32] R. Engelbrecht-Wiggans, E. Haruvy, and E. Katok, "A Comparison of Buyer-determined and Price-based Multi-attribute Mechanisms," *Marketing Science,* vol. 26, pp. 629–641, 2007.

[33] M. T. P. Adam, J. Kramer, and C. Weinhardt, "Excitement Up! Price Down! Measuring Emotions in Duch Auctions," *International Journal of Electronic Commerce,* vol. 17, pp. 7-39, 2013.

[34] J. Hultman and B. Axelsson, "Towards a typology of transparency for marketing management research," *Industrial Marketing Management,* vol. 36, pp. 627-635, 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1103–1110

Verification of ArchiMate process specifications based on deductive temporal reasoning

Radosław Klimek* and Piotr Szwed*
*AGH University of Science and Technology
E-mail: {rklimek,pszwed}@agh.edu.pl

***Abstract*—Formal verification of business models has become recently an intensively researched area. Application of formal methods in this field necessities in overcoming several problems. Firstly, business analyst and designers rarely have enough skills and motivation to manually build abstract and formal specifications, hence, it arises the need to provide tools for an automated translation of business models into a suitable form ready for formal verification. Moreover, notations and languages used to describe enterprises usually have no clear semantics. Finally, the verification itself must be supported by an efficient tool. In this paper we investigate an application of formal and deduction-based techniques to automated verification of behavioral description embedded within ArchiMate models. We describe a set of rules that governs translation of processes specified in ArchiMate language into Linear Temporal Logic (LTL) formulas. The translation step is achieved with the developed software, as a plugin into a popular the Archi modeler. Formal verification of a business process properties is achieved with another tool, the LTL prover based on the semantic tableaux technique. Application of the method is discussed on a small, yet illustrative, example of a taxi service.**

I. Introduction

Formal verification of business models has become recently an intensively researched topic. The growing interest in this area stems to some extend form a historical fact. Issued in 2000 revisions to ISO 9001 and 9004 norms recommended process-oriented approach to quality management. Since then, many organizations have started to identify and describe their processes to fulfill certification requirements, but also they have realized that coherently and unambiguously specified business processes can bring such benefits, as quality of products and services, smaller ratio of operational errors, reduced cost and greater competitiveness.

In this paper we investigate an application of formal deduction-based techniques to automated verification of behavioral description embedded within ArchiMate models. ArchiMate is an lightweight and scalable language supporting analysis and modeling of business and enterprise domains [23]. It has been developed to provide a uniform representation for architecture descriptions. The popularity of ArchiMate language within the enterprise architecture modeling community grows.

Formal methods are understood as a set of principles for precise formulation of important artifacts formed when developing and refining software models. They enable revealing, questioning and removing ambiguities and flaws in specifications. Moreover, the formality of used languages leads to the rigorous analysis and verification. There are two established approaches to formal reasoning and system verification [5], i.e. model checking and deductive reasoning. Model checking is based on the state exploration and is an operational rather than analytic approach [4]. Deductive inference enables analysis of infinite computation sequences.

Temporal logic TL brings in logical symbols for reasoning about varying logical valuations of formulas throughout the flow of time. Two basic unary operators are $\Diamond$ for "sometime (or eventually) in the future" and $\Box$ for "always in the future". Temporal logic is a well-established formalism for specification and verification of reactive and concurrent systems. It allows to describe both temporal relations between reached states or events occurring within a system and to specify expected properties.

Liveness and safety are standard elements of a taxonomy of system properties. *Liveness* means that the computational process achieves its goals, i.e. something good eventually happens. *Safety* means that the computational process avoids undesirable situations, i.e. something bad never happens.

In recent years, a number of temporal logics has been proposed. Temporal logic exists in many varieties, however, these considerations are limited to the *linear-time temporal logic* (LTL). Linear temporal logic refers to infinite sequences of computations considered as linear structures and our attention is focused on the *propositional linear time logic* PLTL. These sequences are formally represented as Kripke structures, which define semantics of TL, i.e. a syntactically correct, or a well-formed, formula can be satisfied by an infinite sequence of truth evaluations over a set of *atomic propositions* AP. The basic issues related to temporal logics and their syntax and semantics are discussed in many works, e.g. [9].

The properties of time structure are fundamental to a logic. Of particular significance is the *minimal temporal logic*, e.g. [25], also known as temporal logic of the class K. The minimal temporal logic is an extension to a classical calculus defining the axiom $\Box(P \Rightarrow Q) \Rightarrow (\Box P \Rightarrow \Box Q)$ and the inference rule $\vdash P \implies \vdash \Box P$. The essence of the logic is the fact that there are no specific assumptions pertaining to the time structure order. The following formulas may be considered as typical examples of this logic: $action \Rightarrow \Diamond reaction$, $\Box(send \Rightarrow \Diamond receive)$, $\Diamond alive$, $\Box\neg(badevent)$, etc. The considerations of this work are limited to this logic since it allows to define many system properties (safety, liveness) and it is also easier

978-83-60810-53-8/$25.00

to build a deduction engine, or use the existing verified provers.

Application of deductive approach to validation of business processes faces the problem of automatic obtaining logical specifications from business models. The need to build them manually can be recognized as a major obstacle to untrained users, due to the fact that the process of specifying of a large collection of formulas is difficult and monotonous, c.f. also the requirements engineering process [13].

For temporal logic, that is a suitable language for expressing behavior and reasoning about it, such specifications are constituted by set of temporal logic formulas $\{F_1, ..., F_n\}$ When the number of formulas is large, what is not an extraordinary situation, then in practice it is not possible to build a logical specification manually. It follows that this process usually requires (very) skilled human intervention. Thus, in order to move the deductive-based formal verification form a pen-and-paper approach to engineers' needs an automation of the generation process seems particularly important.

The motivation for the work is the lack of tools for the deduction-based formal verification of business models. Another motivation is the lack of tools for the automatic generation of logical specifications from Archimate models.

The contribution of the work are the following: rules for automatic generation of logical specifications considered as sets of temporal logic formulas are defined and a complete deduction-based system, which enables an automated and formal verification of Archimate business models is proposed. Reasoning process is performed using the semantic tableaux method for temporal logic. An example of the approach is provided.

The paper is organized as follows: the next Section II discusses approaches to verification of business models, it is followed by Section III briefly describing ArchiMate language. Then, in Section IV an example of a process specification using ArchiMate is provided. Section V defines rules governing translation of ArchiMate models to LTL formulas. In Section VI an architecture of the verification system is described and an example of a checked property is given. Finally Section VII gives concluding remarks.

II. Related work

Recent work by Morimoto [15] surveys formal verification tools for business processes. It discusses in the context of business process management application of such formalisms as: automata, model checking, process algebras and Petri nets. The described approaches can be considered as variations of either model checking or simulation. In particular, model checking seems to be the most often used. There are several reports on application of model checking approach, e.g. to perform verification of of e-business processes, work by Anderson et al. [2], or BPMN models extended with resource constraints, c.f. work by Watahiki et al. [28]. In work by Deutsch at al. [8] verification of data-centric business processes is studied. The correctness problem was expressed in the LTL-FO, an extension to the Linear Temporal Logic, in which propositions were replaced by First Order statements about data objects. A salient consequence of modeling operations on data are infinite domains. Hence, the problem of correctness verification can be undecidable. Application of CTL to verification of BPEL processes was reported in work by Mongiello and Castelluccia [14]. Three types of correctness properties were analyzed: invariants, properties of final states and temporal relations between activities. The first two can be classified as *safeness*, the last as the *liveness* property. Similarly, in work by Fu et al. [11] CTL was applied to the verification of e-services and workflows with both bounded and unbounded number of process instances. Application of deduction based approach is rare in the area of business models verification. The work by Shankar [21] contains a comprehensive study for the area of verification using automated deduction and deduction-based techniques. Up to our best knowledge, no attempts has been made to define formally semantics and perform verification for behavioral elements of ArchiMate. Some suggestions and research direction can be found in an early document [7]. On the other hand, in a few publications [10], [3] ontologies were applied to define semantics for subsets of ArchiMate elements and relations. However, all of the research themes mentioned above are different from the approach presented in the work.

III. Archimate

ArchiMate [23], [26] is a contemporary, open and independent language for description of enterprise architectures. It comprises three main modeling layers: business, application and technology. The *business* layer includes business processes and objects, functions, events, roles and services. The *application* layer contains components, interfaces, application services and data objects. The *technology* layer gathers such elements as artifacts, nodes, software, devices, communication channels and networks. ArchiMate allows to present an architecture in the form of views which, depending on the needs, can include only items in one layer or can show vertical relations between layers, e.g.: a relationship between a business process and a function of the component software. ArchiMate was built in opposition to UML [18], which can be seen as a collection of unrelated diagrams, and Business Process Modeling Notation BPMN [17] which covers mainly behavioral aspect of enterprise architecture. The definition of a language has been accompanied by an assumption, that in order to build an expressive business model, it is necessary to use the relationships between completely different areas, starting from business motivation to business processes, services and infrastructure. ArchiMate goes beyond UML [16]: it defines a metamodel on the basis of which a user can create and illustrate the relationships between elements of different layers.

Archimate provides a small set of constructs that can be used to model behavior. It includes *Business Processes*, *Functions*, *Interactions*, *Events* and various connectors (*Junctions*), which can be attributed with a logical operator specifying, how inputs should be combined or output produced. According to language specification casual or temporal relationships be-

tween behavioral elements are expressed with use of *triggering* relation. On the other hand, Archimate models frequently use *composition* and *aggregation* relations, e.g. to show that a process is built from smaller behavioral elements (subprocesses or functions). In should be also noted that *Business Activity* present in Archimate 1.0 specification was removed in version 2.0. Instead, an atomic process should be used.

Although the set of behavioral elements seems to be very limited when compared with BPMN [17], after adopting a certain modeling convention its expressiveness can be similar [22]. An advantage of the language is that in allows to comprise in a single model a broad context of business processes including roles, services, processed business objects and elements of lower layers responsible for implementation and deployment.

Another process modeling notation that can be almost directly mapped on Archimate constructs is *Event-driven Process Chain* (EPC) [20], [19]. Indeed, all behavioral elements of both languages are exactly the same: events, functions (or processes in Archimate) and various joins and splits (XOR, OR and AND). In spite of almost 20 years presence of EPC tools on a market and thousands of deployments in modeling business organizations, there is no consensus of semantics of EPCs. Analyzes of several semantics variations have revealed certain erroneous patterns, e.g the famous vicious circle [27] resulting in a deadlock caused by improper use of synchronization joins. Due to the correspondence between EPC and Archimate constructs, discussions and discovered problems related to EPCs semantics apply also to Archimate models.

IV. Example of a business model

In this section we give an example of ArchiMate model for a small business process defining Taxi service allowing a client to order a taxi that will carry him or her to the desired destination.

The example shows typical constructs that can be used to model business behavior in ArchiMate. We assume that each high-level process starts and ends with an event. A triggering event originate from the environment, e.g. *Client trip order* in Fig. 1 or *Order cancellation request* in Fig. 4. A complex process can be decomposed into several subprocesses, which are usually presented in separate views. Such processes are linked by intermediate events, e.g. *Trip accepted* appearing on the output of a subprocess in Fig. 1 and on the input of Fig. 3b. Finally, a process stops at *end events*, which may trigger external processes or terminate a thread of execution, as events *Stop 1 – Stop 7*.

For clarity, we omitted in diagrams additional information, which is usually present in ArchiMate models of business layer, e.g. assigned roles and accessed business objects, functions, services, actors and locations. ArchiMate specifications can be much richer, what justifies common practice of decomposing behavior descriptions into smaller patterns presented in multiple views. In this way specifications can cover various important aspects of an enterprise architecture, but control flows become harder to be verified manually, as they require switching between multiple views. In fact, specifications presented in Fig. 1–4 are excerpts restricted to behavioral elements presented in eight views.

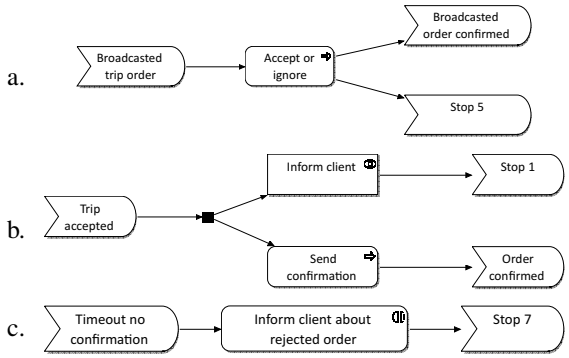


Fig. 3. Extensions to the main operator processes: a. Broadcasting trip order; b. Reaction on accepted order; c. Handling timeout (an order cannot be realized) Broadcast

The main flow of activities within dispatcher's office is shown in Fig. 1.

When a client requests a taxi (*Client trip order*), then the order is handled by a dispatcher. Firstly, all taxicab stands located closest to the present customer's position are searched. If contact with the driver (*Contact taxi driver*), and details of the order, leads to its rejection (*Trip rejected*), then the next taxicab form stands is selected. If a free taxi is found and the trip request is accepted (*Trip Accepted*), then the customer and the driver are provided with information about each other as well as trip's details, and the trip order is stored in the database as approved. If *Look for free taxi cabs* process takes too long time (*Timeout*), then dispatcher broadcasts the trip order to all drivers (in the city area). Such order may also be accepted on the general principles.

Main process of the driver is shown in Fig. 2.

After an order confirmation, the agreed driver goes to the the trip start location (*Reach location*) to pick up the passenger (*Pick up passenger*). Along the way to the pick-up location difficulties may arrise, e.g. heavy traffic or passenger absence. After the completion of the ride (*Trip finished*), its status is changed from approved to finished.

Extensions of the dispatcher processes are shown in Fig. 3 and discussed below.

(a) Waiting for the first declaration which originates from the broadcast.
(b) A kind of a virtual handshake between a taxicab driver (*Send confirmation*) and the passenger (*Inform client*) with an intermediary of dispatcher.
(c) If none of the methods result, the customer is notified and the order is rejected (*Timeout no confirmation*).

Further extensions of the dispatcher processes are shown in Fig. 4 and discussed below.

(a) If there are obstacles when reaching the client, then he/she should be informed about this fact.
(b) When order is canceled, the driver must be notified.

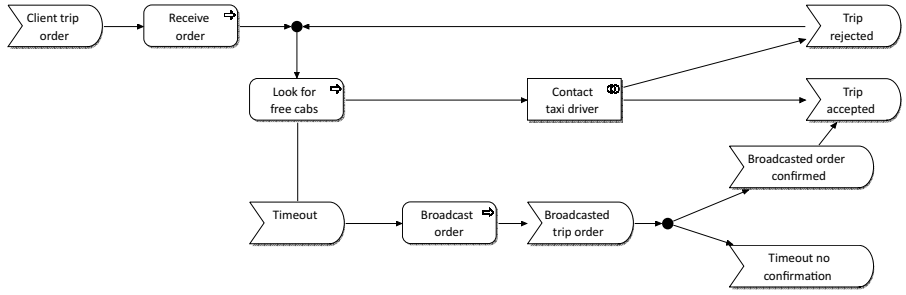


Fig. 1. Main dispatcher process. After receiving a client order, operator looks for registered free cabs. If not found, broadcasts information about the order

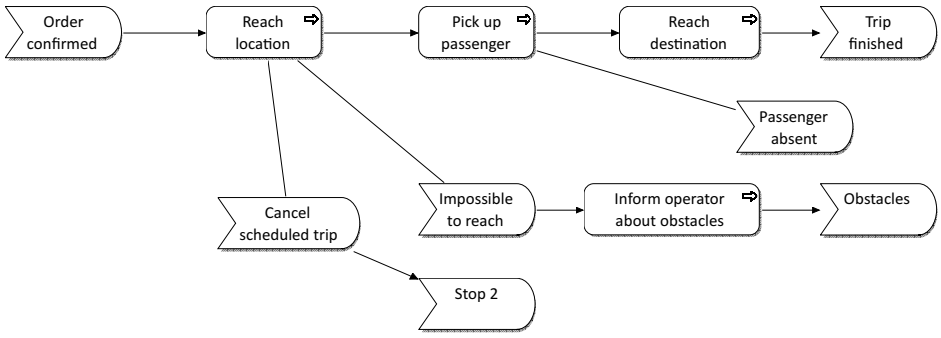


Fig. 2. Main driver process.

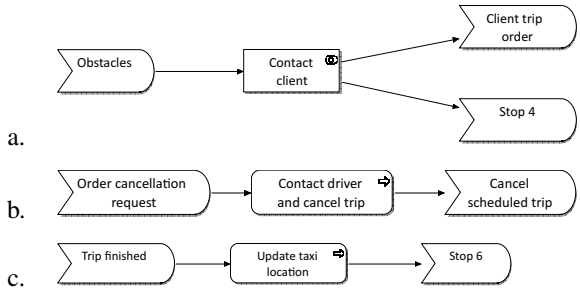


Fig. 4. Further extensions to the main operator processes: a. Contacting client after a taxi driver reports obstacles, e.g. delays ; b. Handling cancellation; c. updating information on taxi location after a finished trip

(c) When the trip is successfully finished, then the new taxicab location is introduced.

V. Modeling Archimate behavioral constructs

This section gives formally defined rules for translation of behavioral elements within an ArchiMate specification into LTL formulas. The internal structure of an Archimate model constitutes a graph of nodes linked by directed edges. Both nodes and edges are attributed with information indicating a type of element or relation. Generating LTL formulas describing behavioral aspects of Archimate model we focus on components of the *Business layer*: processes (or functions), events and various junctions. We apply a linear procedure, which visits nodes, analyzes their successors and generates LTL formulas describing control flow.

It should be noted that Archimate behavioral constructs have no precisely defined semantics. In fact, translation from Archimate specification to LTL assigns a semantics, which, although arbitrarily selected, follows a certain intuition, e.g. how to interpret an activity or an event.

Definition 1 (Archimate model)**.** Archimate model AM is a tuple $\langle V, E, C, R, v, e\rangle$, where

- V is a set of vertices,
- $E \subset V \times V$ is a set of edges,
- C is a set of Archimate element types,
- R is a set of relations,
- $vt\colon V \to C$ is a function that assigns element types to graph vertices
- $et\colon E \to R$ assigns relation types to edges.

As the considerations in the work focus on business layer elements used to specify behavior, it is assumed that $C = \{Process, Function, Interaction, Event, Junction, And\text{-}Junction, OrJunction, Other\}$ and $R = \{triggering, association, composition, other\}$.

A. Modeling atomic activities

By atomic process (function, interaction) we mean a process that is not linked with other elements by a *composition* relation. It represents a basic unit of behavior, which corresponds to the activity concept of other languages, e.g. UML.

A process can be executed if its environment is in a state enabling its activation. After a process terminates, it causes state changes in the surrounding world [12]. While defining

LTL formulas describing processes and other elements, we follow directions of relations and specify only transitions between internal states of elements and caused states. In turn, the reached caused states enable activation of other elements. Hence, after processing all relevant elements, a complete network of states of the whole system specified in LTL is obtained.

To model execution of an atomic process two states (and corresponding propositions in LTL): *start* and *end* are used. A process is considered *imperfect*, even if has a name in imperative mood suggesting achievement, e.g. *register invoice*, *scan document*, *send message*. Once invoked (the *start* state becomes active), the process can successfully complete reaching the *end* state or be interrupted by an event starting an alternative flow of control. Such approach to modeling business processes can be explicitly supported by language constructs. In particular, the BPMN notation allows to attach various types of interrupting events to activities, e.g. timer, error or cancellation. In ArchiMate an association relation between processes and events can be used to distinguish events triggered upon process completion and interrupting a normal flow.

Fig. 5 illustrates this approach. The *end* state and *Interrupting event* are successor of the process *start* state. On the other hand, states of surrounding elements that can be reached by the normal triggering relation are successors of the *end* state.

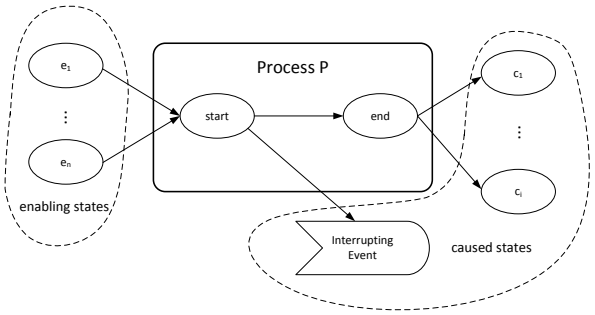


Fig. 5. Two states: *start* and *end* used to model a process

Atomic processes, functions and interactions (ArchiMate equivalents to activities) are imperfect and require two states (and propositions) to model their behavior. In turn events and junctions are perfect and their activation can be modeled by singular states (truth values of propositions). To describe all behavioral elements in an uniform manner we define two functions $start$ and end that map vertices from Archimate model V to a set of propositions $Props$. It is assumed, that if a certain vertex v represents a process, a function or an interaction, i.e. $v(v) \in \{Process, Funtion, Interaction\}$, then $start(v) \neq end(v)$. For other elements: events and junctions $start(v) = end(v)$ holds. We extend these functions to sets of vertices, i.e. $start(X) = \bigcup_{v \in X} start(v)$ and $end(X) = \bigcup_{v \in X} end(v)$.

By $T(v) = \{v' : (v, v') \in E \wedge et(v, v') = triggering\}$ we will denote a set of behaviors that are triggered by v. $A(v) = \{v' : vt(v') = (Event) \wedge (v, v') \in E \wedge et(v, v') = association\}$ is a set of events linked with v by association relation. $C(v) = \{v' : (v, v') \in E \wedge et(v, v') = composition\}$ is a set of children of v as defined by composition relation.

Let $\mathcal{F}(Props)$ be a set of LTL formulas obtained from a set of propositions $Props$ by applying classical or temporal operators and parentheses. For the brevity of notation we will further omit $Props$ and write simply $\mathcal{F}$

We define two auxiliary functions $\delta_{ij}(p)$ mapping formulas $\mathcal{F} \times \mathbb{N} \times \mathbb{N} \rightarrow \mathcal{F}$ (1) and $oneof(P)$ converting a set of propositions P into a formula in disjunctive normal form (2).

$$\delta_{ij}(p) = \begin{cases} p, & \text{if } i \neq j \\ \neg(p), & \text{if } i = j \end{cases} \tag{1}$$

$$oneof(P) = \bigvee_{i=1}^{|P|} \bigwedge_{j=1}^{|P|} \delta_{ij}(p_j) \tag{2}$$

B. Atomic process, function or interaction

LTL formulas defining temporal relations for atomic activities (processes, functions and interactions) are generated according to Rule 1. Rules for other ArchiMate elements have similar form. Each rule contains precondition part separated from its postcondition by a horizontal line. Generated formulas are placed in brackets $[\![\ ]\!]$.

Rule 1. Atomic process, function or interaction

$$\frac{\begin{array}{c} v \in V, \\ vt(v) \in \{Process, Function, Interaction\}, \\ C(v) = \emptyset \end{array}}{\begin{array}{c} [\![\Box(start(v) \Rightarrow \Diamond oneof(start(A(v)) \cup \{end(v)\})]\!] \in \mathcal{F} \\ p \in A(v) \cup \{end(v)\} \rightarrow [\![\Box\neg(start(v) \wedge p)]\!] \in \mathcal{F} \\ T(v) \neq \emptyset \rightarrow [\![\Box(end(v) \Rightarrow \Diamond oneof(start(T(v))))]\!] \in \mathcal{F} \\ p \in T(v) \rightarrow [\![\Box\neg(end(v) \wedge p)]\!] \in \mathcal{F} \end{array}}$$

LTL formulas describing the behavior for the sequence of two active elements *Look for free cabs* and *Contact taxi driver* in Fig. 1 are presented in Fig. 6 (original transcription is preserved). They were generated according to Rule 1.

```
% BusinessProcess (Look for free cabs)
[] (Look_for_free_cabs_start =>
   <> ((Look_for_free_cabs_end & ~Timeout)
   |  (~Look_for_free_cabs_end & Timeout))),
[]~(Look_for_free_cabs_start & Look_for_free_cabs_end),
[]~(Look_for_free_cabs_start & Timeout),
[] (Look_for_free_cabs_end => <>Contact_taxi_driver_start),
[]~(Look_for_free_cabs_end & Contact_taxi_driver_start),
% BusinessInteraction (Contact taxi driver)
[] (Contact_taxi_driver_start => <>Contact_taxi_driver_end),
[]~(Contact_taxi_driver_start & Contact_taxi_driver_end ),
[] (Contact_taxi_driver_end =>
   <> ((Trip_accepted & ~Trip_rejected)
   |  (~Trip_accepted & Trip_rejected ))),
[]~(Contact_taxi_driver_end & Trip_accepted),
[]~(Contact_taxi_driver_end & Trip_rejected),
```

Fig. 6. An excerpt of generated formulas for the main dispatcher process

C. Event

According to Archimate specification [23] business event is something that happens and influences behavioral elements

(processes, functions and interactions). Events has no duration, thus they can be modeled as single boolean variables. Functions $start(v)$ and $end(v)$ map an event v to the same proposition, which change value to true if the event occurs.

An event can be linked by triggering relations with multiple recipients (or *sinks* in the Event Driven Architecture). Events are somehow similar to AndJunctions. An occurrence or the both activates all elements linked by a *triggering* relation. However, we assume that, unlike AndJunctions, activation of elements triggered by an event is not synchronized (c.f. Rule 2).

Rule 2. Event

$$\frac{v \in V \wedge vt(v) = Event}{\begin{array}{c} p \in T(v) \rightarrow [\![\Box(end(v) \Rightarrow \Diamond start(p)]\!] \in \mathcal{F} \\ p \in T(v) \rightarrow [\![\Box(start(p) \Rightarrow \Diamond \neg end(v))]\!] \in \mathcal{F} \end{array}}$$

D. Junctions

Archimate language defines three types of connectors:

- *Junction* that can be considered a typical XOR connector, i.e. it activates exactly one output.
- *OrJunction* being a typical OR connector activating at least one output
- *AndJunction* that can be used in two modes: when used to merge flows on input it requires their synchronization. In the second mode it starts a parallel execution of output flows.

Archimate junctions has counterparts in EPC, BPMN (exclusive, inclusive and parallel gateways) and XPDL transition restrictions [24].

Similarly to events, junctions are modeled by single state variables. If a junction is activated, in a subsequent step elements linked by triggering relations should be activated according to assumed semantics. As there are three types of junctions, we define three rules (Rule 3–5) for translating them to LTL formulas.

Rule 3. Junction

$$\frac{v \in V \wedge vt(v) = Junction}{\begin{array}{c} T(v) \neq \emptyset \rightarrow [\![\Box(end(v) \Rightarrow \Diamond oneof(start(T(v)))]\!] \in \mathcal{F} \\ p \in T(v) \rightarrow [\![\Box\neg(end(v) \wedge p)]\!] \in \mathcal{F} \end{array}}$$

Rule 4. OrJunction

$$\frac{v \in V \wedge vt(v) = OrJunction}{\begin{array}{c} T(v) \neq \emptyset \rightarrow [\![\Box(end(v) \Rightarrow (\bigvee_{z \in T(v)} start(z)))]\!] \in \mathcal{F} \\ T(v) \neq \emptyset \rightarrow [\![\Box\neg(end(v) \wedge (\bigwedge_{z \in T(v)} start(z)))]\!] \in \mathcal{F} \end{array}}$$

Define: $T^{-1}(v) = \{v' : (v', v) \in E \wedge et(v', v) = triggering\}$ - set of input

Rule 5. AndJunction

$$\frac{v \in V \wedge vt(v) = AndJunction}{\begin{array}{c} T^{-1}(v) \neq \emptyset \rightarrow \\ [\![\Box((\bigwedge_{z \in T^{-1}(v)} end(z)) \Rightarrow start(v))]\!] \in \mathcal{F} \\ T^{-1}(v) \neq \emptyset \rightarrow \\ [\![\Box\neg((\bigwedge_{z \in T^{-1}(v)} end(z)) \wedge start(v))]\!] \in \mathcal{F} \\ T(v) \neq \emptyset \rightarrow [\![\Box(end(v) \Rightarrow (\bigwedge_{z \in T(v)} start(z)))]\!] \in \mathcal{F} \\ T(v) \neq \emptyset \rightarrow [\![\Box\neg(end(v) \wedge (\bigwedge_{z \in T(v)} start(z)))]\!] \in \mathcal{F} \end{array}}$$

E. Discussion

In this section formal rules governing translation of basic ArchiMate behavioral elements of business layer into LTL formulas were defined. It should be noted, that ArchiMate syntax defines strictly, how elements of various types can be combined, however, without giving semantics to them (at least in case of behavioral elements). Hence, the rules can be considered to be an assignment of a semantics to language patterns.

After analysis of several examples we decided not to define rules for elements composed of lower level activities. ArchiMate syntax seems to manifest some flaws in this case. An activity modeled as a process, function or interaction can be a part (as defined by the composition relation) of several high-level behavioral elements. Moreover, a complex process can be composed of lower level activities, but junctions and internal events are not included into the composition. Therefore, we decided to treat complex processes as kinds of views helping to organize the models, rather than manageable entities.

VI. Deduction-based verification

System specification in form of LTL formulas $F_1, ...F_n$ obtained by applying rules defined in previous section can be checked for either its validity or entailment: $F_1, ...F_n \models G$. The second case is particularly interesting, as G can express a desired system property pertaining to temporal ordering of states and events. The approach proposed in this work consists in applying a semantic tableaux method to reason about entailment. The method is described briefly in Section VI-A, which is followed by Section VI-B giving an outline of a verification system architecture. Finally we present an example of specification in Section VI-C.

A. Semantic tableaux method

Semantic tableaux is a decision-making procedure for checking satisfiability of a formula. To do so, it shows that the negation of an initial formula cannot be satisfied, hence, the initial formula is a tautology. To verify an entailment $F_1, ...F_n \models G$ it suffice to prove that $\{F_1, ..., F_n, \neg G\}$ is unsatisfiable.

The main principle of propositional tableaux is to "break" complex formulae into smaller ones until complementary pairs of literals are produced or no further expansion is possible The method originates form classical logic but it can be also used for temporal logics [6]. Generally speaking, the method is based on well defined rules of formula decomposition and expansion. They allow to handle each of the logical connectives. When the rules are applied, branches of the inference tree are built. They correspond to alternatives appearing in formulas placed at the tree nodes. An inference tree is *finished*, when no formula can be further broken down, i.e. no complementary pairs of literals can be produced. The branches in the tree can be of two kinds: open or closed. A branch is closed if it can be established that a set of literal formulas, i.e. atomic formulas or its negations, on this branch has no model. In practice, this corresponds to a condition that a pair of

contradictory formulas can be found on the branch. If all branches of the tree have contradictions, the whole inference tree is *closed*. If the negation of the initial formula is placed in the root, this leads to the statement that the initial formula is true. A very simple yet illustrative example of the reasoning tree is shown in Fig. 8. The negation of the initial formula $(a \Rightarrow \Diamond b) \wedge (b \Rightarrow \Diamond c) \Rightarrow (a \Rightarrow \Diamond c)$ is placed in the root of the tree. All branches are closed (red nodes) what means that the initial formula is always satisfied.

B. Deduction based verification system

The architecture of the deduction-based verification system is shown in Fig. 7. The system consists of three components:

1) *Modeler* that allows to prepare and develop business models using ArchiMate language. In this case the Archi software, an excellent free modeling tool [1] was used.
2) *Generator* generates logical specifications from ArchiMate models. We have implemented a component that applies rules described in Section V to elements of a business layer and yields a set (a conjunction) of LTL formulas. It is deployed as a plugin to Archi.
3) *Prover* takes as input logical specifications (a set of temporal logic formulas describing a verified system) and a query, i.e. an examined property represented by a single formula, checks its validity and issues a response (*Yes* or *No*).

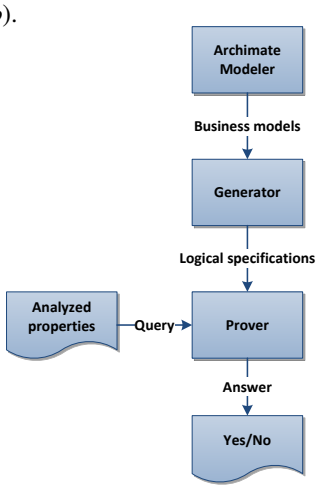


Fig. 7. An architecture of the deduction system

The prover is a crucial component of the verification system. Recently, a prototype reasoning engine for linear and future time minimal temporal logic was implemented[1], c.f. Fig. 8. It allows to examine logical validity for formulas expressing liveness or safeness, as described above. Internally, the prover applies the semantic tableaux method customized to requirements of reasoning on validity of LTL formulas.

An advantage of the described system (Fig. 7) is that it can give instantaneous response whenever the specification of a model is changed or there is a need for a new inference due to a newly introduced property.

[1]The engine was implemented as a students' (Joanna Kulesza and Kamil Łopata) project under the supervision of one of the authors of the work.

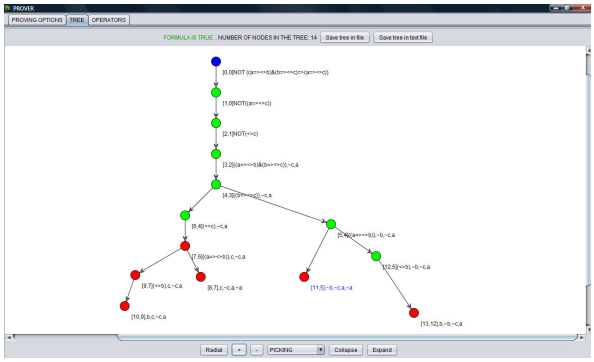


Fig. 8. A prototype system of inference using the semantic tableaux method

C. Example of verification

In this section we return to the ArchiMate model of a taxi service presented in Section IV. For this model 108 temporal formulas were generated. Due to the limited size of the work, it is not possible to show them all. Thus, a subset of the whole logical specification L referring to Fig. 1 and Fig 3.b-c, is shown below.

$$
\begin{aligned}
L = \{&\Box(Contact\_taxi\_driver\_start \Rightarrow \\
&\Diamond Contact\_taxi\_driver\_end), \Box\neg(Contact\_taxi\_driver\_end \wedge \\
&Trip\_accepted), \Box(Broadcasted\_trip\_order \Rightarrow \\
&\Diamond Junction\_0), \Box(Timeout \Rightarrow \\
&\Diamond Broadcast\_order\_start), \Box(Inform\_client\_start \Rightarrow \\
&\Diamond Inform\_client\_end), \Box(Inform\_client\_end \Rightarrow \\
&\Diamond Stop\_1), \Box(Timeout\_no\_confirmation \Rightarrow \\
&\Diamond Inform\_client\_about\_rejected\_order\_start), \\
&\Box(Trip\_accepted \Rightarrow \Diamond And\_Junction), \Box(Junction \Rightarrow \\
&\Diamond Look\_for\_free\_cabs\_start), \\
&\Box(Inform\_client\_about\_rejected\_order\_start \Rightarrow \\
&\Diamond Inform\_client\_about\_rejected\_order\_end), \\
&\Box(Inform\_client\_about\_rejected\_order\_end \Rightarrow \Diamond((Stop\_2 \wedge \\
&\neg Stop\_7) \vee (\neg Stop\_2 \wedge Stop\_7))), \Box(Client\_trip\_order \Rightarrow \\
&\Diamond Receive\_order\_start), \Box(And\_Junction \Rightarrow \\
&\Diamond(Inform\_client\_start \wedge \\
&Send\_confirmation\_start)), \Box(Look\_for\_free\_cabs\_start \Rightarrow \\
&\Diamond((Look\_for\_free\_cabs\_end \wedge \neg Timeout) \vee \\
&(\neg Look\_for\_free\_cabs\_end \wedge \\
&Timeout))), \Box(Look\_for\_free\_cabs\_end \Rightarrow \\
&\Diamond Contact\_taxi\_driver\_start), \Box(Receive\_order\_start \Rightarrow \\
&\Diamond Receive\_order\_end), \Box(Receive\_order\_end \Rightarrow \\
&\Diamond Junction), \Box(Broadcast\_order\_start \Rightarrow \\
&\Diamond Broadcast\_order\_end), \Box(Broadcast\_order\_end \Rightarrow \\
&\Diamond Broadcasted\_trip\_order), \Box(Junction\_0 \Rightarrow \\
&\Diamond((Broadcasted\_order\_confirmed \wedge \\
&\neg Timeout\_no\_confirmation) \vee \\
&(\neg Broadcasted\_order\_confirmed \wedge \\
&Timeout\_no\_confirmation))), \\
&\Box(Broadcasted\_order\_confirmed \Rightarrow \Diamond Trip\_accepted)\}
\end{aligned}
$$

Let us consider a liveness property expressed formally by the following formula

$$\Box(Client\_trip\_order \Rightarrow \Diamond(Stop\_1 \vee Stop\_7)) \qquad (3)$$

which can be understood that, *if a client ordered a trip then sometime in the future the client is informed about assigned cab (Stop_1) or the order is rejected (Stop_7).*

When analyzing if a specification L satisfies the property expressed by the formula (3) a new formula (4) is constructed

and submitted to the prover.

$$C(L) \Rightarrow (\Box(Client\_trip\_order \Rightarrow \Diamond(Stop\_1 \vee Stop\_7))) \quad (4)$$

where $C(L)$ means logical conjunctions of formulas.

Presentation of a full inference tree, which contains more than a thousand nodes, exceeds the size of the work. All branches of the semantic trees are closed, i.e. formula 4 is satisfied in the considered model. If the tree had open branches, this would indicate that the input formula can be not satisfied. In this case the prover would provide information about the source of the error, what can be considered an important advantage of the method.

VII. Conclusion

This paper discusses a problem of automatic verification of behavioral specification embedded within ArchiMate models. We propose to apply an approach consisting in translating ArchiMate specification into a set of LTL formulas and using deductive reasoning technique to check temporal properties of the model. Firstly, on a small example we present language patterns that can be used to model processes, then rules for translation of ArchiMate behavior specification into LTL formulas are formally defined. Finally, we describe the architecture of the implemented reasoning system and show how it can be used to check desired system properties.

Although the considerations in this work are focused on deductive reasoning and semantic tableaux method, automatically generated LTL specifications can be verified with other methods, e.g. the resolution method.

The defined set of rules for transforming ArchiMate models into LTL formulas considers only atomic processes and functions. It is an open question how to give semantics to explicitly specified high-level behavioral elements aggregating low-level behaviors. At present they are treated as views organizing models, however we are analyzing alternative approaches.

Acknowledgment

This work was supported by the AGH UST internal grant no. 11.11.120.859.

References

[1] "Archi, archimate modelling tool," 2013, [Online; accessed 23-April-2013]. [Online]. Available: http://archi.cetis.ac.uk/download.html

[2] B. Anderson, J. V. Hansen, P. Lowry, and S. Summers, "Model checking for e-business control and assurance," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 35, no. 3, pp. 445–450, 2005.

[3] C. Azevedo, J. Almeida, M. Van Sinderen, D. Quartel, and G. Guizzardi, "An ontology-based semantics for the motivation extension to archimate," in *Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International*, 2011, pp. 25–34.

[4] E. Clarke, O. Grumberg, and D. Peled, *Model Checking*. MIT Press, 1999.

[5] E. Clarke, J. Wing, and et al., "Formal methods: State of the art and future directions," *ACM Computing Surveys*, vol. 28 (4), pp. 626–643, 1996.

[6] M. d'Agostino, D. Gabbay, R. Hähnle, and J. Posegga, *Handbook of Tableau Methods*. Kluwer Academic Publishers, 1999.

[7] F. de Boer, M. Bonsangue, H. ter Doest, L. Groenewegen, H. Jonkers, A. Stam, and L. van der Torre, "Analysis of Enterprise Architectures (q4)," 2003. [Online]. Available: https://doc.telin.nl/dscgi/ds.py/Get/File-31618

[8] A. Deutsch, R. Hull, F. Patrizi, and V. Vianu, "Automatic verification of data-centric business processes," in *Proceedings of the 12th International Conference on Database Theory*. ACM, 2009, pp. 252–267.

[9] E. Emerson, *Handbook of Theoretical Computer Science*. Elsevier, MIT Press, 1990, vol. B, ch. Temporal and Modal Logic, pp. 995–1072.

[10] R. Ettema and J. Dietz, "Archimate and demo – mates to date?" in *Advances in Enterprise Engineering III*, ser. Lecture Notes in Business Information Processing, A. Albani, J. Barjis, and J. Dietz, Eds. Springer Berlin Heidelberg, 2009, vol. 34, pp. 172–186. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-01915-9_13

[11] X. Fu, T. Bultan, and J. Su, "Formal verification of e-services and workflows," in *Web Services, E-Business, and the Semantic Web*. Springer, 2002, pp. 188–202.

[12] M. Gruninger and M. S. Fox, "An activity ontology for enterprise modelling," *Department of Industrial Engineering, University of Toronto*, 1994.

[13] R. Klimek, "From extraction of logical specifications to deduction-based formal verification of requirements models," in *Proceedings of 11th International Conference on Software Engineering and Formal Methods (SEFM 2013), 25–27 September 2013, Madrid, Spain*, ser. Lecture Notes in Computer Science, R. M. Hierons, M. G. Merayo, and M. Bravetti, Eds., vol. 8137. Springer Verlag, 2013, pp. 61–75. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40561-7_5

[14] M. Mongiello and D. Castelluccia, "Modelling and verification of BPEL business processes," in *Model-Based Development of Computer-Based Systems and Model-Based Methodologies for Pervasive and Embedded Software, 2006. MBD/MOMPES 2006. Fourth and Third International Workshop on*. IEEE, 2006, pp. 5–pp.

[15] S. Morimoto, "A survey of formal verification for business process modeling," in *Proceedings of the 8th International Conference Computational Science (ICCS 2008), June 23–25, 2008, Kraków, Poland, Part II*, ser. Lecture Notes in Computer Science, M. Bubak, G. D. van Albada, J. Dongarra, and P. M. A. Sloot, Eds., vol. 5102. Springer–Verlag, 2008, pp. 514–522.

[16] M. Nick, "Will there be a battle between Archimate and the UML?" 2009. [Online]. Available: http://blogs.msdn.com/b/nickmalik/archive/2009/04/17/will-there-be-a-battle-between-archimate-and-the-uml.aspx

[17] OMG, "Business Process Model and Notation (BPMN) version 2.0," OMG, Tech. Rep., January 2011. [Online]. Available: http://www.omg.org/spec/BPMN/2.0

[18] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual, The (2nd Edition)*. Pearson Higher Education, 2004.

[19] A.-W. Scheer and M. Nüttgens, "ARIS architecture and reference models for business process management," in *Business Process Management*. Springer, 2000, pp. 376–389.

[20] A. Scheer, *Aris - Business Process Modeling*, ser. ARIS - Business Process Modeling. Springer, 1999, no. v. 2.

[21] N. Shankar, "Automated deduction for verification," *ACM Computing Surveys*, vol. 41 (4), pp. 20:1–20:56, 2009.

[22] P. Szwed, W. Chmiel, S. Jedrusik, and P. Kadluczka, "Business processes in a distributed surveillance system integrated through workflow," *Automatyka/Automatics*, no. to appear, 2013.

[23] The Open Group, "Open Group Standard. Archimate 2.0 Specificattion," 2012. [Online]. Available: http://www.opengroup.org

[24] The Workflow Management Coalition, "Process Definition Interface - XML Process Definition Language, Workflow Management Coalition Workflow Standard, version 2.1a," The Workflow Management Coalition, Tech. Rep., 2008. [Online]. Available: http://www.wfmc.org/Download-document/WFMC-TC-1025-Oct-10-08-A-Final-XPDL-2.1-Specification.html

[25] J. van Benthem, *Handbook of Logic in Artificial Intelligence and Logic Programming*, ser. 4. Clarendon Press, 1993–95, ch. Temporal Logic, pp. 241–350.

[26] H. Van Den Berg, H. Bosma, G. Dijk, H. Van Drunen, J. Van Gijsen, F. Langeveld, J. Luijpers, T. Nguyen, R. Oosting, Gerand Slagter, and et al., "ArchiMate made practical," *Work*, 2007.

[27] W. M. van der Aalst, J. Desel, and E. Kindler, "On the semantics of EPCs: A vicious circle," in *Proceedings of the EPK*, 2002, pp. 71–80.

[28] K. Watahiki, F. Ishikawa, and K. Hiraishi, "Formal verification of business processes with temporal and resource constraints," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1173–1180.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1111–1118

Design of Financial Knowledge in Dashboard for SME Managers

Jerzy Korczak, Helena Dudycz, Mirosław Dyczkowski
Wroclaw University of Economics
ul. Komandorska 118/120 53-345 Wroclaw, Poland
Email: {jerzy.korczak, helena.dudycz, miroslaw.dyczkowski}@ue.wroc.pl

Abstract — The article presents the approach to develop the economic and financial knowledge used for the Intelligent Dashboard for Managers. The content of the knowledge is focused on essential concepts related to the management of micro, small and medium enterprises. Knowledge-based functions, not previously available in commercial systems, increase the quality, effectiveness, and efficiency of the decision making process. The Intelligent Dashboard for Managers contains six ontologies describing areas of Cash Flow at Risk, Comprehensive Risk Measurement, Early Warning Models, Credit Scoring, Financial Market, and General Financial Knowledge. The ontology design process and examples of topic maps and usage in financial data analysis are illustrated and discussed.

I. Introduction

Decision makers of small and medium enterprises (SMEs) using Business Intelligence systems frequently need appropriate knowledge about the economic situation of the enterprise and its environment. Knowledge about key dependences between various financial ratios is essential, because they can indicate important trends and alert one to anomalies and dangers [23, p. 85]. Decision makers in SMEs, in comparison to managers of big companies, may have no access to much essential strategic information. Usually, financial expertise is either not available or too expensive. For financial and personnel reasons most SMEs cannot afford these types of facilities. Furthermore, SMEs operate in a definitely more uncertain and risky environment than big enterprises, because of a complex and dynamic market that has much more important impact on SMEs' financial situation than on big companies'. Tolerance of mistakes is narrower (see among others [12, p. 74-91]). In these conditions, SME's decision makers often act intuitively and as a result, the rationality of their decisions is decidedly smaller. Moreover, the statistics show that SME's decision makers often don't have a solid knowledge of economics and finance.

In general, most existing Business Intelligence (BI) and Executive Information Systems (EIS) provide the functionality of data aggregation and visualization. Many reports and papers in this domain underline that decision makers expect new ICT solutions to interactively provide not only relevant and up-to-date information on the financial situation of their companies, but also explanations taking into account the contextual relationships.

The aim of this article is to present the approach to developing the economic and financial knowledge used in the Intelligent Dashboard for Managers (called further InKoM). The InKoM system has been developed by a consortium consisting of the Wroclaw University of Economics (WUE), which is the leader, and a company UNIT4 TETA BI Center Ltd. (TETA BIC). Credit Agricole Bank Polska S.A. also participates in the project.

Figure 1 presents the main components of the InKoM: a comprehensive description of the TETA BI system with examples of its application is available on the website: [28] (see also [3], [31]). It can be seen that the InKoM uses TETA BI mechanisms for extracting source data from transactional systems (ETL), its data warehouse, and analytical database. However, the available solutions – in particular the standard analyses, reports and analytical statements generated by the system – are complemented by economic and financial knowledge – most importantly ontologies and topic maps – and financial data mining algorithms, including mechanisms for extracting business knowledge from the deep Web. This enables a dynamic, on-line, interactive analysis of key business indicators.

The transactional data obtained from external sources, supplemented with planning data, e.g., budgets, form multidimensional data structures, or cubes, which are stored in a TETA BI Analysis Services database and provide a basis for the on-line, interactive creation of standard analytical queries and/or reports. The InKoM system complements and extends these processes[1]. By providing

The work is supported by the National Research and Development Centre within the Innotech program (track In-Tech), grant agreement INNOTECH-K1/IN1/34/153437/NCBR/12.

[1] The InKoM architecture and functionalities have been presented in [18]; [19].

978-83-60810-53-8/$25.00

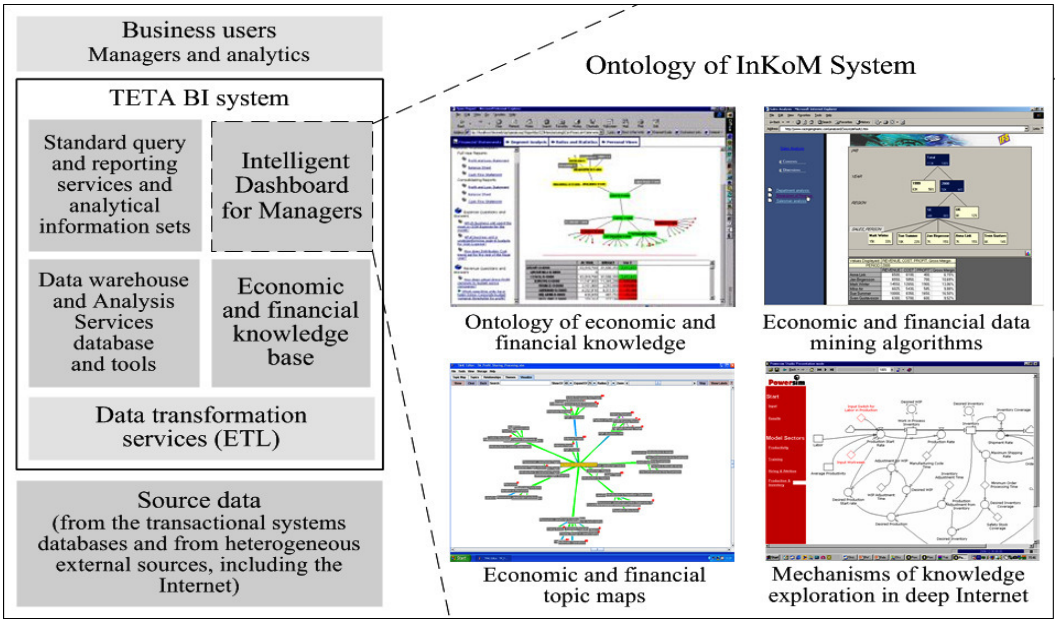


Fig. 1. Components of the Intelligent Dashboard for Managers and their location in the TETA BI system

economic and financial knowledge stored in ontologies and presented in the form of topic maps to facilitate the perception of concepts, InKoM can make the analysis more comprehensive and simpler. This is particularly important for users who are not specialists in the analysis and interpretation of economics and finance.

Over a dozen methods of building ontologies have been developed. Among methods listed in literature, the following are worth noting (more widely characterized inter alia in [13]; [22]): METHAONTOLOGIA (based on the standard IEEE 1074-1995), the method of Noy and McGuinness, On-To-Knowledge, SENSUS, TOVE (TOronto Virtual Enterprise), UPON (Unified Process for ONtology building) and the method of Ushold and King. Noy and McGuinness emphasized that the best model of ontology is the one that best cooperates with existing information system, accomplishes a set of goals, is intuitive and easy to maintain [Noy, McGuinness 2005]. So far there is no single standard of design because creating an ontology is dependent on its application and the needs of specific users (see [22]). However during last years the standardization activities in terms of information have been undertaken at European level; an example is the SMEST project (see http://www.cencenelec.eu/sme/SMEST/).

The structure of the paper is the following: In the next section the ontological approach to knowledge design is described. The third section presents the ontology structure. The fourth section characterizes the ontology design process in the InKoM. To illustrate the use of the InKoM, a case study of ontology for credit scoring is described. The last section summarizes the work already carried out and points out the further research and development tasks.

II. Ontological Approach to Knowledge Design

The main goal of any BI system is to access the right data at the right time to allow proactive decision-making (see among others [9], [26], [30]). The users of BI systems expect access to useful information through an interface easy to understand and use. However, existing BI solutions are designed primarily for users who are able to understand the business data models (see [25]). BI systems provide simple, personal analytical tools which support the exploration of data sources, retrieval of information based on predefined economic and financial relations, and do not require a priori knowledge about data structures and methods of data accessing (see among others [21], [25], [26]).

Today the development of new BI systems is oriented towards BI 2.0 (using semantic search) and 3.0, Service Oriented Architecture (SOA) and Software as a Service (SaaS) (see [21], [25], [26]). The typical features of the systems include: proactive alerts and notifications, event driven (real time) access to information, advanced and predictive analytics, mobile and ubiquitous access, improved visualization, and semantic search information (see also [21]).

One of the main part of modern BI systems is the ontology (see [21]). Ontology in information technology means „an explicit specification of a conceptualization“ [15, p. 907]. In general, the ontology is used to create the necessary knowledge models for defining functionalities in analytical tools. In the development of InKoM, many new features are integrated, such as domain ontology covering key concepts of corporate finance and economics, knowledge discovery algorithms, semantic search

mechanisms, explanation facilities, and tools for visual navigation in domain knowledge.

In the InKoM project, six ontologies were built. Choosing scopes of created ontologies was a difficult task, because of domain complexity and the necessity to develop numerous links between the theory of finance and the contents of BI system databases. The task needed cooperation in pragmatic way with experts on the analysis of finance, economics and business informatics. Experience of our industrial partner, UNIT 4 TETA BI Center, gained from implementing the TETA BI system in enterprises, was also essential in choosing these areas. The result of this work is six ontologies covering economic and financial areas: Cash Flow at Risk, Comprehensive Risk Measurement, Early Warning Models, Credit Scoring, the Financial Market, and General Financial Knowledge. The ontologies will be detailed in the next section.

Integration of these ontologies into the BI systems assures:

- support for the definition of business rules in order to get proactive information and advice in decision-making;
- a semantic layer describing relationships between the concepts and indicators;
- relevant information according to the different kinds of users that can be found in an organization;
- effective usage of existing data sources and data warehouse structure.

The knowledge representation layer is the most critical aspect of a BI system, since it broadly shapes the core understanding of the information displayed on their screen [30]. In InKoM design, the basic assumption of navigation was that managers should be able to view focus and context areas at the same time to present the relevant knowledge structure [27].

Visual exploration in InKoM is based on a standard Topic Map (TM – ISO/IEC 13250:2003). TM enables the representation of complex structures of knowledge bases [4] and the delivery of a useful model of knowledge representation (see [20, p. 174]), where multiple contextual indexing can be used. TM is a relatively new form of the presentation of knowledge, which puts emphasis on data semantics and the ease of finding desired information (see also [1]; [24, p. 30]). The application of topic maps permitted us to separate the data of the enterprise's information system from operational business activities (see among others [17]). Developed topic maps for analysis of economic indicators (see among others [7], [8], [17]) have demonstrated that the system [6]:

- can be easily used for the representation of economic knowledge about economic and financial measures,
- can express the organizational structure,
- can be adapted to new applications and managers' needs,
- can be supportive of the managerial staff by facilitating access to a wide range of relevant data resources,
- can assure a semantic information search and interpretation for non-technically-minded users,
- can visualize different connections between indicators that make possible the discovery of new relations between economic ratios constituting knowledge still unknown in this area,
- can improve the process of data analysis and reporting by facilitating the obtaining of data from different databases in an enterprise, and finally
- can be easily extended by users who are not IT specialists, e.g. by experts in economic analysis (using tools for creating a topic map application).

The preliminary evaluation of ontologies is very encouraging The knowledge of corporate finance is very useful in the interpretation and explanation of data presented in financial reports of BI systems.

III. Ontology Structure

The essential element of the InKoM systems' knowledge is a part containing topic maps for six created ontologies for selected areas called: Cash Flow at Risk, Comprehensive Risk Measurement, Early Warning Models, Credit Scoring, Financial Market, and General Financial Knowledge.

CFaR is considered to be one of the best adjusted ratios for the needs of complex risk measurement in enterprises. According to the newest trends in enterprise management and on companies' economic practice, it was determined that the appropriate ratio for complex risk management in enterprises is Cash Flow at Risk (CFaR) using the RiskMetrics Group rules. The choice of Cash Flow at Risk meets managers' needs in the scope of creating single risk ratio illustrating risk level connected with an enterprise's operating activity. Of course, the prerequisite for the manager being able to use this information is the appropriate knowledge level on CFaR ratio.

The ontology of Comprehensive Risk Measurement involves various variants of the model used to estimate the Cash Flow at Risk ratio. The developed ontology concerns the way of understanding and defining inter alia: risk variables, risk models, risk management process.

The ontology of the Early Warning System contains models of cautionary forecasting in supporting a manager's decision. The developed concept uses a group of tools for early warning model creation, in which mainly data from financial reports are used. By using this data, an objective view of company's situation can be quickly and easily obtained, which is essential in managers' efficient and accurate decision making.

The ontology of Credit Scoring describes the model of the credit procedure carried out by a bank. Credit scoring evaluation is an integral part of bank credit procedures (credit processes) and it is an essential stage in conditioning the granting of credit. It is realized by a bank's

organizational units and supporting infrastructural-system solutions, and the company trying to get credit (potential borrower) is the subject of analysis (quantitative and qualitative). The internal logic and details of credit procedures in a specific bank is its unique know-how and is not usually made public.

In literature many approaches to credit scoring can be found, such as stochastic models (e.g. Bayesian models, regression, Markov chain), artificial intelligence techniques (e.g. expert systems, neural network models, genetic algorithms), data mining methods [5]; [10]; [11]; [16]; [32]. The approach implemented in the InKoM system is based on the rules of thumb generally accepted in the practice of financial institutions. Of course there are still open questions which might be considered: which methodology to choose? what models might work? should one look at the profit on each product in isolation or the total profit over all possible products? In general there is no overall best method of credit scoring. The InKoM system is open to include more advanced models and rules, specific to the particular type of company or financial institution.

The Financial Market ontology involves information about the financial market and financial instruments. Knowledge on this topic is an essential element capable of aid manager in making investment decisions and securing from market risk. Financial market may be used by managers to regulate liquidity by using money market instruments.

The ontology of General Financial Knowledge concerns essential economic-financial knowledge which is required to analyze issues of listed ontologies. This ontology includes a set of supplementary topics to other ontologies and will be used in calculating the value of Cash Flow at Risk (the ontology of Cash Flow at Risk), basic economic ratios (the ontology of Early Warning System) and indicators used by banks in the process of credit scoring (the ontology of Credit Scoring).

To sum up, the domain knowledge about relations between economic and financial ratios will make the analysis and interpretation of contextual connections easier. This is very important in the case of SMEs, where a company does not employ experts in economic-financial analysis and using outer consulting is too costly. Reproducing knowledge with the use of a topic map contributes inter alia to better understanding of economic concepts and the interpretation of specific economic and financial indicators.

To navigate in the financial knowledge of InKoM, a semantic search will be applied to avoid difficulties related to decision makers' interpretation of economic and financial information. This gives the opportunity to search data sources taking into account not only structural dependences, but also semantic context.

IV. Ontology Design Process

In the design of the InKoM system five basic stages of creating the ontology were defined. These are:

1. Definition of the goals, scope and constraints of the created ontology. While creating an ontology, assumptions about the created model of knowledge that will apply during its building have to be provided. That requires an answer to the question: what will the created ontology be used for?
2. Conceptualization of the ontology. This is the most important stage in the procedure of ontology development[2]. It includes the identification of all notions, definition of classes and their hierarchic structures (Superclass – Subclass), modeling relations, identification of instances, specification of axioms, and rules of inference.
3. Verification of the ontology's correctness by experts. In this stage the constructed ontology is verified by experts who did not participate in the process of creating it.
4. Coding the ontology in the format compatible with the topic map standard. During this stage the developed ontology is described in the formal language or chosen software. The result of this stage should be the encoded ontology[3].
5. Validation and evaluation of the built ontology. It is the stage during which evaluation of the created ontology meets the needs of the managers.

The important stage in the described procedure of creating an ontology is the conceptualization of the ontology. The process of conceptualization of an ontology is an intellectual activity organizing knowledge from a given field carried out by the person, either an expert or collaborating with an expert, responsible for creating the model of knowledge without the support of automated tools (see inter alia [2, p. 2036]). In this scope, there are few concepts of identification of topics and relations between them within the process of conceptualization. These are the following approaches to carrying out analysis: top-down, bottom-up and middle-out. We used middle-out, because it enables us best to maintain the level of detail control of the created ontology and reduce imprecisions, which translates into reducing iterative work (see [29, p. 21]; see also [2, p. 2036]). Based on literature studies (inter alia [13], [22]), as well as research carried out, the following procedure in conceptualization of the ontology of economic knowledge was used:

1. Identification and definition of all topics.

[2] Independently of the field that is to be modeled by using an ontological approach, it is the most important stage in creating a model based on ontology (see inter alia [2, p. 2036]).

[3] In the InKoM project at first, in order to quickly test developed ontology, it was entered in the program Ontopia. Ultimately it will be realized with the use of software for topic maps developed by the company UNIT 4 TETA BI Center.

A topic, representing any concept, is "a syntactic construct that corresponds to the expression of a real-world in a computer system" [14, p. 60]. In the InKoM project, a topics' list was determined by experts creating ontologies for the given field of economic knowledge. These topics include, beside their names, also their synonyms and descriptions (table 1). The example below illustrates a description of topics related to the case study: Credit Scoring that will be detailed in Section VI.

TABLE 1.
THE EXAMPLE OF TOPICS' LIST

Name	Synonym	Description
Return on Assets	ROA	ROA indicator is a synthetic measure that determines company's assets capability to create profit. It shows the percentage net profit per unit of capital invested in the company.

2. Creating a taxonomy of topics.
Specification of taxonomic relations between distinguished topics and defining classes and subclasses (table 2). This relationship describes the topics generalization. This approach to creating a taxonomy is proposed in METHAONTOLOGIA (see i.e. [13]).

TABLE 2.
THE EXAMPLE OF TOPICS' TAXONOMY

Superclass	Subclass
Indicators	– Debt indicators – Liquidity indicators – Profitability evaluation indicators – Efficiency assessment indicators
Profitability evaluation indicators	– Net Return of Sale – Return on Assets – Return on Equity

3. Definition of all other types of relations between topics.
In the InKoM project, the basic relationship aggregate of (Aggregate – Member) occurring in all six created ontologies was defined. Moreover within each ontology, additional relations were defined, for example: engagement (Engaging – Accession).

4. The list of all the individual relationships existing in the ontology.
The list includes: the name of the relationship, source topic, and target topic (table 3).

TABLE 3.
THE EXAMPLE OF DESCRIPTION OF THE RELATIONSHIPS

Name of relationship	Source topic	Target topic
Engagement	Profitability evaluation	Net Return on Sale
Engagement	Profitability evaluation	Return on Assets
Engagement	Profitability evaluation	Return on Equity

5. Description of functions and rules.
The following function description, specifies the example of the indicator Return on Assets (ROA), implemented in the InKoM system:

Name:
Indicator Return onAssets (ROA)
Input:
- Result of Net profit (NP)
  type: number, value with balance sheet
- Total Assets (TA)
  type: number, value with balance sheet

Output:
- Return on Assets

Initial conditions:
- available data from balance sheet

Final conditions:
- Message 1: *"Value of indicator ROA"*
- If ROA<5%,
  Then to Message *2: "*Low value of ROA*";*
- If 5% < ROA< 10%,
  Then to Message 3: *"*Average value of ROA*";*
- If ROA > 10%,
  Then to Message 4*: "*High value of ROA*";*

Description/formula:
ROA = NP/ TA

6. Description of usage scenarios.
Usage scenarios, also called use case view, describe demonstration analyses of economic topics occurring in this ontology. For example, if a manager is interested in the opportunity of applying for a bank loan:

1. The manager analyzes the semantic network, from which it follows that the credit score assessment is based on the analysis of indicators belonging to four groups: debt indicators, liquidity indicators, profitability evaluation, and efficiency assessment.
2. From the TETA BI system the manager receives the values of financial indicators that make up the credit score assessment. According to them, the company has bad parameters concerning the profitability evaluation, especially the value of ROA indicator.
3. The manager analyzes the semantic network connected to the data from the TETA BI system on account of semantic connections of the ROA indicator with other indicators. The aim of the action undertaken by the manager is to identify causes of unfavorable values from the ROA indicator.
4. Basing on conclusions from the analysis conducted of economic indicators, the manager can undertake corrective actions which may potentially result in improving the company's condition. Improvement of company's parameter essentials in the score assessment can allow actions to be undertaken concerned with receiving bank loan.

That work has required multi-domain expert knowledge, both theoretical and practical, in economics, finance, and informatics.

In InKoM, a semantic search is provided to avoid difficulties related to decision makers' interpretation of

economic and financial information. This gives the opportunity to search data sources taking into account not only structural dependences, but also the semantic context.

V. CASE STUDY – ONTOLOGY FOR CREDIT SCORING

The case study presented illustrates only a fragment of analysis of the company's financial situation with the use of the ontology of the credit scoring and the databases of TETA Business Intelligence system. Let us assume that the manager is interested in the ROA indicator. The ontology shows that the quantitative financial analysis consists of indicators: debt indicators, liquidity indicators, profitability evaluation indicators, and efficiency assessment indicators (Figure 2). On the Figure 2, solid lines denote taxonomic relations (relation Superclass – Subclasses), whereas broken lines denote all other types of relations between topics.

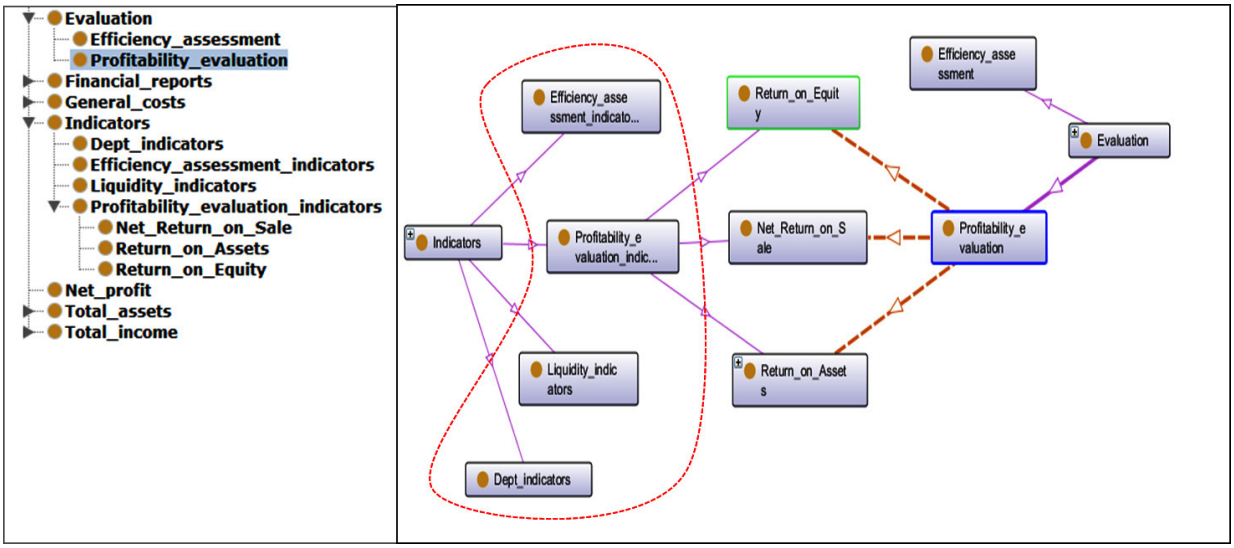


Figure 2. Example of topic map including the quantitative financial indicators

Returning to the ROA, it is known that the more effectively a company manages its assets, the higher the value of this indicator. It is assumed that its value should be greater than 5%. The ROA indicator is calculated on data from the balance report, that is Net profit (NP) and Total Assets (TA) (Figure 3):

$$ROA = NP / TA$$

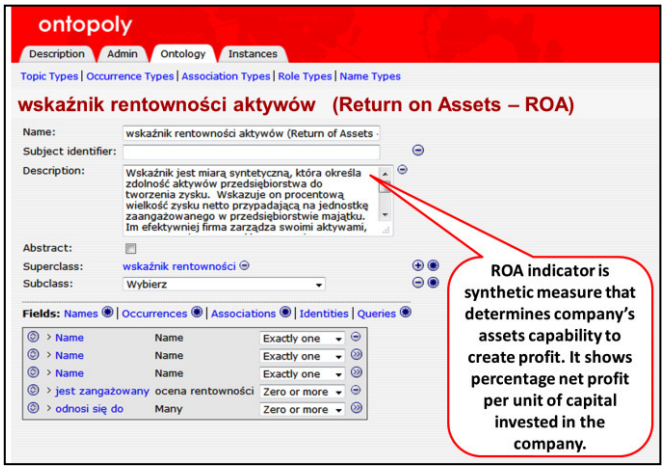


Figure 3. Topic definition – ROA

in year 2011 is equal to 12 908,00, and in the year 2012 equals 14 169,00, whereas the value of Total Assets in the year 2011 equals 184 400,00 and in the year 2012 354 000,00. Based on this data, the ROA indicator in 2011

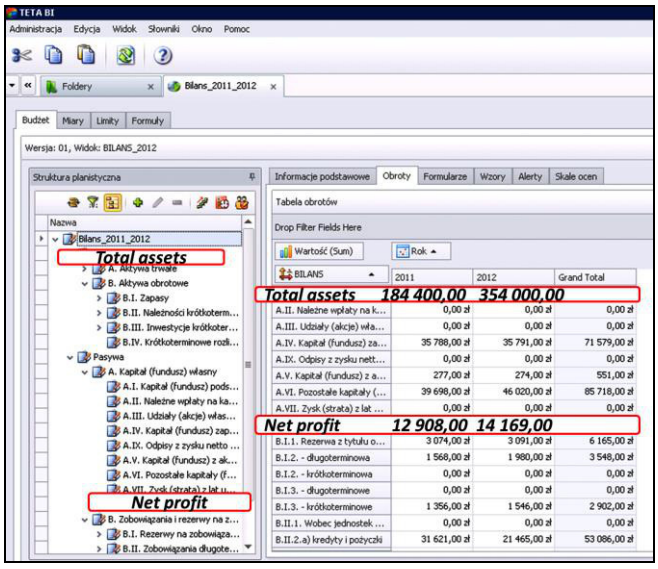


Figure 4. Example of balance sheet extracted from the TETA BI system

On figure 4 is presented an example of a balance report for the years 2011 and 2012 extracted from the TETA BI system. For the analyzed company, the value of Net Profit is equal to 7%, and in 2012 is equal to 4%. This means that the company's condition in 2012 is bad. In order to improve this situation, the manager needs to identify the cause of

these values. Therefore, following the topic map, the manager notes that the value of Net Profit in 2012 was higher than in 2011, so the disadvantageous ROA indicator results from value of Total Assets. To identify the sources of the ensuing situation, the manager has to examine the semantic relations between Total Assets and other financial topics. Exemplary decomposition of the chosen topic is presented on the Figure 5. The screenshot shows the expansion of the selected topic: Total Assets. On the diagram it is the area encircled by a dashed line, with new topics being a subclass of topic Total Assets.

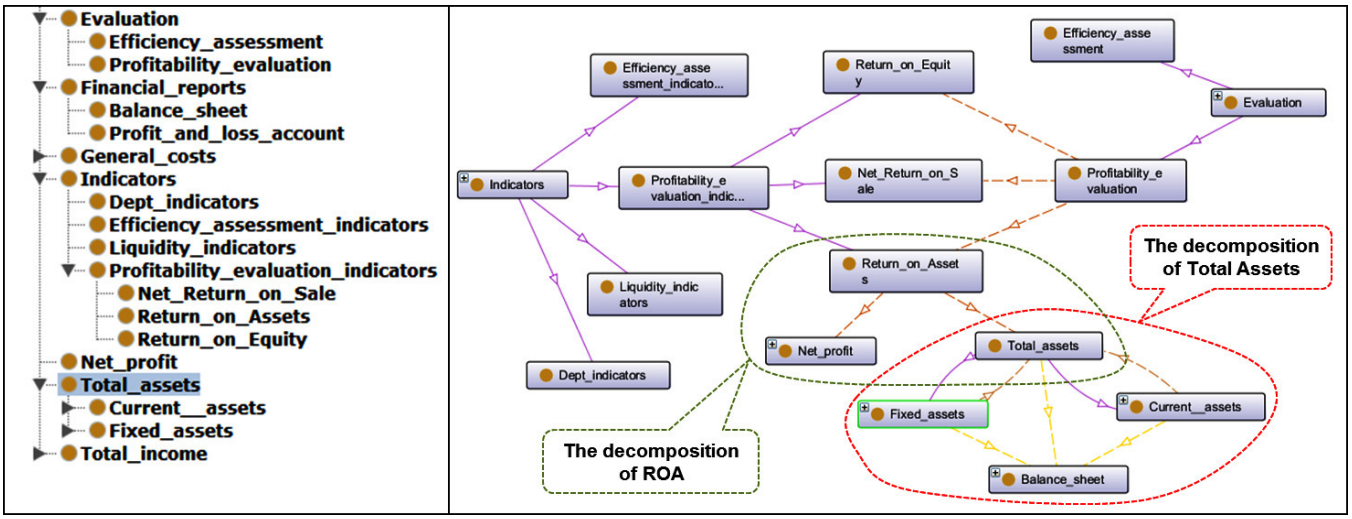


Figure 5. Example of topic map including Total Assets

In the InKoM system the manager can use the topic map application which allows relations between analyzed topics to be shown. This approach can be very useful for managers, because it is difficult a priori to identify causes that show by using only the balance sheet or other financial reports.

To interpret financial reports correctly, many measures and ratios need to be examined that either directly or indirectly influence the final result. Visualizing the relationships explicitly not only it makes the interpretation of indicators easier, but also contributes to finding out explanations of the current values of indicators. The topic map provides a user-friendly interface which allows managers to navigate easily from topic to topic in an interactive manner. Therefore various types of associations are visualized in different ways. For example, the lines which have the same relation have the same color. This enables an easy overview of concepts, visualisations and navigation of hierarchical structures, whilst also providing short definitions on each topic. The visualization is highly interactive: interesting nodes can be put in the foreground with zooms, translations, and rotations. Managers can delete non relevant branches of the graph or expand interesting ones.

VI. Summary And Future Work

In this paper, the approach to developing the ontology of InKoM was presented. The created ontologies are currently integrated in the TETA BI system and will be soon available in beta version. In the next step of the project, the effectiveness of the final system will be examined and assessed by SME managers. Further studies will be conducted on, among others, representation and interpretation of financial indicators, such as NPV and IRR, and methods specific to IT investment, such as TCO.

References

[1] Ahmed K., Moore G., "Applying topic maps to applications", The Architecture Journal, 2006, January, http://msdn.microsoft.com/en-us/library/bb245661.aspx.

[2] Almeida M. B., Barbarosa R. R., "Ontologies in Knowledge Management Support: A Case Study", Journal of the American Society for Information Science and Technology, 2009, nr 10 (60), s. 2032-2047.

[3] *Architektura system. TETA Business Intelligence. Materiały* informacyjne, UNIT4 TETA Business Intelligence Center, Wrocław 2011.

[4] Arndt H., Graubitz H., Jacob S., Topic map based indicator system for environmental management systems, 2008, http://www.iai.fzk.de/Fachgruppe/GI/litArchive.

[5] Desai V.S., Convay D.G., Crook J.N., Overstreet, G.A., "Credit scoring models in the credit union environment using neural networks and genetic algorithms", IMA Journal of Mathematics Applied in Business and Industry no 8, 1997, pp. 323–346.

[6] Dudycz H., "The concept of using standard topic map in Business Intelligence system", in: Proceedings of the 5th International Conference for Entrepreneurs, Innovation and Regional Development – ICEIRD 2012, D. Birov, Y. Todorova, eds., St. Kliment Ohridski University Press, Sofia, Bulgaria 2012, ISBN 978-954-07-3346-3, pp. 228-235.

[7] Dudycz H., "Visual analysis of economical ratios in Du Pont model using topic maps", in: Proceedings of the 4th International Conference for Entrepreneurs, Innovation and Regional Development – ICEIRD 2011, R. Polenakovik, B. Jovanovski, T. Velkovski, Eds., National

Center for Development of Innovation and Entrepreneurial Learning, Ohrid-Skopje, Macedonia 2011, Book of abstract ISBN 978-608-65144-1-9, p. 39 & CD with full papers ISBN 978-608-65144-2-6, pp. 277-284.

[8] Dudycz H., "Research on usability of visualization in searching economic information in topic maps-based application for return on investment indicator", in: Advanced Information Technologies for Management - *AITM'2011. Intelligent Technologies and Applicati*ons, J. Korczak, H. Dudycz, M. Dyczkowski, Eds., Wroclaw University of Economics Research Papers no 206, Wrocław 2011, pp. 45-58.

[9] Dudycz H., "Visualization methods in Business Intelligence systems – an overview", in: Business Informatics (16). Data Mining and Business Intelligence, J. Korczak Ed., Research Papers of Wroclaw University of Economics, 2010, no. 104, pp. 9-24.

[10] Fishelson-Holstine H., "Case studies in credit risk model development", in: Credit risk modeling, E. Mays, Ed., Glenlake Publishing, Chicago, 1998, pp. 169–180.

[11] Fung R., Lucas A., Oliver R., Shikaloff N., Bayesian networks applied to credit scoring, in: Proceedings of Credit Scoring and Credit Control V, Credit Research Centre, University of Edinburgh 1997.

[12] Gibcus P., Vermeulen P.A.M., Jong J.P.J., "Strategic decision making in small firms: a taxonomy of small business owners", International Journal of Entrepreneurship and Small Business, 2009, vol. 7, no. 1, pp. 74-91.

[13] Gomez-Perez A., Corcho O., Fernandez-Lopez M., "Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web", London: Springer-Verlag, 2004.

[14] Grant B. L., Soto M., "Topic maps, RDF Graphs, and ontologies visualization" in: Visualizing the Semantic Web. XML-based Internet and information visualization, second edition, V. Geroimenko, C. Chen Eds., Springer-Verlag London 2010, pp. 59-79.

[15] Gruber T. R., "Toward principles for the design of ontologies used for knowledge sharing*"*, International Journal Human-Computer, 1993, Studies 43, pp. 907-928.

[16] Hand D.J., Henley W.E., Statistical classification methods in consumer credit, Journal of the Royal Statistical Society, Series A 160, 1997, pp. 523–541. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus (83).

[17] Korczak J., Dudycz H., "Approach to visualization of financial information using topic maps", in: Information Management, B. F. Kubiak, A. Korowicki, Eds., Gdansk University Press, Gdansk 2009, pp. 86-97.

[18] Korczak J., Dudycz H, Dyczkowski M., "Intelligent Dashboard for SME Managers. Architecture and Functions", in: Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS 2012. M. Ganzha, L. Maciaszek, M. Paprzycki, Eds., Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, Warsaw, Los Alamitos, CA 2012, pp. 1003–1007.

[19] Korczak J., Dudycz H., Dyczkowski M.: "Intelligent decision support for SME managers – project InKoM", [in:] Business Informatics, J. Korczak, H. Dudycz, M. Dyczkowski, Eds., Wrocław University of Economics Research Papers 2012, no 3 (25), pp. 84-96.

[20] Librelotto G.R., Azevedo R.P., Ramalho J.C., Henriques P.R., "Topic maps constraint languages: understanding and comparing", International Journal of Reasoning-based Intelligent Systems, 2009, vol. 1, no. 3/4, pp. 173-181.

[21] Nelson G.S., Business Intelligence 2.0: Are we there yet?, SAS Global Forum 2010, http://support.sas.com/resources/papers/proceedings10/040-2010.pdf.

[22] Noy F. N., McGuinness D.L., "Ontology Development 101: A Guide to Creating Your First Ontology", 2005, http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html.

[23] Olszak C., "Wybrane technologie informatyczne w doskonaleniu rozwoju systemów Business Intelligence", in: *Zastosowania systemów informatycznych zarządzania*, W. Chmielarz, J. Kisielnicki, T. Parys, O. Szumski Eds., „Problemy Z*arządzania*", zeszyt specjalny 2011, Wydawnictwo Naukowe Wydziału Zarządzania, Uniwersytet Warszawski, Warszawa 2011, pp. 85-96.

[24] Pimentel M.P., Suárez J., Caparrini F.S., "Topic maps for philological analysis", in: Linked Topic Maps. Fifth International Conference on Topic Maps Research and Applications, TMRA 2009, L. Maicher, L.M. Garshol Eds., Leipziger Beiträge zur Informatik, Band XIX, Leipzig, pp. 29-39.

[25] Raden N., Business Intelligence 2.0: simpler, more accessible, inevitable, February 01, 2007, http://www.informationweek.com/news/software/bi/197002610.

[26] Sell D., Cabral L., Motta E., Domingue J., Pacheco R., adding semantics to Business Intelligence, 2008, http://dip.semanticweb.org/documents/WebSpaperOUV2.pdf.

[27] Smolnik S., Erdmann I., "Visual navigation of distributed knowledge structures in groupware – base organizational memories", Business Process Management Journal, 2003, vol. 9, no. 3, pp. 261-280.

[28] TETA Business Intelligence, UNIT4 TETA Business Intelligence Center, http://tetabic.eu/pl/aplikacja.html.

[29] Uschold M., Gruninger M., "Ontologies: Principles, methods and applications", Knowledge Engineering Review, 1996, vol. 11,no. 2, pp. 93-155.

[30] Wise L., The emerging importance of data visualization, part 1, October 29, 2008, http://www.dashboardinsight.com/articles/business-performance-management/the-emerging-importance-of-data-visualization-part-1.aspx.

[31] We change data into knowledge. TETA Business Intelligence. *Materiały informacyjne* UNIT4 TETA Business Intelligence Center, Wrocław 2011.

[32] Yobas M.B., Crook J.N., Ross P., Credit scoring using neural and evolutionary techniques, Working Paper no. 97/2, Credit Research Centre, University of Edinburgh 1997.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1119–1126

Risk avoiding strategy in multi-agent trading system

Jerzy Korczak
Wrocław University of Economics
ul. Komandorska 118/120, 53-345
Wrocław, Poland

Marcin Hernes
Wrocław University of Economics
ul. Komandorska 118/120, 53-345
Wrocław, Poland

Maciej Bac
Wrocław University of Economics
ul. Komandorska 118/120, 53-345
Wrocław, Poland

e-mail: {jerzy.korczak, marcin.hernes, maciej.bac}ue.wroc.pl

***Abstract*—The authors of this paper present an approach to trading strategy design for a multi-agent system which supports investment decisions on the stock market. The individual components of the system, the functionalities, and the mechanism of assessing the individual agents are briefly described. The main component, the supervisor agent, uses as a strategy a consensus method to reduce the level of investment risk. This method allows the coordination of the work of agents, and on the basis of decisions provided by the agents, and presents trading advice to the investor. The strategy testing has been done on FOREX quotes, namely on the pair EUR/USD. The results of the research are described and the directions of the further development of the platform are provided in the conclusion.**

I. Introduction

GENERALLY, the algorithms used in stock trading decision support systems can be based on mathematics, statistics, economics, or artificial intelligence [1;2;3;4;5;8]. Investing in financial instruments is always related to the occurrence of risk as the uncertainty of the future performance of investments. This uncertainty occurs due to links between the functioning of the capital market and factors such as the economic policy of the Government, the level of interest rates, exchange rates, or phases of the business cycle

A very important element of risk management is to measure these risks, to estimate the level of risk that is being taken in relation to the size of the capital which is at the disposal of the investor, as well as the investment limits. In general, the risk measures can be divided into three basic categories [8] :

- volatility measures (i.e. average deviation, average coefficient of variation),
- sensitivity measures (i.e. beta coefficient, delta coefficient),
- downside measures (i.e. Value at Risk).

In order to reduce the risk, diversification is applied. That is, investing in different types of instruments as well as various instruments of the same type. The diversity of investment reduces the risk of the instrument with the greatest level of risk, however, but on other site also lowers the expected investment rate of return . Another technique to reduce the level of risk is to take investment decisions with the use of multiple methods at the same time.

The aim of this paper is to present the trading strategy in a multi-agent system which avoids risky investment decisions due to the integration and cooperation of the agents. In the design of our system, called A-Trader, the accuracy of predictions, the orientation on online trading, the improvement of the financial knowledge base, and the ability to adapt to the changing market environment were all important requirements.

The paper is divided into three main sections. The first one presents the functional architecture of the system. The individual components of the system and the manner of communication between them are discussed. In the second, the consensus strategy used by the Supervisor agent to reduce the level of investment risk is described. The last part is a description of the results of the Supervisor strategy testing and performance analysis on the FOREX quotes. In conclusion, the further development direction of the A-Trader system are also described.

II. Multi-agent system – architecture and functions

The key ideas of A-Trader have been already detailed in our previous papers [9;10]. As a brief reminder: the A-Trader architecture in fig.1 sketches the main agents and components, namely: Notification Agent (NA), Historical Data Agent (HDA), Cloud of Computing Agents (CCA), Market Communication Agent (MCA), User Communication Agent (UCA), Database System (DS), Supervisor (S).

Let us describe briefly each agent of the system. The Notification Agent (NA) ensures efficient communication within the system. The Notification Agent forwards the information on the status change of a given agent to all agents that are recorded in the Notification register as the clients/observers of its signals. The notification is performed by triggering an appropriate Web method (SOAP) in all the agents from the list of listeners to the indicated signal. Next, it records the information on the status change of the

978-83-60810-53-8/$25.00

Notification Agent in the database. This capability of the Notification Agent makes the system flexible and scalable, provides the possibility to add and remove agents easily, and ensures the independence of the system from the physical position of the agent. Notification Agent sends the Base Agents and Intelligent Agents signals (decisions) to the Supervisor Agent. On the basis of these decisions, the Supervisor strategies are realized.

Another system agent downloads financial data from the Database System (SD) and delivers them to the agents according to their needs. This is the role of the Historical Data Agent (HDA).

The next system component is the Cloud of Computing Agents (CCA), consisting of Basic Agents Cloud (BAC), Intelligent Agents Cloud (IAC), and User Agents Cloud (UAC).

The Basic Agents Cloud (BAC) is a group of agents which pre-process data and compute the basic technical analysis indicators. The agents which possess their own knowledge base, which can learn and change their parameters and their internal state, create another component of the agents cloud, called the Intelligent Agents Cloud (IAC). This group of agents includes all the solutions based on artificial intelligence (genetic algorithms, neural networks, expert systems, etc.), agents analyzing text messages, agents observing user behaviour. The result of the operation of Basic Agents and the Intelligent Agents are the decisions which are transferred to the Supervisor Agent.

The User Agents Cloud (UAC), in turn, is the cloud containing the agents created by external users. Separating this part of the system provides the possibility of integrating the User Agents with the system without the necessity of installing the agent on the servers.

The communication of the system with the external environment is ensured by the Market Communication Agents (MCA). On the one hand, these agents supply the news from financial markets and quotations of the available securities. On the other hand, they are responsible for performing open and close position orders.

Fast and easy visualisation of the results of the agents is an important aspect in verifying the correctness of its operation. Such visualisation is possible in the system due to the User Communication Agents (UCA). The Communication Agent allows the user to communicate its own recommendations to the Intelligent Agents. It enables the change of the parameters of a selected agent or the suggestion for the Supervisor on which mechanisms are supposed to influence investment decisions, and to what extent.

The key component of the system is the Supervisor (S). The main goal of this agent is to generate profitable trading advice that reduces the investment risk. The supervisor, by using different strategies, coordinates the computing on the basis of decisions generated by Basic and Intelligent agents, and provides the final decision to the trader. Fig. 2 presents the general functional schema of A-Trader. A few strategies were developed in the system such as the consensus strategy, the rule-based strategy, and the evolution-based strategy.

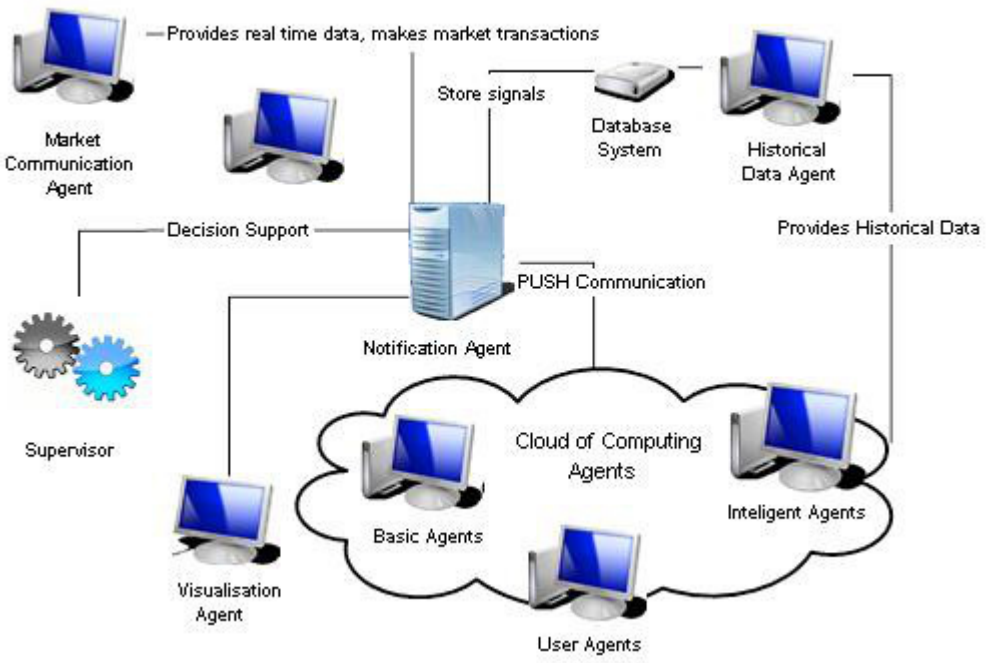


Fig.1 A-Trader system architecture
Source: Own work.

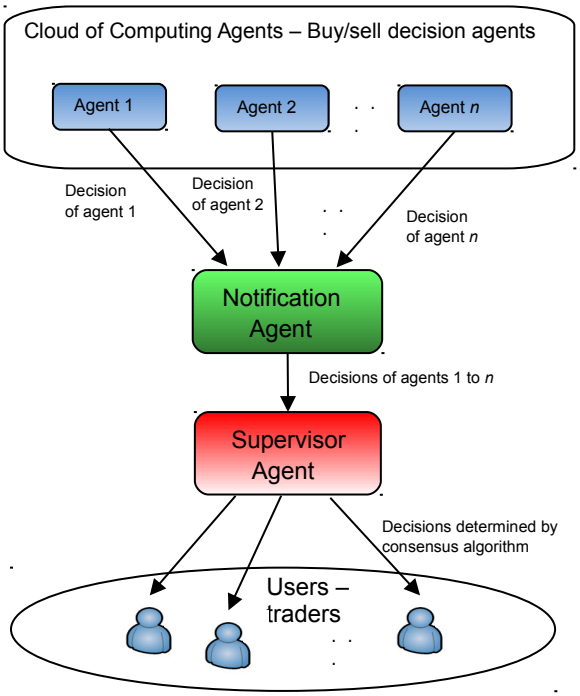


Fig.2 Functional schema of A-Trader
Source: Own work.

The strategies operate on the following assumptions:

1. Cloud of computing agents – Buy/sell decision agents, the intelligent programs, which, on the basis of the signals received from the Notification Agent, take the specified decision on buy/sale. Each agent has been implemented with a different method of computation and decision making. Buy/sell decision agents send the decisions to Notification Agent.
2. The Supervisor Agent – functions on the basis of the strategies that allow the determination of the final decisions on the basis of separate decisions generated by individual agents (read from Notification Agent), which are to be presented to the users (Traders). As a consequence, it is possible to reduce the level of risk associated with investing in a financial instrument.
3. Users – mostly traders who invest on financial markets, or bots (automatic traders).

The multi-agent system is composed of the agents being capable of generating independent decisions. These may be mutually consistent or completely contradictory decisions. Such mutually exclusive decisions are e.g. the open and close positions generated by two independent agents at the same time. The conflicts are resolved by the Supervisor, which observes the decisions of all the Cloud of Computing Agents and the Intelligent Agents, and assesses their effectiveness in investing and risk. The Supervisor determines which agents are taken into consideration when making an investment decision and whose advice is ignored based on the collected information.

The Supervisor may apply various strategies to generate the final trading decision. In the paper, the consensus method is detailed and tested. The relevant literature [13] defines consensus as an agreement and originates from choice theory. Consensus is based on the existing solutions to a given problem, is very close to them, but does not have to be one of these solutions. Hence, the financial decision presented to the user is a decision formed on the basis of the generated trading decisions [11].

The consensus is elaborated in three major stages. In the first stage it is necessary to carefully examine the structure of the set of financial decisions. In the second stage it is necessary to define the distance functions among particular decisions. The third stage is an elaboration of consensus algorithms that generate a decision, that the distance between this decision (consensus), and the individual decisions is minimal (according different criteria) [12].

The specificity of the problem being solved (a large set of open and close position signals) due to the dynamically changing market environment, the implementation requires an extremely high performance of the system. Therefore, from a software point of view, to make the implementation easier, each agent is activated within its container, which isolates it from the environment and which encapsulates the communication with the Notification Agent. Note that multiple containers may be activated on one machine.

The objective of the isolated agents and their transfer to the cloud is to ensure asynchronous cooperation and to enable the performance of specialised operations on the dedicated environment. For example, computing algorithms may be performed synchronously on the computers equipped with multi-processor NVIDIA graphic cards.

III. `Supervisor - Consensus Strategy

The consensus method was implemented as one of the main strategies of the Supervisor. The consensus algorithm runs automatically after providing the decision advices by individual agents.

Each financial decision must be represented by using a concrete structure (the first stage of consensus determining). Such structure was defined in work [6]. In our system, the financial decision consists in a trading position relating to a given quote, such as EUR/USD, USD/GBP, etc. The formal definition of this structure is the following:

Definition 1.

Decision P about finite set of financial instruments $E=\{e_1, e_2, \ldots, e_N\}$ is defined as a set:

$$P=\langle \{EW^+\}, \{EW^\pm\}, \{EW^-\}, Z, SP, DT \rangle \quad (1)$$

where:

1) $EW^\pm=\langle e_o, pe_o\rangle, \langle e_q, pe_q\rangle, \ldots, \langle e_p, pe_p\rangle$

Couple $\langle e_x, pe_x \rangle$, where: $e_x \in E$ and $pe_x \in [0,1]$ denote a financial instrument and this instrument's participation in set EW^+.

Financial instrument $e_x \in EW^+$ is denoted by e_x^+.

The set EW^+ is called a positive set; in other words, it is a set of financial instruments about which the agent knows the decisions to buy, and the volume of this buying.

2) $EW^{\pm} = \langle e_r, pe_r \rangle, \langle e_s, pe_s \rangle, \ldots, \langle e_t, pe_t \rangle$.

Couple $\langle e_x, pe_x \rangle$, where: $e_x \in E$ and $pe_x \in [0,1]$ denote a financial instrument and this instrument's participation in set $EW^{\pm}$.

Financial instrument $e_x \in EW^{\pm}$ will be denoted by $e_x^{\pm}$.

The set $EW^{\pm}$ is called a neutral set, in other words, it is a set of financial instruments, about which the agent does not know that buy or sell. If these instruments are held by an investor, that they should not be sold, or if they are not in possession of the investor, should not be bought by them.

3) $EW^- = \langle e_u, pe_u \rangle, \langle e_v, pe_v \rangle, \ldots, \langle e_w, pe_w \rangle$.

Couple $\langle e_x, pe_x \rangle$, where: $e_x \in E$ and $pe_x \in [0,1]$, denote financial instrument and this instrument's participation in set EW^-.

Financial instrument $e_x \in EW^-$ will be denoted by e_x^-.

The set EW^- is called a negative set; in other words it is a set of financial instruments of which the agent knows that these elements should sell.

4) $Z \in [0,1]$ - predicted rate of return.

5) $SP \in [0,1]$ - degree of certainty of rate Z. It can be calculated on the basis of the level of risk related with the decision.

6) DT - date of decision.

A situation in which the structures of a decision in the system differ, or the values of their attributes are different, is called a knowledge conflict of these agents. This conflict results in the taking by agents of various, often contradictory decisions concerning buying and selling a financial instrument.

Consensus is determined on the basis of a decision generated by different agents working in a system. We call a set of such decisions a profile and define it as follows [7]:

Definition 2.

E set of financial instruments $E = \{e_1, e_2, \ldots, e_N\}$ is given. In the case of A-Trader it is a set of pairs of currencies, e.g. EUR/USD, USD/GBP.

A profile $A = \{A^{(1)}, A^{(2)}, \ldots, A^{(M)}\}$ is called a set of M decisions of finite set of financial instrument E, such that:

$$A^{(1)} = \langle \{EW^+\}^{(1)}, \{EW^{\pm}\}^{(1)}, \{EW^-\}^{(1)}, Z^{(1)}, SP^{(1)}, DT^{(1)} \rangle$$
$$A^{(2)} = \langle \{EW^+\}^{(2)}, \{EW^{\pm}\}^{(2)}, \{EW^-\}^{(2)}, Z^{(2)}, SP^{(2)}, DT^{(2)} \rangle$$
$$\ldots\ldots\ldots\ldots$$
$$A^{(M)} = \langle \{EW^+\}^{(M)}, \{EW^{\pm}\}^{(M)}, \{EW^-\}^{(M)}, Z^{(M)}, SP^{(M)}, DT^{(M)} \rangle \quad (2)$$

In the case of A-Trader the profile is a set of decisions generated by Base Agents and Intelligent Agents. On the basis of these decisions, the Supervisor strategy is executed.

At the A-Trader system, the values Z, SP are provided by Base Agents and Intelligent Agents (e.g. by using statistical forecasting methods, or artificial intelligence methods) or by Supervisor (e.g. on the basis of agent performance evaluation). The values DT are generated, by all agents, together with the signals (decisions).

The Supervisor strategy is carried out according to the following consensus algorithm:

Algorithm 1.

Data: The profile A= $\{A^{(1)}, A^{(2)}, \ldots A^{(M)}\}$ consists of M agents' decisions.

Result: Consensus

$CON = \langle CON_+, CON_{\pm}, CON_-, CON_Z, CON_{SP}, CON_{DT} \rangle$ according to A. The consensus is a decision generated by the Supervisor Agent. This decision consists of the same attributes as the decision of the agents (e.g. CON_+ mean consensus of the EW^+ set), but the values of these attributes differ.

Begin

1: $CON_+ = CON_{\pm} = CON_- = \emptyset, CON_Z = CON_{SP} = CON_{DT} = 0$
2: $j := 1$.
3: $i := +$.
4: If $t_i(j) > M/2$ then $CON_i := CON_i \cup \{e_j\}$
Go to:6.
5: If i=+ then $i := \pm$
If $i = \pm$ then i:=-
If i=-, then Go to:6
Go to:4.
6: If $j < N$ then $j := j+1$ Go to:3
If $j \geq N$ then Go to:7.
7: $i := Z$.
8: Determine $pr(i)$.
9: $k_i^1 = (M + 1) / 2$, $k_i^2 = (M + 2) / 2$.
10: $k_i^1 \leq CON_i \leq k_i^2$.
11: If $i = Z$ then $i := SP$
If $i := SP$ then $i := DT$
If $i := DT$ then End
Go to: 8.

End

The computational complexity of this algorithm is O(3NM).

The presented algorithm of consensus proposes a decision to the trader, who does not need to think about the choice of decision generated by Basic Agents and Intelligent Agents, which significantly reduces the time it takes to make a decision. Since a decision is taken on the basis of multiple

agents' decisions, it also reduces the risk of taking this decision, because it eliminates the possibility to make an incorrect decision by one of the agents a-Trader system agents.

The verification of the Supervisor strategy is presented in the next section of the article.

IV. EXPERIMENTS

FOREX (Foreign Exchange Market) is the market where one currency is traded for another. It is one of the largest markets in the world. Currencies are traded against one another in pairs. For instance, the quotation EUR/USD (EUR/USD) 1.3465 is the price of the euro expressed in US dollars, meaning 1 euro = 1.3465 dollars. To evaluate the Supervisor performance the pair EUR/USD is chosen from the FOREX market. In the evaluation the following assumptions have been imposed:

1. The minute-by-minute quotations EUR/USD are randomly selected, covering the following periods:
   I. 17-04-2013 hours: 12:47 to 15:00,
   II. 30-04-2013 hours: 18:36 to 23:16,
   III. 08-05-2013 hours: 21:45 to 23:34,
   IV. 09-05-2013 hours: 00:00 to 2:50.
   For instance, Fig. 3 presents a quotation of the pair EUR/USD in period IV.
2. During verification, the Supervisor uses the decisions (signals buy-value: 1, sell-value: -1, remain unchanged-value: 0) generated by program agents, which operate on the basis of a combination of technical analysis indicators (i.e.. agent no. 1 taking decisions on the basis of RSI, Stochastic Oscillator, MACD indicators combination, agent no. 2 – CCI, WILLIAMS, OBV, etc.). Due to the computational complexity and time constraints, the experiment was illustrated in the article by the Supervisor's signals generated by 6 agents.
3. Final Buy-Sell decisions are taken on the basis of the Supervisor's signals (fig. 4).
4. It is assumed that the initial trader capital equals 1000 USD, and that the investment rate of return shall be calculated as the difference between the initial capital and the amount that the investor will have after the last sales in a given period. The rate of return is expressed in (USD).
5. No transaction costs are taken into consideration.
6. Money management – assume that in each transaction, the investor commits 100% of capital. Money management strategy can be set by the user. The investor invests every time 1000 USD - leverage 10:1.
7. Performance evaluation is based on following ratios:
   a) the number of transactions,
   b) gross profit,
   c) gross loss,
   d) total profit,
   e) the number of profitable transactions,
   f) the number of profitable transactions in a row,
   g) the number of unprofitable transactions in a row,
   h) the average coefficient of variation is the ratio of the average deviation of the arithmetic average multiplied by 100% and is expressed:

$$V = \frac{s}{|E(r)|} \cdot 100\,\% \ . \tag{3}$$

   where:
   V – average coefficient of variation,
   s – average deviation of the rate of return,
   E(r) – arithmetic average of the rate of return.

Fig.3 EUR/USD quotations
Source: Own work.

Fig.4 Decisions of Supervisor, Buy-and-Hold and EMA in the period IV
Source: Own work.

i) Value at Risk – the measure known as value exposed to the risk - that is the maximum loss of the market value of the financial instrument possible to bear in a specific timeframe and at a given confidence level [2].

$$VaR = P*O*k \quad (4)$$

where:
P – the initial capital,
O – volatility - standard deviation of rates of return during the period ,
k –the inverse of the standard normal cumulative distribution (assumed confidence level 95%, the value of k is 1,65).

8. The results obtained by the Supervisor have been compared with the passive strategy Buy-and-Hold and the benchmark using EMA.

The test was carried out in the following way:

1. On the basis of the quotation from the first period, each agent referred to when to buy and when to sell a currency EUR/USD.
2. Next, taking into consideration the decisions of all the agents, the consensus was determined.
3. The performance of the Supervisor and benchmarks Buy-and-Hold and EMA are reported.
4. Next, the steps 1 to 4 were repeated using the next periods of the financial time series.
5. In the final stage, the performance ratio values were calculated corresponding to rates of return resulting from all decisions generated by the Supervisor, Buy-and-Hold and EMA strategies (not only of the final rates of return, but with all the rates of return calculated after each sale decision).

Comparison of final capital and rates of return obtained are shown in table 1.

TABLE 1. COMPARISON OF FINAL CAPITAL AND RATES OF RETURN

	Consensus		**B & H**		**EMA**	
period	Rate of return [USD]	Rate of return [%]	Rate of return [USD]	Rate of return [%]	Rate of return [USD]	Rate of return [%]
I.	10,59	0,011	-19,01	-0,019	-29,64	-0,030
II.	19,09	0,019	11,10	0,011	7,68	0,008
III.	4,56	0,005	3,50	0,004	4,71	0,005
IV.	6,84	0,007	-0,23	-0,0002	4,26	0,004
average	10,27	0,010	-1,16	-0,0001	-3,25	-0,0003

Source: Own work.

Summing up the results obtained through the use of the consensus method, in each period, a higher rate of return is shown compared with the decisions generated by the Buy-and-Hold and EMA. It should also be noted that the average rate of return of the Supervisor's decision is positive (profit), while the average rate of return of the Buy-and-Hold and EMA is a negative value (loss).

The performance analysis (table 2) shows that the Supervisor generated a smaller number of transactions than using the EMA, but with the EMA, however, the gross profit from these transactions is higher than the gross profit generated by EMA and Buy-and-Hold.

At the same time, the gross loss generated by the Supervisor is relatively lower in comparison with the benchmarks. It should also be noted that the Supervisor conducted a 93,33% profitable transactions (Buy-and-Hold 50%, EMA 47,06%). Important is also the fact that the Supervisor does not generate a series of unprofitable transactions in a row, but for instance the EMA generated such transactions. Analyzing the risk of decisions, it can be

TABLE 2.
RESULTS OF PERFORMANCE ANALYSIS

Performance ratio	Supervisor - Consensus	B & H	EMA
Number of transactions,	15	4	34
Gross profit	41,54 USD	14,60 USD	27,60 USD
Gross loss	-0,46 USD	-19,24 USD	-40,59 USD
Total profit	41,08 USD	-4,64 USD	12,99 USD
Number of profitable transactions (%)	14 (93,33%)	2 (50%)	16 (47,06%)
Number of profitable transactions in row	8	1	6
Number of unprofitable transactions in row	1	1	6
Average coefficient of variation	6,29%	7,95%	10,51%
Value at Risk	8,26 USD	18,30USD	18,53 USD

Source: Own work

noticed that the use of consensus methods by the Supervisor allows the lowest level of risk investment. The value of Average coefficient of variation equals 6.29%, while for Buy-and-Hold equals 7,95%, and for EMA 10,51%, instead. The Value at Risk of decisions generated by the Supervisor was 8,26 USD, which means that using the consensus method the trader can lose up to 8,26 USD in about a 2 hours period. Regarding Buy-and-Hold and EMA, this value was appropriately 18,30 and 18,53.

The verification of using the consensus method by the Supervisor agent therefore suggested that the decisions supported by the A-Trader system are the decisions which allow the investor to get satisfactory investor's results. It should be noted, of course, that the consensus method will not necessarily always get the highest rate of return. However, it can be assumed that, as a general rule, it allows the investor to obtain a lower level of risk associated with the investment. Note that if an investor had to make the choice which agent has to "listen to", then, assuming that the probability of selection of the agents by the investor is the same, he could more often choose a decision (hint) of an agent that allows one to get a lower rate of return. Besides, the evaluated agents using simple indicators are characterised by the large disparity in rates of return, confirming, for example, the value of the average coefficient of variation of EMA or Buy-and-Hold.

In conclusion, we can say that financial decisions generated by the consensus method allow to get a higher rate of return in comparison to benchmarks such as Buy-and-Hold and EMA, and get a faster determination of the decision, than if the investor takes the decision himself, among the decisions generated by the agents. Currently, due to the turbulent economic environment, investing in financial instruments must be carried out in close to real time. First and foremost, however, the use of consensus algorithms by the Supervisor allows the investor to decrease the level of risk related to financial instrument investing. Therefore, it also increases the level of usefulness of the decisions, and this can bring the user satisfying benefits.

It should be noted that the agents that were used in this experiment applied only to technical analysis indicators. It should be stressed that the A-Trader system gives a possibility of implementing the agents using fundamental analysis, or behavioral analysis. Work on the extension and variety of the agents' knowledge is in progress.

V. CONCLUSION

The first attempts to implement a multi-agent environment proved encouraging. The Supervisor decreased the investment risk by restricting the independent operations of more risk-taking agents for joint decisions of the entire environment. The cooperating agents made profitable decisions more frequently and close the loss-generating positions considerably earlier.

It should be stressed that the goal of multi-agent financial decision support systems, also the A-Trader system, is not only to maximize the rate of return on investment, but also to limit the level of risk associated with this investment. Taking into account the EUR/USD quotation dealt with in the article, it can be concluded that the level of risk is associated with, among other things, the fact that the financial situation of the euro depends on the economic and political situation in many countries. Whereas the dollar depends on a variety of government regulations and the United States engagement in the world economy.

Of the experiment in the article, it can be concluded that the use by the Supervisor of the consensus method makes it possible to lower the level of investment risk in consequence.

This can lead the investor to achieve a satisfactory investment rate of return.

The platform allows the implementation of different, intelligent or behavioral Supervisor strategies. The described multi-agent system makes testing and validating these new algorithms easier by supplying the basic functionalities and data. It enables the concentration of work on constructing new Supervisor strategies without being concerned about the basic data and message supply mechanisms

The A-Trader platform is now in the testing and expansion phase. The number and scope of the applied methods is being continuously expanded. New agents based on the recent methods are created. Of course, to obtain more objective conclusions about the consensus strategy, the tests should be done on a longer periods, with other quotes.

REFERENCES

[1] Barbosa R.P., Belo O., "Multi-Agent Forex Trading System", in: Agent and Multi-agent Technology for Internet and Enterprise Systems, Studies in Computational Intelligence Volume 289, 2010, pp 91-118.

[2] Chan L., WK WONG A., "Automated Trading with Genetic-Algorithm Neural-Network Risk Cybernetics: An Application on FX Markets", Finamatrix, January 2011, pp.1-28.

[3] Karjalainen, R., "Using Genetic Algorithms to Find Technical Trading Rules", *Journ. of Financial Econ*., 51, 1999, pp.245-271.

[4] LeBaron B., "Active and Passive Learning in Agent-based Financial Markets", *Eastern Economic Journal*, vol 37, 2011, pp. 35-43.

[5] Dempster M., Jones, C., "A Real Time Adaptive Trading System using Genetic Programming", *Quantitative Finance*, 1, 2001, pp. 397-413.

[6] Hernes M*., Metody consensusu w rozwiązywaniu konfliktów wiedzy w wieloagentowym systemie wspomagania decyzji*, Praca dokt., Uniwersytet Ekonomiczny we Wrocławiu, 2011.

[7] Hernes M., Nguyen N.T., "Deriving Consensus for Hierarchical Incomplete Ordered Partitions and Coverings", *Journal of Universal Computer Science* 13(2) /2007, pp. 317-328.

[8] Jajuga K, Jajuga T., *Inwestycje: Instrumenty finansowe, ryzyko finansowe, inżynieria finansowa*, PWN, Warszawa 2000.

[9] Korczak J., Lipinski P., „Systemy agentowe we wspomaganiu decyzji na rynku papierów wartościowych", in:, Rozwój informatycznych systemów wieloagentowych w środowiskach społeczno-gospodarczych, ed. S. Stanek et al., Placet, 2008, pp. 289-301.

[10] Korczak J., Bac M., Drelczuk K., Fafuła A., "A-Trader - Consulting Agent Platform for Stock Exchange Gamblers", in: Proc. FedCSIS, Wrocław, 2012, pp.963-968.

[11] Nguyen N.T., "Using Consensus Methodology in Processing Inconsistency of Knowledge", in: Last M. et al. (Eds): *Advances in Web Intelligence and Data Mining, series Studies in Computational Intelligence*, Springer-Verlag, 2006, pp. 161-170.

[12] Sobieska-Karpińska J., Hernes M.," Consensus determining algorithm in multiagent decision support system with taking into consideration improving agent's knowledge", *Proceedings of the Federated Conference on Computer Science and Information Systems* (FedCSIS), 2012, pp. 1035-1040.

[13] Zgrzywa M., "Consensus Determining with Dependencies of Attributes with Interval Values", *Journal of Universal Computer Science*, vol. 13, no. 2, 2007, pp. 329-344.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1127–1134

Optimising Web-Based Information Retrieval Methods for Horizon Scanning Using Relevance Feedback

Marco A. Palomino and Tim Taylor
University of Exeter
Medical School
Truro, UK
[m.palomino, tjt205]@exeter.ac.uk

Geoff McBride and Hugh Mortimer
STFC Futures Programme
Rutherford Appleton Laboratory
Didcot, UK
[geoff.mcbride, hugh.mortimer]@stfc.ac.uk

Richard Owen
University of Exeter
Business School
Exeter, UK
r.j.owen@exeter.ac.uk

Michael Depledge
University of Exeter
Medical School
Truro, UK
m.depledge@exeter.ac.uk

Abstract—**Horizon scanning is being increasingly regarded as an instrument to support strategic decision making. It requires the systematic examination of information to identify potential threats, emerging issues and opportunities to improve resilience and decrease risk exposure. Horizon scanning can use the Web to augment the acquisition of information, though this involves a search for novel and emerging issues without knowing them beforehand. To optimise such a search, we propose the use of relevance feedback, which involves human interaction in the retrieval process so as to improve results. As a proof-of-concept demonstration, we have carried out a horizon scanning exercise which showed that our implementation of relevance feedback was able to maintain the retrieval of relevant documents constant over the length of the experiment, without any reduction. This represents an improvement over previous studies where relevance feedback was not considered.**

I Introduction

The use of the World Wide Web for futures research has been gaining increasing attention [1-3]. Largely, the aim of futures research is to anticipate and prepare for new and changing risks, and to consider the implications that emerging issues will have on the distribution of resources and existing priorities. Given the current environment of change and uncertainty, both public and private sectors have identified the need to strengthen futures research and integrate it into strategic thinking and planning.

In the UK, the importance of futures research has been highlighted by a series of perceived failures in science and policy, such as the failure to recognise the concerns of the public about genetically modified crops until they emerged in the media, and the inadequate reaction to the outbreak of the foot/hoof and mouth disease in 2001 [4]. As a consequence of these setbacks, the UK Government has emphasised its use of horizon scanning, *"the systematic examination of information to identify potential threats, risks, emerging issues and opportunities, beyond the Parliamentary term, allowing for better preparedness and the incorporation of mitigation and exploitation into the policy making process"* [5]. Explicit objectives of horizon scanning are to anticipate issues, accumulate reliable data and knowledge about those issues and thus inform policy making and implementation [6].

Data collection associated with horizon scanning has blossomed with the availability of electronic databases and Web search engines. Regrettably, the process of searching for potential threats and emerging issues is not transparent. While searching is a retrieval process where the searcher knows in advance what she is looking for, horizon scanning is a process where we are trying to discover what is novel and surfacing without knowing it ahead of time. As explained by Palomino et al [7], we have access to ''search engines'' on the Web, but not to ''scanning engines''.

The impossibility of establishing precisely what is being sought before beginning the search makes it difficult to formulate information queries that are well designed for horizon scanning purposes. This suggests that the first retrieval operation involved in the process of scanning the horizon should be conducted with a tentative, initial query, and should be treated as a trial only, designed to locate a few useful items, which could then be examined for relevance so that later on new and improved query formulations can be constructed with the expectation of retrieving additional useful items in subsequent search operations. This is the reason why we have decided to explore the use of a controlled, automatic process for query reformulation, namely, relevance feedback, a technique utilised by some information retrieval systems [8].

The aim of this paper is to assess the use of relevance feedback as part of a horizon scanning system. To this extent, the remainder of this paper is organised as follows: Section II reviews related work on relevance feedback and briefly outlines previous research on Web-based horizon scanning. Section III details our implementation of relevance feedback in the context of a horizon scanning prototype which we are employing as a proof-of-concept demonstration. Section IV discusses a horizon scanning exercise that was conducted for a European Union Framework 7 project in association with RAL Space [9]—a world-class space research centre—to review current and future technologies for detecting and monitoring diseases in vegetation. We used this exercise as a case study to test our implementation of relevance feedback. Section V reports on the evaluation of the results of RAL Space's exercise, and, finally, Section VI states our conclusions.

978-83-60810-53-8/$25.00

II RELATED WORK

Relevance feedback has been extensively studied since its development in the mid-1960s [8, 10-12]. It refers to an interactive process that helps to improve retrieval performance: when a user submits a query, an information retrieval system would first return an original set of documents that satisfy the query and then ask the user to judge whether these documents are relevant or not; after that, the system would reformulate the query based on the user's judgments, and return a new set of documents. To some extent, relevance feedback is an alternative to save users from articulating queries in a trial-and-error manner.

Most of the research on relevance feedback undertaken thus far has approached its implementation as a supervised learning problem [8, 10, 11], where the key is to optimally balance the original query and the feedback information [13]—a special track to look into the effects of different factors on the success of relevance feedback has been organised by the Text Retrieval Conference (TREC) [14]. However, the use of relevance feedback in the context of horizon scanning has not been investigated yet. References to the applications of horizon scanning and the results of specific scans keep growing [4, 6, 15-21], as the interest in the subject increases, but only a few academic papers describe the methodology to carry out an automated scan [7, 22, 23], and the combined use of horizon scanning and relevance feedback has not been documented until now.

Shaping Tomorrow [24] and Recorded Future [25] are two private firms using Web-based scanning tools. Shaping Tomorrow helps organisations make better decisions through anticipating and preparing for the future. It uses a variety of manual, semi-manual and automated scanning processes to track and share information from around the world. It is first supported by a virtual network of volunteer and client researchers who "scan the scanners"—experts in the field—for material. Shaping Tomorrow also employs its own purpose-built Web-robot to scrape high value future websites and its service has accumulated 100,000 scan hits on emerging change, gathered over ten years from 5,000 plus sources, and 3,600 issues—trends, uncertainties and surprises—evidenced and linked to the scan hits. Shaping Tomorrow will soon release software to read the scan hit and do almost all of the researchers work automatically [26].

Recorded Future is established on the premise that all the information available on the Web is useful to support forecasting methods, financial or otherwise. Recorded Future continuously harvests news from more than 40,000 online sources, ranging from media and government websites to individual blogs and selected twitter streams [25]. Recorded Future aims to create and maintain a database of facts—pairs of timed entities and event instances—to track trends and historical developments and predict future events. As opposed to Recorded Future, we are not interested in predicting the future, but rather in improving resilience and the capability to react to new risks and opportunities.

In the public sector, horizon scanning has proved useful to identify new and emerging health technologies [20, 27, 28]. However, due to the large amount of information published online, it is difficult to recognise valuable data [29]. In an attempt to establish how exactly the Web should be used in health technology assessments, Douw et al [27] circulated a questionnaire among organisations known to use the Web for horizon scanning purposes. The questionnaire focussed on the type of websites scanned, the frequency of the scanning, and the importance of the Web for the identification of new health technologies. Responses to the questionnaire indicated that the organisations surveyed found new information through word of mouth, and links found on websites that they monitor continuously. Even though this highlights the importance of personal networking in horizon scanning, and the expertise of the scanners to choose the best links to follow, our work is directed towards the automation of the human-intensive practice of detecting and summarising emerging information. Hence, rather than surveying organisations, we have concentrated on the methodology to carry out a Web-based scan of the horizon.

Our methodology for Web-based horizon scanning comprises several interlinked components, as described by Palomino et al [7]: emerging information is retrieved—manually or otherwise—and / or received—e.g., via selected RSS feeds—from a variety of Web-based sources—such as, online scientific, peer-reviewed literature and news websites, which were sources of high importance for the work with RAL Space that we will describe in Section IV. Key parts of the retrieved information may be extracted and later on categorised. Afterwards, the information is often archived in a database. Periodically, outputs are presented to decision makers or used to write up reports or newsletters.

III RELEVANCE FEEDBACK

The main idea behind our implementation of relevance feedback consists of choosing important keywords attached to certain previously retrieved documents that have been characterised as relevant by the users, and of enhancing the importance of those keywords in future queries. Correspondingly, keywords included in previously retrieved non-relevant documents could be deemphasised in any future query formulation. Ideally, the effect of this query alteration process is to "steer" the query in the direction of the relevant documents and away from the non-relevant ones, with the expectation of retrieving more useful and fewer non-useful documents in later steps of the search.

Figure 1 shows a general Web-based horizon scanning approach for strategic decision support that uses relevance feedback. It accentuates the importance of the continuous scanning, noting that the processes of retrieving documents, and analysing, categorising and archiving information are iterated as part of a continuous process—static or sporadic scans become outdated quickly. The outputs of horizon scanning can be interfaced with further tools for opportunity and risk analysis [14] and scenario development.

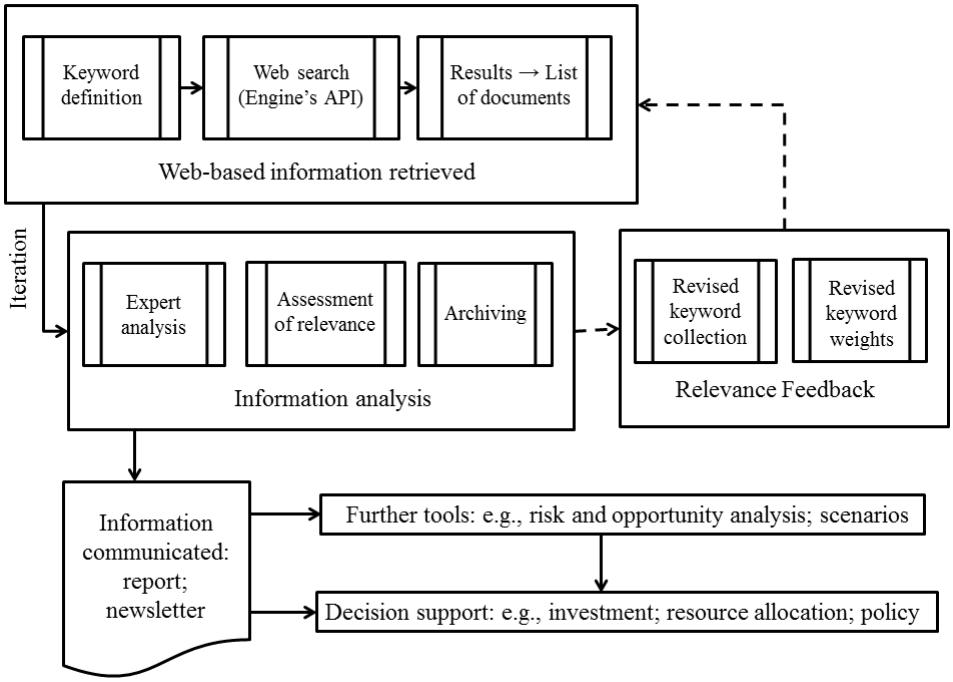


Figure 1. A generalised approach to Web-based horizon scanning for decision support using relevance feedback—based on Palomino et al [7]

Relevance feedback offers the following advantages to the analysts in charge of scanning the horizon:

(i) It frees the analysts from the details of the query formulation process—especially in late stages of the search.
(ii) It splits the search into an organised sequence of steps to reach the desired information gradually.
(iii) It devises useful queries without former analysis of the availability of data on the Web.
(iv) It features a controlled query alteration process designed to emphasise some keywords and deemphasise others, as required to accomplish a particular search.

Relevance feedback was originally developed as a technique to be used in conjunction with vector queries—i.e., queries represented by vectors with as many entries as keywords comprised in the query. Each entry refers to a "weight" symbolising the importance of the corresponding keyword within the query. For example, a particular query Q composed of n keywords may be written as

$$Q = (w_1, w_2, \ldots, w_n),$$

where w_i is the weight of the i-th keyword. Keyword weights are restricted to the range 0 to 1, where 0 means the corresponding keyword is absent from the query and 1 means it is so critical to the query that it has a full weight.

Given a vector such as Q, the relevance feedback process starts by generating a new vector

$$Q' = (w_1', w_2', \ldots, w_n'),$$

where w_i' represents a modified weight for the i-th keyword in the query—new keywords can be introduced to the query, and old keywords can be removed by reducing to 0 its weight. The process continues by creating yet another vector Q'' by modifying the weights of Q' according to new feedback, and so on and so forth until the required documents are found or the process reaches a pre-established number of iterations. Graphically, the relevance feedback process can be depicted as a relocation of the query vector from one place to another in the n-dimensional space defined by the n keywords under consideration.

A poorly conceived query reformulation can result in deterioration in retrieval performance [30]. Hence, a suitable set of keywords to search for information should be selected at each step in the process. We always choose our keywords with the support of software for the automatic extraction of keywords. Specifically, we use Yahoo!'s Content Analysis Web Service [31].

Normally, a scan of the horizon begins by defining the goals of the scan with a few sentences. We then submit those sentences to Yahoo!'s Content Analysis Web Service to automatically extract keywords—when available, entire documents relevant to the scan, called seed documents, are submitted to extract keywords.

These keywords are used to create the initial queries to search the Web for information. Normally, these keywords are combined with terms and phrases such as `new development`, `revolutionary`, `first time`, and others which have been suggested by the UK Defence Science and Technology Laboratory (Dstl) as descriptors of emerging issues [32]. These combinations of automatically extracted keywords and descriptors of emerging issues constitute the queries employed to bootstrap the relevance feedback process—i.e., these are the queries whose formulation we will attempt to refine along the process.

Once we have retrieved a first list of documents as a result of releasing our queries, we proceed to collect feedback. Usually, an expert, or a group of experts, in the field of the scan, or the same people who developed the requirements for the scan, are asked to indicate, for each document in our results, whether it is relevant, very relevant or non-relevant. The documents that are marked as very relevant are submitted to Yahoo!'s Content Analysis Web Service to extract new keywords. Keywords that were not considered in the initial queries, but are at the top of the new list of keywords yielded by Yahoo!'s Content Analysis Web Service are added to the original keywords and used to formulate new queries—keywords at the top of the list are expected to be more characteristic of the documents submitted than those near the bottom [31].

For each document that we retrieve, we keep a record of the keywords that were included in the queries used to retrieve it—note that a particular document can be retrieved as a result of more than one query and therefore be associated with several keywords. The weights of keywords used in queries that retrieved documents that were marked as very relevant are increased by a factor proportional to the number of very relevant documents associated with them. Likewise, the weights of keywords associated with documents marked as non-relevant is decreased by a factor proportional to the number of non-relevant documents associated with them—see Figure 2. The weights of keywords associated with documents marked as relevant—but not very relevant—is not modified and remains the same for the following iteration.

Once the set of keywords has been amended to integrate the initial feedback received, and the weight of each keyword has been adjusted to reflect the number of relevant, very-relevant and non-relevant documents retrieved with them, we proceed to release new queries, whose formulation can be thought of as a refinement of the initial ones, and the entire process can be repeated again until we complete a pre-established number of iterations. To automate our search for documents on the Web, we programmatically released our queries via Google's Custom Search API [33]. We chose Google's Custom Search API, because Google is the most popular search engine [34]; yet, other engines with an API interface could be used too—in other words, we will focus on Google for testing purposes, but the approach described here is not restricted to a specific search engine.

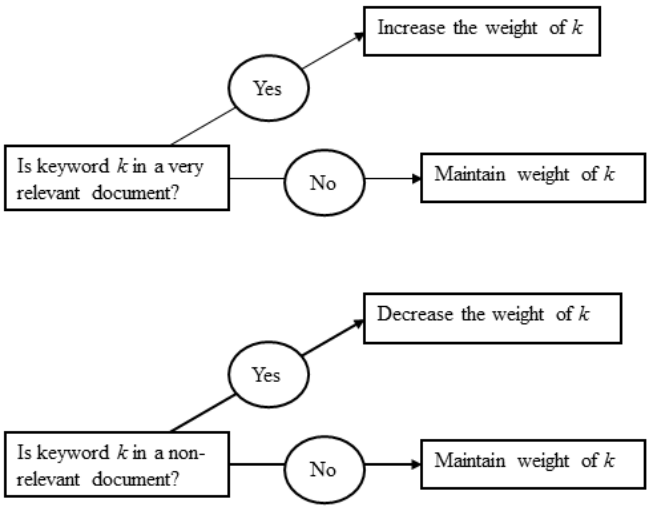


Figure 2. Keyword weight adjustment

Google has one of the largest databases of Web pages, including many types of documents—blog posts, wiki pages, group discussion threads—and document formats—PDF, Microsoft Word or PowerPoint documents, among many others. Despite the presence of all these types of documents and formats, Google's method of ranking on the basis of the PageRank citation algorithm [35] often places relevant documents near the top of the search results, and Google's Custom Search API allows us to query Google's repository directly and frequently in an automated way. Indeed, the frequency with which we query Google's repository can be adapted to the particular needs of the scan.

A. Queries with weighted keywords

A critical aspect of our relevance feedback implementation is the use of weights to express the importance keywords. Appropriately using those weights is what guarantees that our process reaches the desired information gradually; otherwise, the continuous extraction of keywords from newly retrieved documents would simply increase the number of keywords and queries, which would in turn increase the number of collected documents, without guaranteeing that we are actually gathering more useful information. Devising a way to adequately use the weights so that subsequent queries assign higher importance to keywords with greater weights is one of the most challenging features to accomplish.

Our implementation is based on using the weights of the keywords to decide how we should employ those keywords to look for documents:

(i) Keywords with low weights are used to search for documents that include the keywords anywhere in the text—not necessarily in prominent places.
(ii) Keywords with high weights are used to search for documents that include the keywords in their titles—according to Page et al [35], titles are more descriptive of the contents of a document than the rest of the text.

(iii) Keywords with very high weights are used to search for documents which are referenced to by hyperlinks whose text includes the keywords—Page et al [35] have stated that the text contained in the hyperlinks that point to a document, also known as the anchor text, link text, or link title, is greatly descriptive of the contents of the document referred to.

(iv) Keywords whose weights have been reduced to 0, which means that they have no relevance at all to the search, are preceded by the "minus" operator in our queries to explicitly indicate that they must not appear in the retrieved documents.

(v) All keywords have the same weight at the start, when the first search takes place and no feedback has been gathered yet. For the first iteration, all keywords are used to search for documents that contain them anywhere in the text.

(vi) Keywords that are meant to be descriptors of emerging issues—for instance, `ground breaking` and `closer to reality`—have constant weights that are never modified through the entire process. We always search for documents that contain these keywords anywhere in the text.

Our implementation of relevance feedback ensures that keywords with higher weights are looked for in places which are expected to have higher importance and therefore be more descriptive of the documents that contain them. Table I displays the association between weight ranges for keywords and the locations—hyperlinks, titles, or general text—where we search for those keywords to retrieve new documents that contain them.

TABLE I. KEYWORD RANGES AND KEYWORD LOCATIONS

Weight range	Keyword location
0	Nowhere in the document
(0,0.33]	Anywhere in the text
(0.33,0.66]	In the title
(0.66,1]	In the anchor text

In order to illustrate the relevance feedback process in detail, we will use an example. The example derives from a horizon scanning exercise proposed by RAL Space in October 2012 and it is explained in the following section.

IV RAL SPACE SCANNING EXERCISE

In October 2012, RAL Space, based at the Rutherford Appleton Laboratory (RAL), undertook a review for the European Union Framework 7 project Q Detect: Developing Quarantine Pest Detection Methods for use by National Plant Protection Organizations (NPPO) and Inspection Services [36]. In this review, RAL Space looked into current and future aerial platform technologies and instrumentation options for detecting and monitoring diseases in vegetation, and the mapping of pests through the use of aerial platforms.

The review aimed to assess the efficacy of remote sensing techniques—such as direct imaging and spectrally resolving reflected light—from different aerial platforms—ranging from small unmanned aircraft to low altitude satellites—to evaluate and monitor the health of plant life over long periods of time with little human inspection. The report was not meant to target specific plant diseases, but to provide an overview of various, if not all, potential diseases, whilst providing a thorough examination of the state-of-the-art in remote sensing instrumentation and platform technology.

As part of the review, RAL Space assessed how low, medium and high-altitude platforms integrated with high spectral and spatial resolution instrumentation could be used to come up with different performance metrics within a specific user requirement framework, which included cost, endurance, spatial resolution and frequency of measurement. RAL Space's review contributed to compare the economic benefit and practical realisation of present and forthcoming technology to assist in the detection of quarantined disease remotely. Since decision making on the uptake and use of emerging technology for disease monitoring has to be supported by timely and high quality information, RAL Space made use of horizon scanning to produce the review.

The horizon scanning exercise began by establishing the seed documents. These documents—listed in Table II—were mostly academic papers chosen by RAL Space.

TABLE II. SEED DOCUMENTS

Carter, G.A. & Knapp, A.K. 'Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration.' American Journal of Botany, 88: (2001)

Cloutis, E.A. 'Agricultural crop monitoring using airborne multi-spectral imagery and C-band synthetic aperture radar'. International Journal of Remote Sensing. Volume 20, Issue 4. (1999)

Coops, N.C., Goodwin, C., Stone, C. Sims, N. 'Assessment of forest plantation canopy condition from high spatial resolution digital imagery'. Canadian Journal of Remote Sensing. 32: (2006)

Lelong, CD, Burger, C., Jubelin, G. Roux, Labbé, S. & Baret, F. 'Assessment of Unmanned Aerial Vehicles Imagery for Quantitative Monitoring of Wheat Crop in Small Plots'. Sensors 8. (2008)

Moran, S.M. 'Thermal Infrared Measurement as an Indicator of Plant Ecosystem Health.' Journal remote sensing. (2003)

Rock, B., Vogelmann, J., Williams, D., Vogelmann, A., & Hoshizaki, T. 'Remote Detection of Forest Damage.' BioScience, 36. (1986)

Sharples, J.A. 'The Corn Blight Watch Experiment: Economic implications for use of remote sensing for collecting data on major crops'. LARS information note 110173.

The text of all the abstracts of the academic papers in Table II was submitted to Yahoo!'s Content Analysis Web Service, and a large list of keywords was produced in return. Together with an analyst from RAL Space, we chose the keywords that we considered most useful and grouped them into three different categories:

(i) Subject keywords, which refer to the main subject of RAL Space's review—for example, `crop monitoring` and `plant health`.
(ii) Technology keywords, which refer to different technological alternatives for detecting and monitoring diseases in vegetation—for example, `satellite` and `remote sensing`.
(iii) Descriptors of emerging issues, which are keywords defined by Dstl to capture "fresh" information on relevant subjects.

Table III shows the precise set of keywords that we use to start the process. Combinations of these keywords produced a total of 140 queries: each query included one, and only one, keyword from each category. Those 140 queries were used to start the search.

TABLE III. INITIAL SETS OF KEYWORDS

Subject	Technology	Emerging issues
crop disease	aerial platforms	breakthrough
crop monitoring	remote sensing	closer to reality
environmental monitor	satellite	first time
forest monitoring	unmanned aerial vehicle	ground breaking
plant health		new development
		novel
		revolutionary

Although we set up our prototype to limit to 64 the number of results per query, this still allowed up to 8,960 documents to be retrieved for each automatic release of the 140 queries employed in the initial search—indeed, nearly 4,000 unique documents, approximately, were retrieved per iteration. It would be unmanageable for a RAL Space analyst to review all those documents, given the short time allocated to this activity. Hence, we committed to deliver 50 documents, exclusively, per iteration to RAL Space, because this was the number of estimated documents that could be reviewed by a RAL Space analyst per iteration.

We assumed that the documents of most importance—i.e., those of greatest relevance—would be the ones that consistently appear at the top of the search results. We thus presented a ranked list of documents to RAL Space, with the ranking being based on the number of times that the document was retrieved by Google's Custom Search API over the course of each iteration—i.e., cumulative retrieval occurrences from programmatic releases of queries—see Palomino et al [22] for more details regarding the use of Google's Custom Search API.

Once the top-ranked 50 documents per iteration were chosen, we divided them into three different categories: academic papers, news articles and standard documents. The academic papers comprised, mostly, peer-reviewed papers relevant to the scan. The news articles were, mostly, press releases and news articles available on the Web; and the list of standard documents consisted of documents retrieved as a result of our queries that were not published by news websites or online academic journals. All the documents that we delivered, regardless of the category, were published between 2010 and 2012, exclusively.

V RESULTS

We previously conducted a benchmarking study between September and October 2010 in collaboration with Lloyd's of London [37], one of the global leaders in the insurance market. The goal of that study was to use our prototype for framing decision making on novel risks—specifically risks associated with space weather and how these might affect terrestrial and near-Earth insurable assets [22]. As part of the study, we were able to identify several documents that Lloyd's Emerging Risks Group analysts considered very relevant to assess insurance exposure; yet, the number of very relevant documents retrieved per week decreased as the experiment progressed, while the number of non-relevant documents retrieved increased over the same period [22].

Table IV displays the precise numbers of very relevant, relevant and non-relevant documents retrieved weekly in our study with Lloyd's of London—relevance feedback was not employed in that study and the relevance of the documents was evaluated according to the criteria developed by Lloyd's analysts —see Palomino et al [22] for full details.

TABLE IV. LLOYD'S EVALUATION RESULTS

	Week 1	Week 2	Week 3	Week 4
Very relevant	29	19	11	5
Relevant	66	64	74	74
Non-relevant	5	17	15	21

Although there were reasons to justify why most of the very relevant documents retrieved in our Lloyd's study were discovered in the first week, one of the major goals of the current study, and a motivation for our interest in relevance feedback, was to improve the performance of our prototype to make sure that the retrieval of relevant documents remains constant over the length of the experiment.

The scanning exercise undertaken with RAL Space comprised three iterations between 12 and 19 October 2012. Table V shows the exact number of very relevant, relevant and non-relevant documents retrieved per iteration. Table V shows that the number of very relevant documents decreased by one in the second iteration but then remained constant, which is an improvement over the results of the Lloyd's experiment, where the number of very relevant documents decreased by 10 after the first set of results and kept decreasing afterwards—see the first row in Table IV.

TABLE V. RAL SPACE EVALUATION RESULTS

	Iteration 1	Iteration 2	Iteration 3
Very relevant	16	15	15
Relevant	15	23	24
Non-relevant	19	12	11

As explained above, the 50 documents that we delivered per iteration to RAL Space were divided into academic papers, news articles and standard documents—all of them published between 2010 and 2012, exclusively. The specific breakdown per category and iteration is shown in Table VI.

TABLE VI. RAL SPACE EVALUATION RESULTS PER CATEGORY

First iteration

	Very relevant	Relevant	Non-Relevant
Academic	8	8	6
Standard	6	5	4
News	2	2	9

Second iteration

	Very relevant	Relevant	Non-Relevant
Academic	7	15	5
Standard	7	6	5
News	1	2	2

Third iteration

	Very relevant	Relevant	Non-Relevant
Academic	8	14	4
Standard	5	7	3
News	2	3	4

Due to the involvement of RAL Space in the Q-Detect project, academic papers were considered of particular importance for RAL Space's review. Table VI shows that the number of very relevant academic papers discovered by our prototype decreased only in the second week—decreased by one—but remained almost constant for the entire length of the experiment, which shows the potential of relevance feedback for searches within online journals.

To further evaluate the performance of our prototype, we used precision, one of the most common measures for evaluating the performance of information retrieval systems [38]. Precision is defined as the fraction of retrieved documents that are relevant to the search. For this experiment, we computed precision by considering all the documents evaluated by the analyst as being relevant or very relevant to be at least relevant, and compared these to the total number of documents presented to RAL Space each week—i.e., 50. Table VII displays the precision of our prototype per iteration. The final column shows the overall precision value for the entire experiment—namely, 72%. Note that the precision of the prototype actually increased on a weekly basis. Also note that the number of non-relevant documents—as indicated in Table V—decreased over the experiment, though not by much.

TABLE VII. PRECISION MEASURED PER ITERATION

	Iteration 1	Iteration 2	Iteration 3	Overall
Precision	62%	76%	78%	72%

A possible explanation as to why the number of relevant documents decreased as the Lloyd's experiment progressed is related to the timescale of the evolution of space weather documents on the Web. A period of four weeks might be insufficient to capture a significant number of additional newly published documents on space weather after our first search—i.e., after the first release of queries has been made. Consequently, the very relevant documents retrieved in the first week of the experiment were likely to be the most relevant ones for the entire experimental period of one month. To support this, we were able to verify that most of the documents marked as very relevant by Lloyd's Emerging Risks Group analysts were discovered in the first week of the experiment, but we could not include them in the results for the first week because we were restricted to a maximum of 100 documents per week.

As opposed to the case of the Lloyd's experiment, in the horizon scanning exercise undertaken with RAL Space, where we experimented with the use of relevance feedback, we can confirm that none of the documents delivered to RAL Space in the final iteration was discovered previously, and only two of the relevant documents delivered in the second iteration were discovered in the first week. The reason why we were able to find new documents and maintain the number of very relevant documents per week was that our relevance feedback implementation allowed us to modify the queries to reach different areas of the Web that we would not have been able to approach by releasing the same queries for all the iterations of the experiment.

Ideally, we would have liked to use recall as well to evaluate the performance of the prototype [38]. However, it is infeasible to measure recall for a Web-based system, since it is very difficult to determine all the existing documents on a given topic that are available online at a particular time. In addition, it should be noted that the horizon scanning prototype proposed here is not designed to return all relevant documents, but instead 50 documents per iteration.

VI CONCLUSIONS

Relevance feedback provides a method for reformulating queries based on previously retrieved relevant and non-relevant documents. A simple vector modification process that adds new keywords to queries and scales up or down the importance of existing keywords seems very useful. In view of its simplicity, we recommend that this process should be incorporated into operational text retrieval for horizon scanning systems and applications. Poorly processed feedback may lead to deterioration in retrieval effectiveness, which is a major limitation for relevance feedback implementations, but, when properly employed, the overall precision is improved, as shown in Section V.

As an opportunity for future work, we are considering mining social networks—particularly Twitter [39]—as a potential source of data for horizon scanning work. We are aware of the use of Twitter in financial applications, such as those employed by Derwent Capital Markets [40] and Palantir Technologies [41], whose foundations rely on the work by Bollen et al [42], and we realise that relevant information for horizon scanning that has been published originally by science and technology websites has appeared in Twitter streams. Thus, it is worth contemplating the monitoring of such streams for horizon scanning purposes.

ACKNOWLEDGMENTS

We are very grateful to Michael Jackson—founder member, ambassador and consultant of Shaping Tomorrow—for reading our manuscript and clarifying text on Shaping Tomorrow.

Palomino, Taylor and Depledge are staff at the European Centre for Environment and Human Health—part of the University of Exeter Medical School. This is part financed by the European Regional Development Fund Programme 2007 to 2013 and European Social Fund Convergence Programme for Cornwall and the Isles of Scilly.

REFERENCES

[1] Chan, S.W.K. and J. Franklin, A text-based decision support system for financial sequence prediction. Decision Support Systems, 2011. **52**(1): p. 189-198.

[2] Gheorghiu, R., et al., Web 2.0 and the emergence of future oriented communities. Economic Computation & Economic Cybernetics Studies & Research, 2009. **43**(2): p. p1.

[3] Linstone, H.A. and M. Turoff, Delphi: A brief look backward and forward. Technological Forecasting and Social Change, 2011. **78**(9): p. 1712-1719.

[4] Sutherland, W.J. and H.J. Woodroof, The need for environmental horizon scanning. Trends in Ecology & Evolution, 2009. **24**(10): p. 523-527.

[5] Chairman of the Joint Intelligence Committee, Review of cross-government horizon scanning, 2012, Cabinet Office: London, UK.

[6] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2013. Trends in Ecology & Evolution, 2013. **28**(1): p. 16-22.

[7] Palomino, M.A., et al., Web-based horizon scanning: Concepts and practice. Foresight, 2012. **14**(5): p. 355-373.

[8] Salton, G. and C. Buckley, Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 1990. **41**(4): p. 288-297.

[9] RAL Space. RAL Space website. 2013; Available from: http://www.stfc.ac.uk/ralspace/default.aspx.

[10] Robertson, S.E. and K.S. Jones, Relevance weighting of search terms. Journal of the American Society for Information Science, 1976. **27**(3): p. 129-146.

[11] Rocchio, J., Relevance Feedback in Information Retrieval, in The SMART Retrieval System. 1971. p. 313-323.

[12] Harper, D.J., Relevance Feedback in Document Retrieval Systems: An Evaluation of Probabilistic Strategies. 1980: University of Cambridge.

[13] Lv, Y. and C. Zhai. Adaptive relevance feedback in information retrieval. in CIKM '09: Proceedings of the 18th ACM conference on Information and knowledge management. 2009. Hong Kong, China: ACM.

[14] Text REtrieval Conference (TREC). TREC Tracks. 2013; Available from: http://trec.nist.gov/.

[15] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2010. Trends in Ecology & Evolution, 2010. **25**(1): p. 1-7.

[16] Sutherland, W.J., et al., Horizon scan of global conservation issues for 2011. Trends in Ecology & Evolution, 2011. **26**(1): p. 10-16.

[17] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2012. Trends in Ecology & Evolution, 2012. **27**(1): p. 12-18.

[18] Carlsson, P. and T. Jorgensen, Scanning the Horizon for Emerging Health Technologies: Conclusions from a European Workshop. International Journal of Technology Assessment in Health Care, 1998. **14**(04): p. 695-704.

[19] Douw, K., H. Vondeling, and W. Oortwijn, Priority setting for horizon scanning of new health technologies in Denmark: Views of health care stakeholders and health economists. Health Policy, 2006. **76**(3): p. 334-345.

[20] Palomino, M.A., et al., Web-based Horizon Scanning: Recent Developments with Application to Health Technology Assessment. Business Informatics, 2012. **3**(25): p. 139-159.

[21] O'Malley, S.P. and E. Jordan, Horizon scanning of new and emerging medical technology in Australia: Its relevance to Medical Services Advisory Committee health technology assessments and public funding. International Journal of Technology Assessment in Health Care, 2009. **25**(03): p. 374-382.

[22] Palomino, M.A., A. Vincenti, and R. Owen, Optimising Web-based information retrieval methods for horizon scanning. foresight, 2013. **15**(3): p. 159-176.

[23] Palomino, M.A., T. Taylor, and R. Owen. Towards the development of an automated, Web-based, horizon scanning system. in Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on. 2012.

[24] Shaping Tomorrow. Shaping Tomorrow website. 2013; Available from: http://www.shapingtomorrow.com/.

[25] Truvé, S., Big Data for the Future - Unlocking the Predictive Power of the Web, 2011, Recorded Future: Cambridge, MA.

[26] Jackson, M., Personal communication, 2013.

[27] Douw, K., et al., Use of the Internet in scanning the horizon for new and emerging health technologies: A survey of agencies involved in horizon scanning. Journal of Medical Internet Research, 2003. **5**(1): p. e6.

[28] Douw, K., et al., "The future should not take us by surprise": preparation of an early warning system in Denmark. Int J Technol Assess Health Care, 2004. **20**(3): p. 342-350.

[29] Wild, C. and T. Langer, Emerging health technologies: Informing and supporting health policy early. Health Policy, 2008. **87**(2): p. 160-171.

[30] Salton, G. and C. Buckley, Term Weighting Approaches in Automatic Text Retrieval, 1987, Cornell University.

[31] Yahoo! Developer Network. Yahoo! Content Analysis Web Service. 2013; Available from: http://developer.yahoo.com/search/content/V2/contentAnalysis.html.

[32] Wilson, J.C. and D.J. Holland-Smith, White Paper: Dstl S&T horizon scanning, 2008, Defence Science and Technology Laboratory (Dstl).

[33] Google. Custom Search. 2012; Available from: https://developers.google.com/custom-search/.

[34] Purcell, K., J. Brenner, and L. Rainie, Search engine use 2012, 2012, The Pew Research Center's Internet & American Life Project: Washington, DC.

[35] Page, L., et al., The PageRank Citation Ranking: Bringing Order to the Web, in Stanford InfoLab1999.

[36] Q-DETECT. Q-DETECT: Developing tools for on-site phytosanitary inspection. 2013; Available from: http://www.qdetect.org/0_home/index.php.

[37] Lloyd's of London. Emerging Risks Special Interests Group. 2013; Available from: http://www.lloyds.com/the-market/tools-and-resources/research/exposure-management/emerging-risks/emerging-risks-special-interests-group.

[38] Manning, C.D., et al., Introduction to Information Retrieval. 2008: Cambridge University Press. 496.

[39] Twitter. Twitter: The fastest, simplest way to stay close to everything you care about. 2013; Available from: https://twitter.com/about.

[40] Wikipedia. Derwent Capital Markets. 2013; Available from: http://en.wikipedia.org/wiki/Derwent_Capital_Markets.

[41] Palantir Technologies. Palantir. 2013; Available from: http://www.palantir.com/.

[42] Bollen, J., H. Mao, and X. Zeng, Twitter mood predicts the stock market. Journal of Computational Science, 2011. **2**(1): p. 1-8.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1135–1140

Software Implementation of Common Criteria Related Design Patterns

Dariusz Rogowski
Institute of Innovative Technologies EMAG ul. Leopolda 31, 40-189 Katowice, Poland
Email: drogowski@emag.pl

***Abstract*—Writing evidence documents for evaluation and certification processes according to the Common Criteria security standard is a very difficult, time-consuming and complex task. Nowadays there are only a few, limited solutions based on templates and software tools which can efficiently support developers in preparing evaluation deliverables. This paper describes the results of an R&D project whose aim was to work out a computer-aided tool with built-in design patterns. Firstly, according to all security assurance requirements the design patterns in a paper version were prepared. Secondly, they were verified and validated by the developers in order to make some amendments and improvements. The conclusions were used as the source of functional requirements for a computer-aided tool. As a result a complete computer system was designed which implements the design patterns, knowledge base, evaluation methodology, and additional external supporting software. That solution facilitates and speeds up the development of the evidence documentation.**

I. Introduction

SECURITY features of IT products have received much attention in recent years due to quickly rising numbers of cyber-attacks on important data and information. That is why there is a big demand coming from governments and private users for trusted IT products countering such threats. These products can be more reliable thanks to the evaluation and certification of their built-in security functions. Assessment processes should be conducted by an independent licensed laboratory which can use a security standard with requirements for a product development and documentation.

Here, the Common Criteria for Information Technology Security Evaluation standard (referred to as "Common Criteria" or "CC" throughout this paper), also known as ISO/IEC 15408 [1]–[3], provides a set of development rules and evaluation requirements [4] for the security measures applied in IT products. The results of CC-based evaluation are accepted in countries which joined the Common Criteria Recognition Arrangement (CCRA). This arrangement allows end users to recognize certificates regardless of the country in which they were issued. Therefore the certified products of different vendors can be easily compared by the users which can choose the best option. On the other hand we should remember that CC does not define the product security features or functionality but it provides assurance that the process of specification, implementation and evaluation of the product has been made in a rigorous manner. This assurance is assigned to one of seven assurance levels reflecting the requirements met in the development process of the product.

The product is the subject of the evaluation against the given EAL requirements and then it is called the Target of Evaluation (TOE). Evaluation Assurance Level (EAL) reflects the degree of confidence a user can have in the results of the evaluation and performance of the TOE. The lower assurance levels, EAL 1 through 4, concern most products, and do not require evaluation of the software, only of the development process and documentation. These lower levels are recognized under CCRA whereas the higher EALs are generally country-specific [5] and require a source code of the product to be analyzed.

However, it has been found by many developers of IT secure products that preparation for the Common Criteria evaluation process is very difficult, time-consuming and needs a lot of knowledge of CC requirements, all due to the fact that a lot of evidence documents have to be prepared. That problem is very important and has to be solved in order to make the whole evaluation process cost-effective and developers-friendly.

One solution was based on a series of guidelines and supporting documents issued by German Federal Office of Information Security (BSI) – the leader of researches in the field of the Common Criteria standard. This guidance documentation gives some advice on the structure and contents of evidence documents. Some templates were issued and could be used by developers but still too much work has to be done on their own [6], [7].

The second solution was based on very few software applications which could support users in the preparation of evidence documents and security development process. Unfortunately, these applications provided only basic functionality. They are limited to making only two main security specification documents (Security Target – ST, Protection Profile – PP), discussed later in this paper [8], [9]. Moreover, some of the software tools are not supported and developed by their producers any more.

Although the solutions mentioned above were offered a few years ago, relatively little attention has been paid to other evidence documents needed for the evaluation process. The guides and computer-aided tools are focused mainly on the preparation of STs and PPs documents. Apart from that, there is weak integration of the guidance knowledge within the software tools. In addition, if the developers want to create the evidence document only by using the guidelines and

978-83-60810-53-8/$25.00

templates, they still have a lot of work to do by themselves. They have to plan the structure of the document and find out what kind of information they should write down in the given section. That is why preparing the documentation is still difficult and not effective enough to encourage the developers to do this task.

This paper presents a complete and integrated solution which was worked out in the CCMODE R&D project (Common Criteria compliant, Modular, Open IT security Development Environment) carried out by the Institute of Innovative Technologies EMAG. The aim of the project was to work out a methodology and tools to develop and manage development environments of IT security-enhanced products for the purposes of their future Common Criteria certification. As a result a set of design patterns (the core of the methodology) was developed and next implemented in the computer-aided system CCMODE Tools. Thanks to the software implementation of the design patterns the developers receive one complete solution which facilitates production processes of the TOE and related documentation.

The paper is organized as follows. Section II presents the state of the art. Section III describes shortly the basic evidence documents required for the CC evaluation process. Section IV explains the methodology used for working out the design patterns and their implementation into the computer tool. Section V gives an overview of the CCMODE Tools main modules and their functionality used for evidence documents preparation. Section VI contains conclusions and experiences gained during the usage of the tool with built-in design patterns.

II. State of the art

The best starting point for building the design patterns is the Common Criteria standard which comprises three parts [1]–[3]. The current version of CC was issued in 2012. The first part is a general introduction to the CC methodology with explanation of basic terms and definitions. The second part describes security functional requirements which determine the desired security behavior of a TOE. The third part, the most important for building the patterns, defines the assurance requirements for a TOE and evaluation criteria for PPs, STs and other evidence documents. There are many companion documents to the CC standard. One of them is the Common Evaluation Methodology (CEM) [4] which helps evaluators to conduct the TOE assessment process. It defines evaluation activities to be done by the evaluators and presents work units – the most granular level of evaluation work – that help to issue verdicts about the quality of security implemented in the TOE. Other documents like technical reports and users guides explain step by step how to build evidence documentation. For instance, the ISO/IEC Technical Committee for Information Technology issued a technical report that is a guide for the production of PPs and STs [10]. This report provides methodologies, techniques and practical tips that developers can use to prepare security specification documents in an efficient and consistent manner. BSI issued a guide for developers of the STs and PPs [11]. Apart from that there is a guide that offers assistance to less experienced developers by extracting the information about the evidence from CC [12]. It explains requirements concerning the structure and contents of documents to be provided for the CC evaluation process. Another guide concerns evaluation reports according to CC and gives some advice and recommendations on the structure of information provided in these reports [13]. The CC standard and all guides mentioned above are used by the developers in common practice for writing evidence documents. But this way of work is very inconvenient because the developers must carefully read recommendations, check requirements and think about the necessary information to be provided in every new document each time they begin a project.

Although this guidelines-based approach helps the developers to work out documentation, it still does not allow to get rid of inefficient and time-consuming work. That is why some software aiding tools were applied to enhance the work with design patterns. Most of the software tools are dedicated only to preparing security specification documents (ST, PP) [14]. For example, an MS Windows application "CC Toolbox" sponsored by the National Information Assurance Partnership (NIAP, the US government initiative) used to assist users in writing ST and PP but is not longer supported and available. In [15] a generator of security target templates, named "GEST" was presented that can automatically generate security target templates from already evaluated and certified security targets. One of the Spanish CC licensed laboratories "Applus" presented a tool that reduces and automates some developer's activities of evidence documents preparation [8]. Another software tool, "TL SET", was introduced by Trusted Labs [16]. It is a smart editor for Security Targets and Protection Profiles. It integrates predefined libraries of the Common Criteria functional and assurance requirements and a user-friendly graphical interface to fill out the documents. There are also tools with built-in OWL language (OWL – Web Ontology Language) [17]–[20]. These tools are dedicated to build functional specification of the TOE and security problem definition.

So far all the solutions based on guidelines and computer tools have concerned mainly two basic documents: ST and PP. This paper presents a solution which allows to produce all the necessary documents and uses the context-sensitive help based on the CC standard and supplementary documents. The next section describes how complex a task of preparing all the evidence documents can be due to the fact that several documents have to be created.

III. Evidence documents

In the Common Criteria evaluation process security functions of the TOE are evaluated according to security assurance requirements (SARs) in the given EAL. Many documents should be prepared for the needs of the evaluation.

The most important is the Security Target (ST). The ST describes a specific TOE and is written by the developer. The ST can be based on a document called the Protection Profile (PP). The PP describes the general requirements for a TOE type and is used as a template for many different ST

documents. The ST consists of a security problem definition; security objectives; security requirements; a summary specification – showing how the security functions are implemented in the TOE. The ST document claims conformance with the declared EAL and this determines all requirements which have to be fulfilled by the product and described in evidence documentation.

The EAL package consists of assurance components which are organized into classes and families. The following descriptions of classes also include their short names (in brackets) which are commonly used in the CC standard. The Protection Profile Evaluation (APE) and Security Target Evaluation (ASE) classes describe the content and presentation of the PP and ST documents. The Development (ADV) class encompasses six families and nineteen components; it provides information about structuring of the TOE security functionality. The Guidance Documents (AGD) class is divided into two families (with one component for each family); it provides the requirements for preparative and operational user guides. The Life-cycle Support (ALC) class consists of seven families and twenty one components; it concerns the aspects of establishing discipline and control in the TOE development and maintenance during its whole life-cycle. The Tests (ATE) class encompasses four families and twelve components; it provides assurance that the TOE security functions were tested and they operate according to their design descriptions. The Vulnerability Assessment (AVA) class has only one family with five components; it addresses the possibility of exploitable vulnerabilities introduced in the TOE or in its development or operational environment. The Composition (ACO) class encompasses five families and eleven components; it assures that the TOE composed of other evaluated TOEs will operate securely. For instance, the EAL 3 package has fifteen components and for each one a proper evidence document has to be prepared.

On the basis of all assurance components taken from EAL packages the design patterns were worked out and then implemented into the computer tool.

IV. Methodology

In the first part of the CCMODE project a set of design patterns in the form of MS Word documents with predefined chapters and sections was prepared. The patterns were validated and assessed by independent experts in the field. Although they assessed the patterns as very helpful, they proposed some difficult and repeatable operations which can be automated by a computer tool. These insights allowed to make some functional assumptions for the CCMODE Tools system.

In the project there were design patterns created for all components of security assurance requirements (SARs). Next the patterns were verified and validated by developers chosen from the software and hardware industry. The validation was made upon the use cases method. The developers used selected design patterns to make evidence documentation of their software and hardware IT products. As a result of the validation, necessary changes and amendments were incorporated into the patterns. Furthermore, the developers concluded that some automation features should be implemented into the patterns.

In the next project stages a prototype of the software was developed. The prototype was next validated in two selected development environments of software and hardware IT products by using the case study method. A few evidence materials were prepared by developers. Documents for the TOE (ADV class) and for the environment (ALC class) were prepared. The case studies showed what else should be implemented in the computer tool to make the work with documents more effective and easier.

As a result, the CCMODE Tools system was worked out. The system integrates: modules of the development environment management, design patterns, knowledge base, evaluation methodology, and external supporting software. The system can be integrated with other security standards, like an information security management standard (ISMS, based on ISO/IEC 27001) or business continuity management standard (BCMS, based on BS 25999) [21].

Next chapters describe functionality of the software system which implements the design patterns.

V. Applying design patterns in a computer system

Developers use the software tool to start the project of an IT product in accordance with the chosen EAL level. They configure necessary external systems and deliver basic information about the type of the product, roles and duties of the system users, life-cycle model of the TOE, software and hardware tools used, security standards and regulations. Consequently, the following developers' actions can be automated by the software tool:

- verification of the development environment conformance with the CC standard;
- developing a security specification of the TOE in the ST document;
- providing the security problem definition that has to be solved by the TOE;
- specifying security objectives and security functional requirements to resolve the security problem;
- preparing evidence documentation with the use of the design patterns;
- defining life-cycle models for different types of IT products;
- testing the TOE and flaws remediation; establishing communication channels for flaws reports.

The actions mentioned above were next implemented in dedicated modules of the CCMODE Tools system. The following subsections describe the main modules.

A. CCMODE Tools system

A general model of the system is depicted in Fig. 1. The model consists of the Environment Management Tool (EMT), documents generator (GenDoc), knowledge base, evaluation module, external supporting systems, optional security systems (BCMS or ISMS) which can be used as an additional source of assurance to the whole development environment [22].

EMT is the main module which supports the configuration and management of the IT products that are to be built in the development environment. EMT makes it possible to define the system users and their roles in the project, the desired EAL level and life-cycle model.

The knowledge base is a source of context-sensitive help about the CC requirements and guidelines. It includes design patterns, terms and definitions which can be obtained by other modules. It also comprises the guidelines that help to resolve typical security problems with the use of predefined security objectives, threats, assumptions, and security policies.

There are also external systems in CCMODE Tools which support:

- assigning a version number to files and documents – Subversion (SVN) application;
- modeling, development and analyses which are made with the use of UML (Unified Modeling Language) – Enterprise Architect (EA);
- flaws reporting and flaws remediation – Redmine;
- management and planning of TOE tests – TestLink.

The evaluation module is used to verify the development environment against the CC requirements and to evaluate evidence documents according to the CEM methodology.

If the project of the IT product is completely configured then creating evidence documents can start by using the documents generator called GenDoc for short.

B. Documents generator (GenDoc)

GenDoc is used for editing evidence documents based on the design patterns. In order to evaluate the TOE, an ST document and accompanying documents must be prepared. These additional documents are determined by the chosen EAL and its SAR components.

This section shows on the example of an ST document how the software tool is used for filling in the patterns. Fig. 2 depicts the example of the GenDoc window with the ST design pattern. The general structure of the patterns, context-sensitive help and data fields were described. The precise details of the security development procedure and working out the evidence documents in the context of biometric devices can be found in [23].

Every pattern in GenDoc was prepared as a tree of data fields which represent chapters, sections and subsections of the output document. The tree is based on the requirements of the given CC component. The colors of branches show which fields have to be filled in by the user (red ones), which are already filled (black ones), and which are without any data (brown ones). The gray colored fields are automatically filled in with the information taken from the knowledge base and external modules: EA, EMT, SVN, TestLink.

In order to complete the document, the user must follow all the tree branches and find out which fields have to be completed. Every field has its own context-sensitive help which gives necessary guidelines and hints about the information to be delivered.

C. Context-sensitive help

Preparation of data fields content can be facilitated by context-sensitive help. This help is accessible from the main window of GenDoc by the link “Help – access to knowledge base”. There were five types of help applied: “ready to use” – it comprises a text which is ready to use by the user without the necessity to change any information in it; “Common Criteria help” – it comprises all the information and requirements taken from CC; “hints” – these are interpretations, tips and guidelines; “example” – it is an optional text which illustrates what kind of data can be written in the given field; “data source” – it indicates an external system which is the

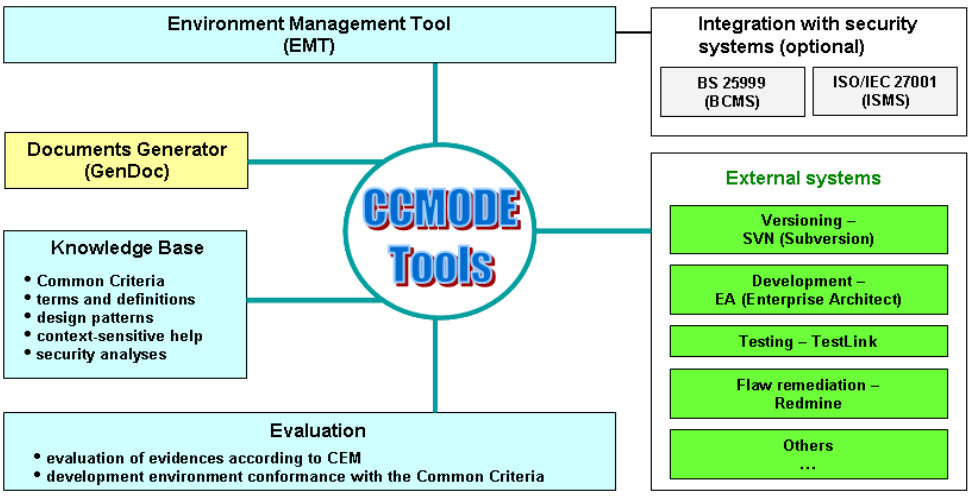


Fig. 1. The general model of the CCMODE Tools system

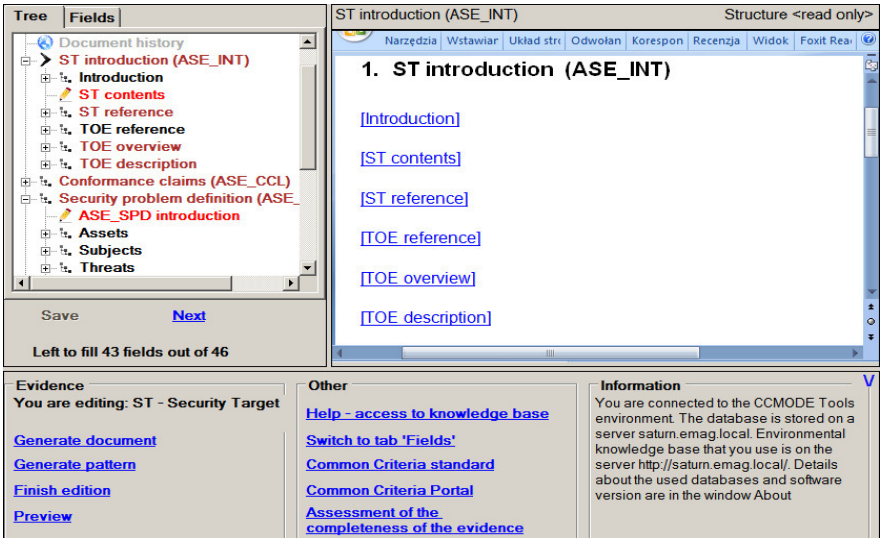


Fig. 2. The GenDoc window with the ST design pattern

source of data for the given data field. All the design patterns implemented into the CCMODE Tools system have similar representations. They contain data fields with precise instructions how to generate a complete evidence document.

At every stage of the edition process the data fields can be reviewed and checked. Verification of the document can be done with the use of the evaluation module as it is described in the next section.

D. Generation and revision of evidence documents

After completing all the information in the pattern, the user can verify the output document by using an evaluation module which is a part of the EMT system (Fig. 3). This module enables to check the document according to the CEM evaluation methodology.

In general, the methodology specifies elements which describe evaluation tasks to be done by the evaluator. These tasks give precise information how each security assurance component should be checked. Every task consists of a set of questions referring to the content and form of the evidence document. These questions are grouped in the so called work units. The answers lead to work units verdicts which can have one of three possible states: pass, fail or inconclusive. Each verdict needs short justification. All verdicts are initially inconclusive and remain so until either a pass or fail verdict is assigned. Verification of the evidence document is positive when all the verdicts are passed.

The evaluation module consists of work units with their detailed descriptions and has an answer form with a built-in justification field as it is depicted in Fig. 3. The developer has to answer all these questions which pertain to the verified document.

The enhanced version of the evaluation module was applied in GenDoc where the work units are directly connected to the relevant chapters and subsections of the evidence document in order to make the verification process easier and faster. This way the developer can see the content to be checked and the relevant work unit in one GenDoc window.

After verification, the complete document can be generated as an MS Word document and saved in the SVN repository. This document can be edited in a standard MS Word editor. The document has a fixed structure with chapters, sections and subsections. It contains also footnotes with hints and guidelines.

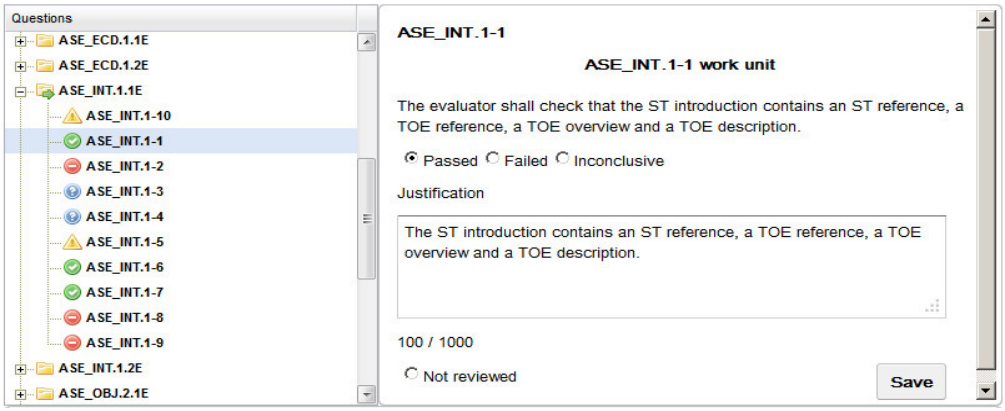


Fig. 3. The evaluation module of the EMT system

VI. Conclusions

This work presented software implementation of the design patterns which were worked out in the CCMODE project. The patterns were positively checked and validated by developers of IT secure products but at the same time they demanded for some automation features. This is why the CCMODE Tools system and the documents generator GenDoc were developed in order to support preparing of evidence documentation.

The proposed solution based on the patterns-based approach improves the IT security development process. It overcomes a lack of knowledge and experience of the user. The context-sensitive help connected to every field of the pattern allows the developers to concentrate only on writing the proper content.

The software tool facilitates and speeds up the IT security development process and improves the quality of evidences, which become more consistent and include all details required by the CC assurance requirements. The CCMODE Tools system gives a great chance to prepare all documentation for successful Common Criteria evaluation process. Additional self-evaluation and verification enhanced functions are also implemented in the GenDoc tool. These offer a practical way of documents evaluation according to the CEM methodology but it will be the topic of the next paper.

Future work will be focused on building a standalone, independent GenDoc application which could work without the EMT framework. It is demanded by some developers who elaborate only evidence documentation. In the future work it will also be considered to adapt the documents generator to produce documents according to different typesetting systems.

References

[1] *Common Criteria for Information Technology Security Evaluation (Version 3.1, Revision 4) Part 1: Introduction and general model (ISO/IEC 15408-1), September 2012*.

[2] *Common Criteria for Information Technology Security Evaluation (Version 3.1, Revision 4) Part 2: Part 2: Security functional requirements (ISO/IEC 15408-2), September 2012*.

[3] *Common Criteria for Information Technology Security Evaluation (Version 3.1, Revision 4) Part 3: Part 3: Security assurance requirements (ISO/IEC 15408-3), September 2012*.

[4] Common Methodology for Information Technology Security Evaluation (Version 3.1, Revision 4) Evaluation Methodology, September 2012.

[5] W. Jackson "Under attack." (GCN), August 10, 2007.

[6] A. Bialas, "Semiformal Common Criteria compliant IT security development framework." Studia Informatica 2008, 29, No. 2B(77); Silesian University of Technology Press Gliwice, Poland.

[7] D. Rogowski, P. Nowak, "Pattern based support for Site Certification." *W. Zamojski et. al. (Eds.): Complex Systems and Dependability*, AISC Vol. 170, pp. 179-193, Springer-Verlag Berlin Heidelberg 2012.

[8] I. Kane, "Automated tools for supporting CC design evidence," 9th International Common Criteria Conference, Jeju, South Korea, 2008.

[9] 13th International Common Criteria Conference, Paris, France, 2012.

[10] *ISO/IEC TR 15446: 2009 "Information technology – security techniques – guide for the production of Protection Profiles and Security Targets"*.

[11] "The PP/ST guide," Version 1, Revision 6.2, BSI, August 2007.

[12] "Guidelines for developer documentation according to Common Criteria Version 3.1," BSI, 2007.

[13] "Guidelines for evaluation reports according to Common Criteria Version 3.1," Version 2.00 for CCv3.1 rev. 3, BSI, 2010.

[14] Higaki, Wesley Hisao: *"Successful Common Criteria evaluations. A practical guide for vendors"*, 2010.

[15] D. Horie, K. Yajima, N. Azimah, Y. Goto, J. Cheng, "GEST: A generator of ISO/IEC15408 Security Target templates". *Computer and Information Science* 2009, pp 149-158.

[16] www.trusted-labs.com: Accessed May 2013.

[17] A. Bialas, "Specification means definition for the Common Criteria compliant development process – an ontological approach." In: *Zamojski W., Mazurkiewicz J., Sugier J., Walkowiak T., Kacprzyk J. (Eds.): Complex Systems and Dependability*; AISC Vol. 170, Springer-Verlag: ISBN 978-3-642-30662-4, pp. 37-54, 2012.

[18] A. Bialas, "Security-related design patterns for intelligent sensors requiring measurable assurance." *Electrical Review*, ISSN 0033-2097, vol. 85 (R.85), Number 7/2009, pp. 92-99, Sigma-NOT, Warsaw.

[19] A. Bialas, "Common Criteria related security design patterns for intelligent sensors – knowledge engineering-based implementation. In: *SENSORS*, Volume 11, Issue 8, Pages: 8085-8114, DOI: 10.3390/s110808085, 2011.

[20] A. Bialas, "Patterns improving the Common Criteria compliant IT security development process." In: *Zamojski W, Kacprzyk J, Mazurkiewicz J, Sugier J, Walkowiak T (Eds.) Dependable Computer Systems*; AISC, Vol. 97, pp. 1-16 Springer-Verlag: Berlin Heidelberg, 2011.

[21] J. Baginski, A. Bialas "Validation of the software supporting information security and business continuity management processes." In: *Zamojski W., Mazurkiewicz J., Sugier J., Walkowiak T., Kacprzyk J. (Eds.): Complex Systems and Dependability*; AISC, Vol. 170, Springer-Verlag: Heidelberg, 2012, pp. 1-18.

[22] A. Białas, "Computer support for the development process of security-enhanced IT products," Original Polish title: Komputerowe wspomaganie procesu rozwoju produktów informatycznych o podwyższonych wymaganiach bezpieczeństwa. Wydawnictwo Instytutu Technik Innowacyjnych EMAG, financed by UE POIG 1.3.1, Katowice ISBN 978-83-932737-8-2, , 2012.

[23] A. Białas, "How to develop a biometric system with claimed assurance," Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 775–780, ISBN 978-1-4673-4471-5 (Web), IEEE Catalog Number: CFP1385N-ART (Web).

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1141–1144

IT Security Threats in Cloud Computing Sourcing Model

Artur Rot
Wroclaw University of Economics
ul. Komandorska 118/120
53-345 Wroclaw, Poland
Email: artur.rot@ue.wroc.pl

Malgorzata Sobinska
Wroclaw University of Economics
ul. Komandorska 118/120
53-345 Wroclaw, Poland
Email: malgorzata.sobinska@ue.wroc.pl

***Abstract*— New information technologies have been developing nowadays at an amazing speed, affecting the functioning of organizations significantly. Due to the development of new technologies, especially mobile ones, borders in the functioning of modern organizations diminish and models of running business change. Almost all organizations are involved in some way in sourcing activities, and each of them develops a sourcing relationship that suits its particular needs. In this article different kinds of outsourcing models were discussed, which are applied in the contemporary management, with particular emphasis put on cloud computing.**

The main aim of this article is to present the most important risks related to the introduction of management models based on the most recent IT technologies, e.g. cloud computing, and emphasizing the role of appropriate IT security management in the times of globalization of organization virtualization.

I. Introduction

Information resources have nowadays strategic significance and have key influence on gaining the competitive advantage by all types of enterprises. Organizations are forced to look for still better and more effective IT solutions that enable for example: IT cost reduction, access to the best technology and best IT hardware and software experts, IT systems security etc. One of such new models referring to IT services is cloud computing.

Ongoing research projects investigate client and vendor capabilities required to successfully implement these sourcing models and initiatives, and how to manage knowledge and expertise in various sourcing contexts to improve efficiency and outcomes of sourcing engagements. Organizations are facing a large variety of possibilities to choose from when making sourcing decision. They should take into consideration a lot of factors (both positive and negative) to be able to make the right decision.

The London School of Economics's research regularly finds that firms that outsource give away too much of their technical capability. It is a challenge to retain skilled people in house paying them the market rate and offering them interesting, value-adding work. The alternative to such an "invest to save" HR approach is to put at risk the long-term health of the deal [4, p.10]. This can be especially true dealing with immature markets, such as cloud services.

In the next part of the paper we will discuss what makes cloud computing popular and what are the main risks of cloud computing sourcing model.

II. Evolution of Sourcing Models

Oshri, Kotlarsky and Willcocks, who have observed outsourcing market since years, notice that various types of global sourcing models have begun to emerge. The major difference between these models lies in whether the function is performed by a subsidiary business unit of the firm or an external vendor (or by both, as a joint effort), and also whether the function is performed on the firm's premises (i.e., on-site) or off-site, which can be onshore (in the country where the organization is located), nearshore (in a neighbor country), or in an offshore location [3, p. 25].

Insourcing means managing the provision of services internally, if needed- through buying in skills that are not available in-house, on temporary basis (for example by staff augmentation). Offshore or nearshore outsourcing means outsourcing contract with vendors situated in a different country from the client organization. Out-tasking- it is outsourcing on a small scale. It usually implies ongoing management of and support for selected packaged applications. Joint venture in the outsourcing or offshoring context – means a partnership between a client firm and offshore vendor whereby the parties contribute resources to the new deal/project Shared services- it is an operational approach of centralizing administrative and business processes that were once performing in separate divisions or locations- for example: finance, IT, human resources. In literature there are listed also the following sourcing models based on Internet delivery of products or services: cloud computing, software as a service, crowdsourcing, and microsourcing. Sourcing decisions should be made jointly by business and IT executives [4, p.11].

III. Attributes of Cloud Computing Model

Although the idea of cloud computing has been around for quite some time, it is an emerging field of computer science. Cloud computing can be defined as a computing environment where computing needs by one party can be

978-83-60810-53-8/$25.00

outsourced to another party and when need be arise to use the computing power or resources like database or emails, they can access them via Internet. Cloud computing is a recent trend in IT that moves computing and data away from desktop and portable PCs into large data centers.

During the past few years, cloud computing has grown from being a promising business idea to one of the fastest growing parts of the IT industry. IT organizations have expresses concern about critical issues (such as security) that exist with the widespread implementation of cloud computing. These types of concerns originate from the fact that data is stored remotely from the customer's location; in fact, it can be stored at any location. Security, in particular, is one of the most argued-about issues in the cloud computing field. Comparison of the benefits and risks of cloud computing with those of the status quo are necessary for a full evaluation of the viability of cloud computing.

In Hauke's and Owoc's opinion facilities available via cloud computing are strictly determined by properties of this technology based on Internet resources [2, p. 125]. One can identify two essential concepts in "cloud" environment-abstraction and virtualization and several properties that can be expressed as secondary (such as scalability, flexibility, availability, measurability, efficiency, low costs of services, low barrier to entry, security).

Security – the most "sensitive" and disputable feature of CC. Theoretically all potential problems should disappear (all necessary tasks are performed by specialized partner). No doubts, problems with database recovery should be served professionally, however, a risk of data loosing or data leaking can occur [2, p.126].

Cloud computing solutions are offered by such large organizations as IBM, Microsoft, Amazon, Google and others. They can give a lot of benefits but they have also some limitations. There are numerous challenges facing organizations when considering cloud computing. Willcocks and Lacity, in their analysis, focus on the four challenges which seem particularly critical in the development of cloud use within organizations: weighing up the security and legal risks, defining the relationship through contracting, the lock-in dilemma, managing the cloud [5, p.290-296].

Cloud represents a great opportunity, but there are also strong challenges to take if an organization wants to use its potential for business advantage.

IV. New Challenges for IT security management

The existence of a company in cyberspace and an opportunity to communicate with it via electronic media is often a minimum condition for the company to be perceived as a reliable and solid partner. Unfortunately, the development of IT, e-commerce and new business models (including various kind of IT sourcing) carries new risks apart from huge benefits. There are new threats, often incomprehensible and underestimated by company management. The basic premises indicating the emergence of new kinds of risk accompanying the functioning of management IT systems, include e.g. the following facts and circumstances [6, p. 11-13] [7, p. 80-82]:

- information has become one of the most important goods on the market, which hence increases its price, meanwhile generating a risk of its unauthorized interception,
- ruthless chase after information, which is especially typical of business and media environments, blurs the border between legal and illegal actions aimed at acquiring it,
- increasing the availability of IT systems, which are considered to be the condition of society civilization development expansion, which facilitates the development of cybercrime,
- most of documents and information, which were dispersed so far, now are stored in one place – in a computer, which makes them easily accessible, but in case of unauthorized access the scope of damage is extensive,
- technological complexity of company IT systems and their security, which makes it impossible for an average user to use them rationally in order to minimize all threats,
- common lack of knowledge or unawareness of information systems threats, which results in lack of compliance to certain requirements, procedures etc.,
- data gathered in IT systems remain under the supervision of system administrators, which results in the fact that a few people have an insight into very important data and can modify them practically without anybody noticing,
- security systems are expensive and happen to be neglected for the sake of efficient performance of an individual,
- companies offering security tools for IT systems are often uncertified, which makes their products fallible and often of low quality,
- companies developing various IT solutions often offer fallible systems, which are full of mistakes,
- the Internet brings in new threats, since it is where you can easily find software for hacking IT systems.

Nowadays specialized institutions (e.g. CERT, Computer Security Institute, etc.) publish statistical data concerning the probability of occurrence of given kinds of threats [see: 8, 9]. According to various statistics, intentional or unintentional actions of organization staff (negligence, lack of concentration, incompetence) and purposeful actions of dissatisfied or sacked employees.

It is worth discussing here conclusions derived from the report 2011 TMT Global Security Study, prepared by the counseling company Deloitte. According to this report, one fifth of companies from the technology, media and telecommunication sector (TMT) find personnel mistakes to be the major threat for the company IT security. Another huge risk for the company IT systems security is the use of

mobile devices by employees (personal smartphones, tablets, laptops) at work. The risk is in this case related to data confidentiality, application popularization and IT support.

Similar conclusions can be derived from studies conducted by the counseling company PricewaterhouseCoopers (PwC). The 2012 Global State of Information Security Study was conducted in 2011 all over the world. The results were obtained on the basis of answers provided by more than 9600 managers, vice-presidents and IT and information security directors from 138 countries, including Poland. According to them, current or former employees are perceived by companies as the major source of risks for IT systems. However, respondents pay more and more attention to a different category of external risks: 17% of respondents indicate clients, and 15% - business partners and suppliers as the key risk sources. In Europe during the past two decades the percentage of companies which require that the suppliers adjust security policies to their requirements, dropped from 31% to alarming 22%; only 18% of companies keeps records of all suppliers processing personal data of customers of employees.

Threats which are less likely to occur include ICT networks and IT systems failures and natural disasters (fire, flood, hurricane and earthquake). Threats related to unauthorized access (of e.g. hackers) and activities of malicious software are relatively unlikely to happen. When assessing the frequency of occurrence of such risks, it is recommended to take into account the specificity of the given company and its environment. The basic classification of IT systems security threats is presented in table 1.

V. THREATS RELATED TO VIRTUALIZATION AND CLOUD COMPUTING

Ernst & Young company conducted the 'Global Information Security Survey' in 2011 [12]. The study group consisted of 1700 organizations, including the largest and the most dynamic companies from 52 countries, from different trades (banking, finance, insurance, motorization industry, public administration, transport, health service, trade). The company has been conducting such studies for 14 years, and their aim is e.g. identification of the most critical kinds of risk in the field of new IT technologies. As it can be concluded from this study, respondents recognize trends related to new kinds of risk, since more than 72% of them estimates that the level of risk related to the development of such technologies, as the aforementioned mobile devices, cloud computing and social networking sites, is larger and larger. 46% of organizations recognize also increasing risk levels related to internal company threats. Polish results are consistent with global results as far as the assessment of the key threats is concerned. The study conducted by the company in 2010 [12] revealed that more than 64% of respondents found data safety to be one of the five key risk areas. It results directly from the fact that for 73% of Polish

TABLE I.
CLASSIFICATION OF IT SYSTEMS SECURITY THREATS

Criterion	Classification of threats	Examples of threats
Role of a man	Threats independent on a man	Atmospheric discharge, flood, fire, humidity etc.
	Threats dependent on a man	Illegal modification of software or data, disclosure or deletion of data, illegal copying and installation of software, damage or deletion of software or data, stealing computer equipment or accessories, storage of resources prohibited by law, mistakes made due to the lack of knowledge, unintentional loss, damage, deletion or disclosure of data to unauthorized persons, etc.
Subject of influence	Threats related to computer systems	Interruptions in electric energy supply, intentional and unintentional human actions (e.g. mechanical damage, configuration errors, incompetent use or maintenance, etc.), unexpected failures of mechanical and electronic elements of computer equipment, etc.
	Threats related to software	Mistakes made by the software manufacturer, mistakes made intentionally or unintentionally by employees or third parties (e.g. incorrect installation, configuration, implementation, deletion or modification of software, introducing malicious programs, blocking correctly functioning applications, illegal access, illegal usage or copying of software, etc.
	Threats related to data	Unauthorized access to data, unauthorized modification of data (e.g. damage, change of content, deletion, etc.), unauthorized copying of data, monitoring (phishing) of data, introducing incorrect data, denying the reception/sending of data, etc.
	Threats related to ICT networks	Intentional or unintentional human actions (e.g. stealing network components, physical damage of a network, wrong configuration, partial or complete blockage of network activity, phishing or unauthorized use of a network, etc.), ICT network failures caused by external factors (e.g. atmospheric discharge, fire, etc.), unexpected damage of electronic elements of a network, etc.
	Threats related to people	Failing to keep business and trade secrets by economic entity personnel (e.g. unintentional release or transfer of data due to the so called 'social engineering), informing workmates or third parties about security systems used in a company, sudden loss or resignation from work by the personnel as well as a situation, when employees having access to confidential information start working for a competitor.
	Threats causing financial losses	Loss of clients, business partners, decrease in turnover and company share in the market, interruptions in the functioning of a business entity, the necessity to exchange the offered products (especially in case of bank services), loss or damage of technology and software, financial sanctions, increase of insurance premiums, decrease in process efficiency and activities carried out in a company, necessity to hire additional employees, costs of outsourcing, judicial costs, penalty interest for breaching agreements)
Action results	Threats causing intangible damage	Loss of prestige and good name, loss of economic subject credibility in the eyes of clients and business partners (due to e.g. media interest in data safety breach), organizational chaos, loss of IT infrastructure efficiency, incorrect decisions made on the basis of falsified or incomplete data.
	Threats independent on a man	Atmospheric discharge, flood, fire, humidity etc.

Source: [16, p. 160-161]

respondents (and 53% of global respondents) the protection of brand and reputation is the most important aim of organizational safety policy, even more important than ensuring compliance with regulations in this field (60% of respondents in Poland and 56% around the world).

An example of a technology, which is developing very dynamically nowadays, is cloud computing, described in more details in point 3 of this article. The biggest problem for entrepreneurs interested in cloud services is issues related to the loss of data safety. According to the study conducted by Harris Interactive center, as many as 91% of respondents worry about the safety of public clouds, and approximately 50% of them indicate that safety issues are the largest obstacle in popularizing cloud solutions. Eelastic-security.com portal also decided to analyze this topic and surveyed suppliers of such services. The aim of the study was to check, how suppliers and cloud users perceive safety issues. The results how divergent the expectations of these groups are. 69% out of 127 surveyed suppliers claimed ensuring safety of cloud services was the sole responsibility of users, who thought the contrary. Most of them said that service providers are responsible for cloud safety or it is the resultant of suppliers' and users' action [13].

Although cloud computing has been known for a few years, it still remains a mysterious technology for most users, and hence it is perceived as highly risky [14]. As Gartner Group analysts confirm in their report entitled 'Safety risk assessment in cloud computing', processing data outside a company is related to a risk, therefore in order to minimize it, only checked solutions have to be applied [15]. Also the study conducted by E&Y explores the topic of cloud computing technology safety. 52% of respondents are concerned about data leakage, and 39% of them worry about the loss of control over information processed in a cloud.

Materialization of the above kinds of risk is one of the greatest threats for organizations using modern technologies. The consequence could be the loss of clients and business partners, decreased turnover and company's market share, the necessity to exchange the offered products.

Tangible damage of an organization may include damage of IT infrastructure, loss of valuable data and hence the necessity to restore it. Another consequence is the inaccessibility of IT systems, which may trigger financial losses, e.g. additional operational costs, loss of profits, claims of business partners, suppliers and clients for not performing services or performing them improperly, increased insurance premiums, claims under civil law resulting from torts, e.g. disclosing personal data, punishments imposed by public institutions, etc.

CONCLUSION

Nowadays, the dominant source of risk for an organization is the fallibility of IT systems, and one of the major sources – the level of data security. The presented studies confirm that new technologies, and with them – new business models/tools – generate new, so far nonexistent, threats, and are a source of new types of risks. Significant changes in the functioning of an organization – which result from ongoing globalization, increasing competition, automation, and in particular – development of IT and virtualization, become the fundament of a new perspective of the risk management process concerning IT security in organizations.

The cloud computing discussed in the paper may dominate IT services market; however, it has several drawbacks. Cloud computing does not remove the need for a sound process [1, p. 17]. As discussed in this paper, it may bring some opportunities, but even if organizations themselves feel "cloud ready" they must anticipate the capacity requirements in the cloud, be aware of new risks and manage IT security in accordance with new operation conditions.

REFERENCES

[1] *Strategies To Improve IT Efficiency In 2010.Using Predictive Analysis To Do More with Less*, April 13, 2010, A Forrester Consulting Thought Leadership Paper Commissioned By TeamQuest, http://www.teamquest.com/pdfs/whitepaper/forrester-it-efficiency-2010.pdf (Access: 18.04.2013).

[2] K. Hauke, M.L. Owoc, *Properties of cloud computing for small and medium sized enterprises*, [In:] Advanced Information Technologies for Management- AITM 2011, editors: J. Korczak, H. Dudycz, M. Dyczkowski, Wroclaw University of Economics Research Papers no 205, ISSN 1899-3192, Wroclaw 2011, p. 123-130.

[3] I. Oshri, J. Kotlarski, L.P. Willcocks, *The handbook of global outsourcing and offshoring*. Second edition, Palgrave Macmillan Ltd. – Houndmills Basingstoke Hampshire (UK) 2011.

[4] *Professional outsourcing*, Issue 7 Winter 2011, www.professionaloutsourcingmagazine.net (Access: 5.10.2012).

[5] L.P. Willcocks, M.C. Lacity, *The new IT outsourcing landscape. From innovation to cloud computing*, Palgrave Macmillan Ltd. – Houndmills Basingstoke Hampshire (UK) 2012

[6] J. Grzywacz J. (ed.) *Information systems security in banking institutions in Poland* (In Polish), Oficyna Wydawnicza SGH in Warsaw, Warsaw 2003, p. 11-13

[7] A. Barczak, T. Sydoruk *Management Information Systems Security* (In Polish), Dom Wydawniczy Bellona, Warsaw 2003, p. 80-82

[8] Online CERT security reports: http://www.cert.pl/raporty

[9] Online Computer Security Institute reports: http://gocsi.com/members/reports

[10] J. Muszynski *New Services, New Threats*, Networld of 15.12.2008, http://www.networld.pl/artykuly/329757/Nowe.uslugi.nowe.zagrozenia.html (Access: 18.09.2010)

[11] *Into the cloud, out of the fog - Ernst & Young's 2011 Global Information Security Survey*, http://www.ey.com/Publication/vwLUAssets/Into_the_cloud_out_of_the_fog-2011_GISS/$FILE/Into_the_cloud_out_of_the_fog-2011 GISS.pdf (Access: 21.02.2012)

[12] *The level of IT threats is rising in connection with the development of new technologies fog - Ernst & Young's 2010 Global Information Security Survey,* Feb. 2011, http://www.ey.com/PL/en/Newsroom/News-releases/PR11_Raport-GISS-2010 (Access: 24.02.2012)

[13] B. Matuszewska *Security in the cloud* (In Polish), Gazeta.pl 05.10.2011,http://komputerwfirmie.gazeta.pl/itbiznes/1,54790,10412168,Bezpiecznie_w_chmurze.html (Access:01.03.2012)

[14] *Cloud computing – Is It Secure In The Cloud?* (In Polish), 18.10.2011, http://internet-news.com.pl/cloud-computing-bezpiecznie-w-chmurze/ (Access: 10.03.2012)

[15] M. Bienkowski *Seven Threats for Cloud Computing Security* (Polish) http://webhosting.pl/Siedem.zagrozen.bezpieczenstwa.dla.komputerowych.chmur (Access: 29.01.2012)

[16] Nowicki A. Turek T. (ed.) Information Technologies for Economists. Tools. Applications (In Polish), Published by the University of Economics in Wroclaw, Wroclaw 2010, p. 160-161

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1145–1152

Modeling the Bullwhip Effect in a Multi-Stage Multi-Tier Retail Network by Generalized Stochastic Petri Nets

Bidyut Biman Sarkar
Department of Computer Applications, Techno India, Salt Lake, Kolkata, India
Email: bidyutbiman@gmail.com

Agostino Cortesi
DAIS
Università Ca' Foscari Venezia, I-30170 Venezia, Italy
Email: cortesi@unive.it

Nabendu Chaki
Department of Computer Science & Eng., University of Calcutta, 92 APC Road, Kolkata, India
Email: nabendu@ieee.org

Abstract— **Bullwhip effect (BWE) refers to the accumulation of stock flowing up and down along the supply chain management (SCM). It reduces the operating efficiency of the chain and blocks the operating resources. Some of the common causes of BWE are demand order variations, long lead times, competence defects between supply chain links, lack of communication among links in the chain, etc. There have been efforts to overcome these issues. However, very little work has been reported based on formal representation and analysis of resource flow in the supply chain system. In this work, a novel framework is proposed using Generalized Stochastic Petri-net (GSPN) model towards handling this issue in a distributed scenario. The analysis on the stochastic nets allows identifying the bottlenecks in the supply chain echelons along with customer relationship management (CRM). This has been used to rebuild infrastructure with the end-objective of reducing the BWE.**

Keywords: **SCM, BWE, BWE-Index, GSPN, Retail Network, CRM**

I. INTRODUCTION

Bullwhip effect (BWE) [37] is one of the trickiest evils for supply management systems. Sustainability for small and medium size enterprises becomes an issue when BWE is moderately high. Prompt delivery of product is more and more a key issue that provides competitive advantage of collaborative process [22]. Business challenges have increased along with the globalization with a view to adopt the changes and advance. A single stage, single tier chain with single customer is the simplest form of SCM, where there is no BWE, e.g., when a government. is a buyer of an item from a single supplier and no other roll players are involved in the process, or in case of a bend market where only manufacturer and buyer exists. However, we are getting to global collaborative retail chains like food retail, health and beauty products, clothing, durable goods, etc, where the success of the chain primarily depends on customer satisfaction, and the different actors play their role in different, distributed, geographical contexts, e.g., a South African garment manufacturing company may have retailers in America or Europe; orders issued from Europe, raw material is purchased in Korea, the material is dyed in Taiwan, and finished garments is assembled in Thailand.

The main aim of this work is to investigate the reasons of BWE at every stage of Supply Chain (SC) and measure it. Some of the common observations of BWE are communication gap, information twisting, and gaming for destabilizing the chain among the SC partners [20]. This leads to disproportionate inventory, fund blockage, loss of revenue and inept customer service [2]. Prompt decision making, demand order processing and monitoring of delivery logistics, a large distributed multi dimensional real time data repository is necessary [10]. The existing works on BWE points out qualitative aspects, and suggested measures to reduce BWE based on such qualitative analysis. The first quantitative measure of BWE was attempted using BWE index [11]. The present work provides experimental evidence by a significant simulation of the impact of stochastic Petri Nets enhanced with data mart controlling the entire Supply Chain operations in order to reduce the BWE.

Section 2 presents a brief literature survey on related scientific contributions. Section 3 introduces multi stage multi tier retail chain for demand analysis; Subsection 3.1 provides the GSPN model of the retail chain. Section 4 presents a revised GSPN model of the Retail Network. Section 5 describes some warehouse applications. The paper ends with concluding remarks in section 6.

II. LITERATURE SURVEY

A good number of works published in leading journals in last 15 years dealing with BWE are briefly reviewed and analyzed in this section. BWE can be seen as a random measure of performance of a SCM over some specific period of time. SCM performance is measured as a ratio of demand against delivery. If the ratio is greater than unity, the performance is considered to be negative [11], i.e., BWE prevails. The four broad classes of BWE studies Behavioral, Analytical, Industrial, and Dynamic approaches are presented herein.

Behavioral: Uneven customer demand is identified and capacity adjustment to do keeping inventory level unchanged is studied in [5]. Demand forecasting and order lead times are identified as root causes of BWE [6]. US census bureau data is used to observe dependency on the order variance ratio, and demand distortion of BWE [7]. Seasonal BWE is discussed and three forecasting levels are

This work has been carried out in the Department of Computer Science & Engineering, University of Calcutta, India.

978-83-60810-53-8/$25.00

used to monitor lead time to control BWE [8]. When seasonal variation is larger than demand uncertainty then demand to order variance ratio becomes insignificant for measuring BWE [12]. Rationing, Gaming, and Order Batching are the causes and VMI is used to control the BWE [16]. Amplified demand, data incompleteness and data aggregation are the primary causes of BWE discussed in [19]. Four operational aspects of BWE are demand forecast updating, order batching, price fluctuation, rationing and shortage due to gaming are identified [20]. Five basic functions of SC; plan, source, make, deliver and return are mapped to SCOR model and measure SC performance to monitor BWE [23]. A class room model with four echelon is used to explain the BWE, no formalism suggested [38].

Analytical: Logistic cost minimization techniques are discussed to control the BWE [1]. Transfer function plot is used to monitor the BWE on replenishment rules [15]. A nonlinear optimization model is proposed to reduce BWE by using a four stage echelon system on demand distortion, misperception of feedback, price variation, batch ordering, and strategic decisions. [21]. BWE is measured using the statistical technique ANOVA over information enrichment percentage, time to adjust inventory; time to adjust work in progress; production delay; and sales exponential smoothing {25]. Time delays are major causes of BWE at every stage of SCM. A mixed integer model is used to measure and monitor the BWE [34]. BWE propagation is measured using Fourier Transform [42].

Industrial: Surveyed thirty fast moving consumer goods (FMCG) to emphasize quantitative presence of BWE in SCM. In [18], as many as eleven key design guidelines are suggested to optimize SC. BWE studied over perishable products, customer demand, retailers and wholesalers order cycle varies, and delay in delivery at various touch points occurs. By adjusting the order cycle and delivery delay time the model can be optimized to stabilize the convergence function [45].

Dynamic: BWE at packaging business is studied. A stochastic demand reduction SC model suggested and confirms shorter the lead time, lesser will be the BWE [30]. IT enabled Service Oriented Architecture (SOA) proposes performance measurement (PM) matrices to evaluate SC performance and corresponding BWE effects over multiple channels [32]. An ERP framework is recommended for understanding all types of enterprise applications that maintain the SC and its corresponding BWE at each echelon [39]. A four stage tree structured SC model and a software simulation is designed and suggested the causes of reverse BWE [47].

In a single or a multi echelon SC one culminating point is that either from the behavioral side or from analytical viewpoint or industry specific models or it is IT enabled software driven systems there are some issues in common for occurring the BWE. However, not much holistic approach to control the BWE in a dynamic, multi-aspect decentralized, stochastic and non-linear environment is yet to be explored. It requires efficient modeling using appropriate tool suitable for distributed environment to embed stochastic behavior of demands in the model itself.

In this work, a high-level net model is proposed analyzed and used for demand driven SC to substantially control the BWE, if not eliminated. We choose Petri Net for modeling due to its strong mathematical formalism. A short review on Petri net and its high level extensions are presented in subsection 2.1.

2.1 HIGH LEVEL REVIEW ON NET MODELS

Petri Nets are directed bipartite graphical notations for modeling of discrete dynamic systems developed by Carl Adam Petri. It is often used as a meta-modeling tool to study and analyze the complexity among the large number of collaborative devices, business processes, distributed communications, and various process simulations. It provides two models: Structural and Behavioral [29].

A Structural Model is a directed graph representing the static part of the system. There are two kinds of nodes, places and transitions, represented by circles and rectangles. Places represent state variables and transitions represent transformers. The net is said to be ordinary when all arc weights are equal to one, i.e. each occurrence of adjacent transitions consumes one token from input to output place. Behavioral Models captures the dynamics of the system behavior using evolution rules for the marking. Markings are represented by tokens. The token at a place is its state value. The values change in adjacent states with the occurrence of transitions. Modeling of some service with simultaneous arrival of tokens to a queue ordinary net can't model such situations: concurrency is not handled.

Extension of ordinary Petri net is therefore needed, where a transition is enabled when its pre-conditions hold but not post-conditions. However, the boundedness is ensured with a maximum token capacity at each place. Similarly, the ordinary net or the elementary net can't model the time dependent system flows or workflows or delayed time transition flows or uncertainties in transitions. In order to model real time systems, various time extensions are proposed through Petri Nets. Time Petri Net model (TPN) is a powerful formalism and conciliation between modeling power and verification complexity. The timed nets are classified as timed place Petri Nets (TPPN) and timed transition Petri Nets (TTPN) depending on whether the timing bounds annotate places or transitions. TPN are useful for performance evaluation and can be implemented using stochastic or constant timing [3]. The stochastic timing is useful to model events where time is important and all the enabled transitions in the model are equally likely to occur. Constant timed Petri Nets are useful to model time dependent events where transitions occur after some predetermined time [43].

Another dimension in high-level net modeling is Color. The Colored Petri Nets (CPN) associate attributes with the individual token. It is commonly used in network protocol simulation and of course, mathematically well-founded. CPN provides a fitting formalism for the description, construction and analysis of distributed and concurrent

systems [27]. Good models depend on data quality, tools and clear thinking. The non-definitive nature of the large retail networks is perfectly suitable for modeling such systems using the Generalized Stochastic Petri Nets (GSPN). The GSPN supports a unique combination of graphical as well as mathematical formalism to model a dynamic system. Rich mathematical foundations permit in-depth analysis of work-flow nets [13].

A generalized SPN (GSPN) supports both immediate and timed transitions. A GSPN with initial markings can be described uniquely by a 6-tuple: GSPN (P, T, I, O, M, λ), where, (P, T, I, O, M) is a marked PN. A marked Petri Net is a 5-tuple composed of Places; P and Transitions; T. The sets P and T are mutually exclusive. I is the Input function, where the value I (p, t) is the number of directed arcs from the place p to the transition t. O is the output function, where the value O(t, p) is the number of arcs from the transition t to the place p. M is the Initial marking of places, where the value M(p) is the number of tokens that are located in the place p. λ = (λ1, λ2, λ3, … , λn) are marking dependent firing rates associated with transitions. Firing delay is an elapse time associated with every transition.

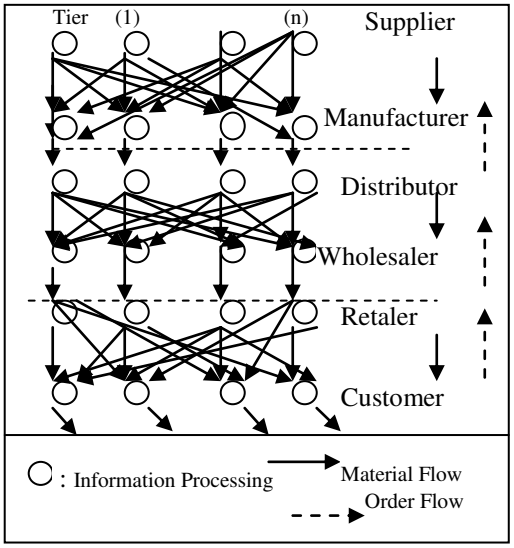


Fig. 1 Multi Stage Multi Tier SC

This delay is a random variable with negative exponential probability density function. For any marking dependent transition ti with associated firing rate λi, can be expressed as λi(mi) and the average firing delay of transition ti in marking mi is [λi (mi)]-1[9].

III. MULTI STAGE MULTI TIER RETAIL CHAIN

There are some frequently used SCM models like SCOR [23]. Epicor a SOA based SCM software, AMT is a SCM ERP application (http://www.softwareadvice.com) may be referred for FlexRFP, which is a web based ERP application.

Figure 1 is the proposed schematic of a multi echelon and multi tier system to deal with multiple products, multiple suppliers in a location independent manner. Managing the inventory in a stochastic natured multi echelon model is a complex process consisting of a set of virtually linked upstream and downstream flows of products, services, finances and information meeting customer's demand.

The supply-chain echelon of our proposed model is composed of customer, retailer, wholesaler, distributor, manufacturer, and supplier. The multiple tiers and its sequence are horizontally presented from left to right at each echelon. Multi stage is presented vertically. Top down flow indicates material flow and bottom up direction represents information flow, e.g., the Beer game is a four Stage model [38], [28]. However, not much discussion is found in the existing literature on multi-tier issues within one supply chain.

This paper aims to model the schematic of figure 1using a GSPN model and reduce or eliminate Bullwhip effect by analyzing the proposed model. The proposed model is simulated using PIPE 2.5 simulation tool.

3.1. GSPN MODEL OF THE RETAIL CHAIN

Let us now model the multi stage multi tier SC using high level net. The GSPN model as shown in figure 2 symbolizes P0, P1, P2 as suppliers, P3, P4, P5 as manufacturers, P6, P7, P8, P9 as distributors, P10, P11, P12, P13, P14 as wholesalers, P15, P16, P17, P18 as Retailers and P19, P20, P21 as customers in tires. {T0 to T18} are the corresponding timed transitions with rate=1.

The net is designed with a constrain that material flow will be to any of the role players in the succeeding layer and information flow will be to its preceding layer only. There is no restriction of tokens at any of the places. The process is initiated with inhibitor arcs from the initial suppliers P0, P1, P2. The PIPE 2.5 tool, used for the GSPN simulation classifies that the proposed is an extended simple type ordinary net as all arc weights are unity.

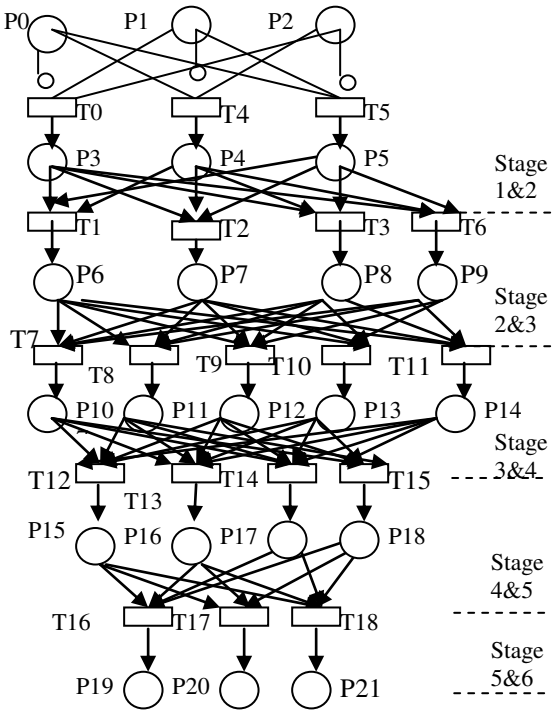


Fig 2: GSPN Model of the Retail Network

The state space analysis pronounces that the GSPN model is not safe. Besides as all the places in the net are not directly connected to each other, so it is not strongly connected. Also there is no live markings and no nonzero safe marking for the proposed model.

Table 1: Token Accumulation after Random Trials

LOC	OB1	OB2	OB3	OB4	Total
P0	6.93	16.83	84.16	94.06	201.98
P1	7.92	17.82	89.11	99.01	213.86
P2	8.91	10.89	54.46	56.44	130.69
Supplier	**23.76**	**45.54**	**227.72**	**249.51**	**546.54**
P3	9.90	12.87	77.23	80.20	80.20
P4	0	5.94	29.70	35.64	35.64
P5	0	7.92	39.60	47.52	47.52
Manufacturer	**9.90**	**26.73**	**146.54**	**163.37**	**346.54**
P6	5.94	9.90	49.50	53.47	53.47
P7	5.94	3.96	19.80	17.82	17.82
P8	1.98	2.97	14.85	15.84	15.84
P9	4.95	6.93	34.65	36.63	36.63
Distributer	**18.81**	**23.76**	**118.81**	**123.76**	**285.15**
P10	4.95	4.95	24.75	24.75	59.41
P11	4.95	4.95	24.75	24.75	59.41
P12	0	1.98	9.90	11.88	23.76
P13	0	1.98	9.90	11.88	23.76
P14	4.95	4.95	9.90	9.90	29.70
Wholesaler	**14.85**	**18.81**	**79.21**	**83.17**	**196.04**
P15	9.90	6.91	19.80	16.81	53.43
P16	3.95	4.94	13.86	14.85	37.60
P17	0	1.98	5.94	7.92	15.84
P18	0	1.98	7.92	9.90	19.80
Retailer	**13.85**	**15.81**	**47.52**	**49.49**	**126.67**
P19	4.95	4.95	4.95	4.95	19.80
P20	4.95	4.95	4.95	4.95	19.80
P21	4.95	3.96	3.96	2.97	15.84
Customer	**14.85**	**13.86**	**13.86**	**12.87**	**55.45**

The GSPN model in figure 2 is not bounded and there can be deadlocks. This is because the number of tokens at any place is a positive integer and each transition gets an output place in the immediate succeeding layer [6], [40].

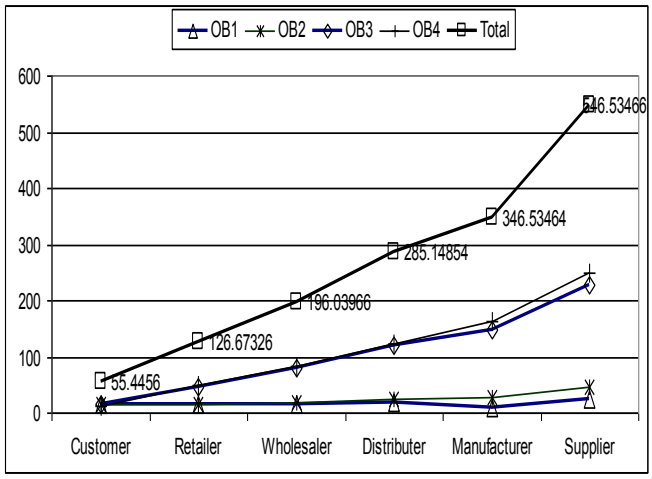


Fig. 3: Simulated Stock Status at different SC Partners

A random number of continuous observations were performed on the GSPN model of figure 2 over four stages OB1 to OB4. These are recorded in Table 1. Figure 3 clearly shows that as the operation progresses, the performance of the system becomes non-linear.

Further, the Bullwhip effect indexes are measured between the pair of communicating role players in the supply chain like distributor and wholesaler, wholesaler and retailer, and retailer and customer. The corresponding bar graphs are presented in figure 4(a) thru 4(c).

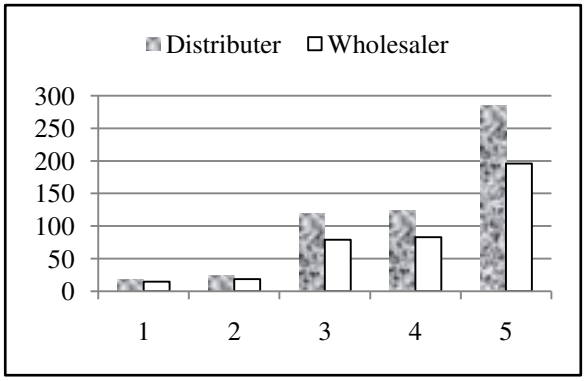


Fig. 4a BWE between distributor and wholesaler

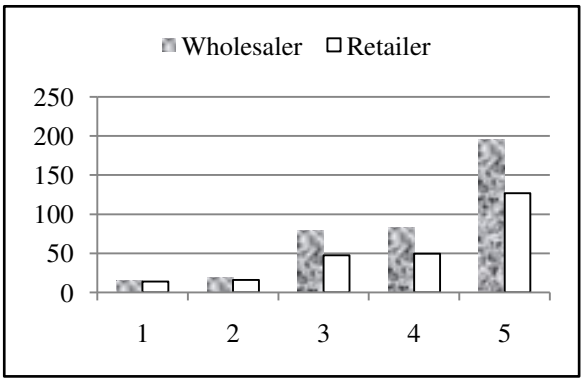


Fig. 4b: BWE between Wholesaler and Retailer

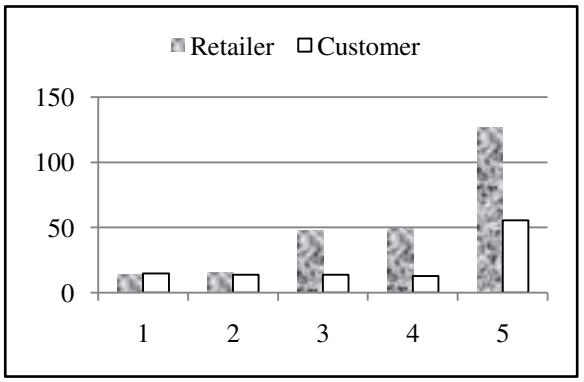


Fig. 4c: BWE between Retailer and Customer

The fact that the BWE index is greater than 1 shows the presence of BWE in the system. Our objective is to model the system in such a way that the BWE should be minimum without affecting the customer's interest. Hence the model demands some major revisions. The revised model is presented in section IV.

IV. THE REVISED GSPN MODEL

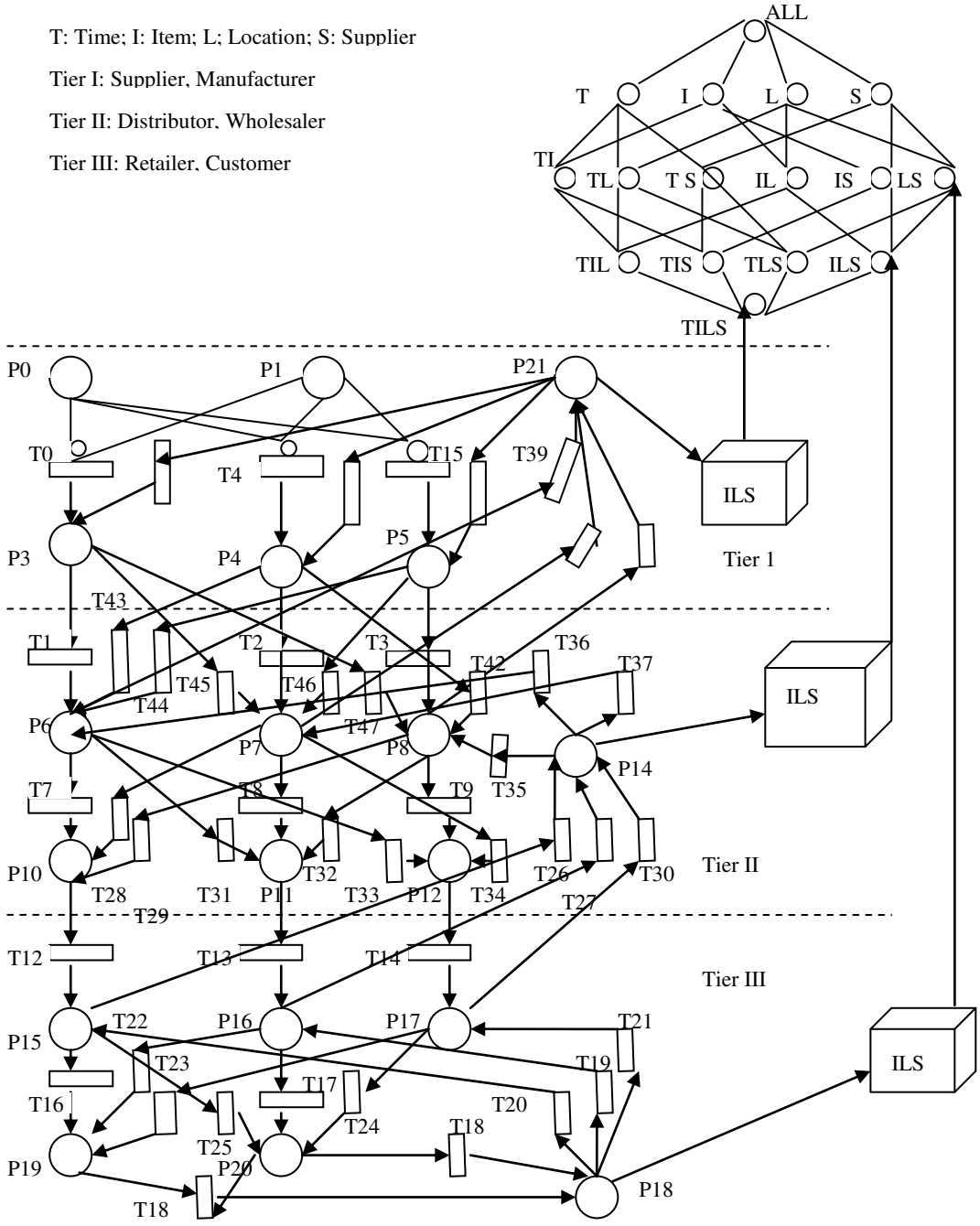


Fig. 5: Revised GSPN model of Retail Network

In this section we introduce an enhanced GSPN model for Retail Network. Its main objective is to remove the non linearity from the SCM process and control the BWE. In order to revise the process, tier wise separate data marts are planned to cater the retailing need of the particular tier. These data marts finally encompass an enterprise warehouse to check for the global retailing needs.

Data marts are subjective and time variant to the respective tiers provides up-to-date requirement need of the immediate succeeding tier and also update the enterprise warehouse for global monitoring and controlling the BWE. With this view we introduce three data marts each at layer 1, layer 2 and at layer3.The success of the retail chains depends on customer relationship and managing the issue the presence of data marts and the corresponding warehouse is inevitable.

With this view the revised model of figure 5 holds two suppliers, three manufacturers, distributors, retailers, and two customers separated over three tires. Each tier holds one data mart with the demand information over the concept hierarchy of item, location, and supplier. The data mart at tier 3 is linked with the server at location P18 dealing with customer and retailer, tier 2 data mart is linked with the server at P19 dealing with distributor and wholesaler, tier 1

data mart is linked with the server at P21 dealing with manufacturer and supplier, and finally the operational data marts of tier I to tier III builds the enterprise warehouse with the concept hierarchy time, item, location, and supplier.

Table 2: Token Accumulation in revised model

LOC	OB1	OB2	Total
P0	21.78218	49.50495	71.28713
P1	38.61386	63.36634	101.9802
Supplier	**60.396**	**112.871**	**173.267**
P3	43.56436	65.34653	108.9109
P4	33.66337	53.46535	87.12872
P5	28.71287	72.27723	100.9901
Manufacturer	**105.941**	**191.089**	**297.03**
P6	29.70297	49.50495	79.20792
P7	31.68317	51.48515	83.16832
P8	26.73267	46.53465	73.26732
Distributor	**88.1188**	**147.525**	**235.644**
P10	28.71287	48.51485	77.22772
P11	35.64356	55.44554	91.0891
P12	25.74257	45.54455	71.28712
Wholesaler	**90.099**	**149.505**	**239.604**
P15	21.78218	41.58416	63.36634
P16	39.60396	59.40594	99.0099
P17	27.72277	47.52475	75.24752
Retailer	**89.1089**	**148.515**	**237.624**
P19	47.52475	67.32673	114.8515
P20	41.58416	61.38614	102.9703
Customer	**89.1089**	**128.713**	**217.822**

The tire wise servers and corresponding data marts are used for communication, material flow and the business intelligence applications. In order to control the entire SC operations and customer relationship management (CRM), some retail chain demands a centralized source of information for effective operation of the chain and the warehouse so built will serve that purpose.

The warehouse server can be placed and maintained in the cloud for efficient distributed application. The properties of the net in figure 5 remains same as that of figure 2 analyzed by the tool [6], [40]. A random number of continuous observations were performed on the high level net of figure 5 over two stages OB1 and OB2 and the data is recorded in table 2.

It clearly shows that with the progresses in operation the non linearity of figure 3 is substantially reduced in figure 6. Further we have measured the BWE index between the pair of communicating role players and presented the results in figure 7(a) – 7(c). The bullwhip effect between distributor and wholesaler as well as between Wholesaler and Retailer are eliminated.

These are demonstrated in figure (7a) and in figure (7b) respectively. However, the BWE between the retailer and customer is reduced but could not be completely eliminated as shown in figure (7c). Table 2 reflects that BWE index between other pairs of operators like supplier-manufacturer-distributor or wholesaler-distributor, in most of the cases are less than unity.

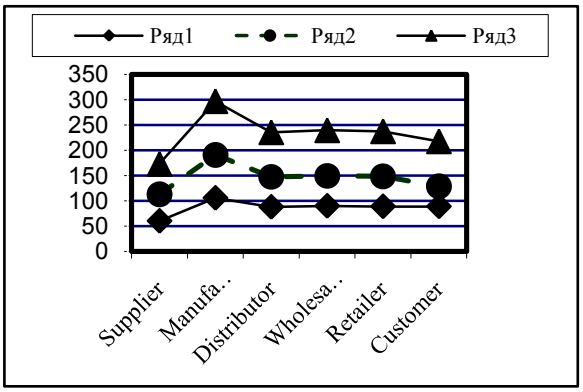


Fig. 6: Token Accumulation for the Revised Model

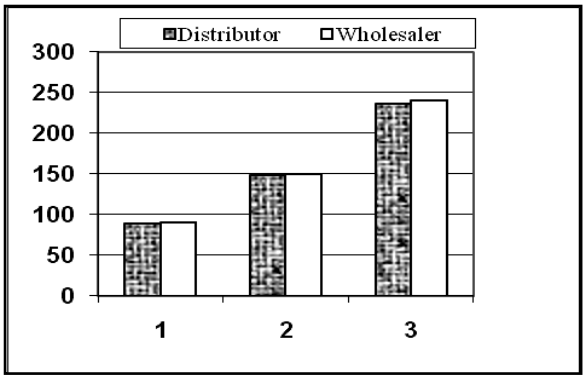


Fig. 7a: Revised BWE for Distributor and Wholesaler

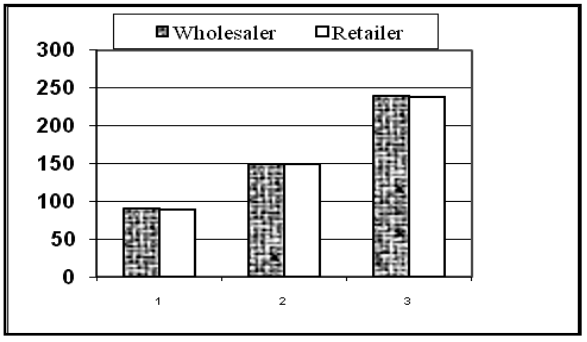


Fig. 7b: Revised BWE for Wholesaler and Retailer

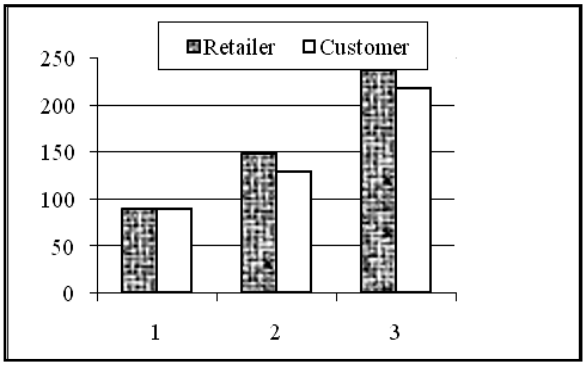


Fig. 7c: Revised BWE for Retailer and Customer

The empirical observations were random and the outcomes are tabulated to reflect the facts that BWE has been reduced due to the inclusion of the data mart in the form of a server to share the information and material flow. However, some existence of BWE may be needed on a player to player basis as a part of CRM strategies. In section V, we discuss some of the derived advantages of the warehouse.

V. WAREHOUSE APPLICATIONS

CRM is the key to success of the retailing business. Some of the CRM queries that can be generated from the multi dimensional data cube are:

- Which distribution channels contribute the greatest revenue and gross margin?
- Which customers are most profitable based upon gross margin and revenue?
- How many unique customers are purchasing this year compared to last year?

The first query generates the total amount sold and the second query checks for the location in which maximum sales has taken place. There could be many such queries can be generated by adding customer as one of the dimension of the warehouse, which will be an added overhead to the warehouse. Instead, local data marts can be used to generate more interesting queries for efficient CRM operations. The objective of the lattice of figure 5 is to carry out online analytical processing (OLAP) of the multi echelon SC. The local OLAP operations can be performed from the data marts of the respective tiers. The data warehouse model can be used to find the most valuable customer through Recent access, Frequency, and Monetary value (RFM) for CRM activities defined as a strategy to enable the organization proactive and profitable. It helps the organizations to have right focus and allocate sufficient resources to where is needed [14]. The star schema of the model for query processing is presented in figure 8.

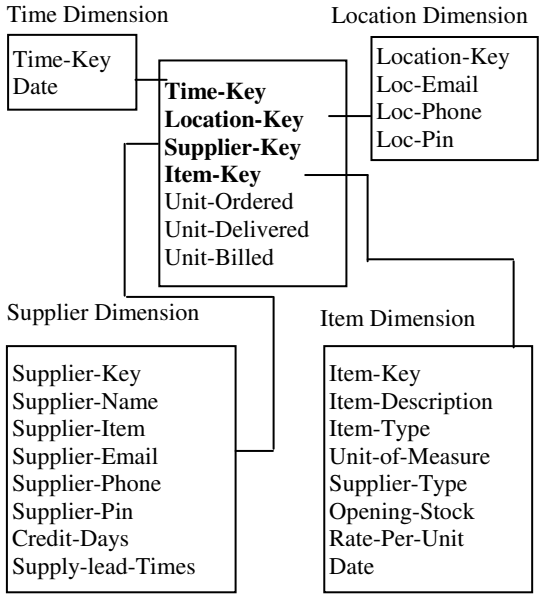


Fig 8: Star schema of the Lattice of figure 5

VI. CONCLUSION

Autonomous forecasting by the Supply Chain players invites BWE. Information visibility and stage wise centralized repositories reduce the BWE. The proposed model is capable of handling demand driven real time activities among the SCM partners. Operational soundness can be pioneered with the augmentation of multi dimensional data cubes and a central warehouse, such components helps in real time CRM applications. Initially a multi echelon, multi-tier retail network has been proposed in section 3 and after a set of simulation run, it is observed that inventory is getting piled along the SC partners, which provokes risks of uncertainty of demand and supply and obsolescence of product. In figure 5, the revised GSPN model is proposed with a demand driven procurement strategy along with stage wise data marts and a central warehouse to avoid the effect of demand variability and real time CRM application.

The proposed virtual multi-tier retail network is collaborative, data intensive and distributed. The large repository will support real time demand analysis, market forecast and strategic decision making. From the operational view point there are only a few global large size retailers who may be interested in maintaining the integrated chain but for the sake of scalability of the proposed model stage wise data marts are also employed for medium and small size retailers. It will not be out of place to mention that ERP based SCM chains can handle transactions and control BWE but beyond transaction management is CRM as SCM starts with the customer and ends with the customer and our proposed model is capable of handling real time CRM operations.

REFERENCES

[1] Agrawal, S., Sengupta, R, N., & Shanker, K., (2009), "Impact of information sharing and lead time on bullwhip effect and on-hand inventory", European Journal of Operational Research,vol:192(2), pp:576-593

[2] Aigbedo, H., & Tanniru, M., (2004), "Electronic Markets in Support of Procurement Processes along the Automotive Supply Chain", Journal of Production Planning & Control,vol:15(7),pp: 688-695

[3] Ajmone-Marsan, M., Conte, G., & Balbo, G.,(1984), "A Class of generalized Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems", ACM Transactions on Computer Systems,vol:2(2), pp:93-122

[4] Alvarado, Karla. & Rabelo, Luis. P., (2008), "Stakeholder value mapping framework for Supply Chain improvement when implementing IT solutions", Proceedings of the Industrial Engineering Research Conference, Vancouver, vol:24(3),pp:24-31

[5] Anderson (Jr), E. G.,& Morrice, D. J.,(2000), "A simulation game for service-oriented supply chain management: Does information sharing help managers with service capacity decisions. Production and Operations Management", International Journal of Logistics Management, vol: 9(1), pp: 40–55

[6] Bonet, P., (2007), A Petri Net Tool for Performance Modelling, PIPE (CLEI), vol. 2 (5).

[7] Cachon, G. P., Randall, T., & Schmidt, G.M., (2007), "In Search of the Bullwhip Effect Manufacturing & Service", Journal of Operations Management, vol:9(4), pp:457–479.

[8] Centeno, Martha, A., Pérez, J. E.,(2008), "Quantifying the Bullwhip Effect in the Supply Chain of small-sized companies", Sixth LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2008)

[9] Chaki, Nabendu, & Bhattacharya, S., (2006), "Performance analysis of

multistage interconnection networks with a new high-level net model", Journal of Systems Architecture, Elsevier, vol. 52(1), pp:56-70.
[10] Chaki, Nabendu.,& Sarkar, B. B., (2010), "Virtual Data Warehouse Modeling Using Petri Nets for Distributed Decision Making", Journal of Convergence Information Technology (JCIT), vol:5(5),pp:8-21.
[11] Chen, F., Drezner, Z., Ryan, J.K., and Simchi-Levi, D.,(2000a), "Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information", Management Science, vol. 4(3), pp. 436-443.
[12] Chen, L. and Lee., H.L., (2009). "Bullwhip effect measurement and its implications", Duke university Working Papers Series (Currently under revision for Management Science).
[13] Corman, Thomas. H., Leiserson, Charles. E.,& Rivest, Ronald. L., (2003), "Introduction to Algorithms", 2nd edition, PHI Publications, chapter 34.
[14] Cunningham, Colleen, et al. (2003), "Design and Research Implications of Customer relationship Management on Data Warehousing and CRM Decisions". In Proceedings of the 2003 Information Resources Management Association International Conference (IRMA 2003), pp:82–85.
[15] Dejonckheere, J.,Disney, S,M., M.R Lambrecht,& Towill ,D.R., (2004), "The impact of information enrichment on the Bullwhip effect in supply chains: A control engineering perspective", European Journal of Operational Research, Vol:153(3), Pages 727–750.
[16]. Disney, S. M. & Towill, D.R., (2003), "Vendor-managed inventory and bullwhip reduction in a two-level supply chain", International Journal of Operation and Production Management, vol:23(6), pp:625-651.
[17] Edmund, Clarke. (2001), "Bounded Model Checking Using Satisfiability Solving", Journal of Formal Methods in System Design, vol:19(1),pp:7–34.
[18] Eleonora, Bottani. & Roberto, Montanari. (2010), "Supply Chain Design and Cost Analysis through simulation", International, Journal of Production Research, vol:48(10), pp. 2859-2886.
[19] Fransoo, Jan. C., & Wouters-Marc, J. F., (2000), "Measuring the bullwhip effect in the supply chain", Supply Chain Management: An International Journal, vol:5(2),pp:78–89.
[20] Hau, L. Lee., Padmanabhan, V., & Whang, Seungjin. (1997), "The Bullwhip Effect in Supply Chains", Sloan Management Review, vol. 38(3), pp: 93-102.
[21] Hohmann, Susanne., & Zelewski, Stephan, (2011), "Effects of Vendor-Managed Inventory on the Bullwhip Effect", International Journal of Information Systems and Supply Chain Management (IJISSCM), vol:4(3), pp:1-17.
[22] Hugos, Michael. H., (2006), "Book Title:Basic Concepts of Supply Chain Management", Wiley& Sons, 2nd Edition.
[23] Huan, S., Sheoran, S. & Wang, G. (2004). "A review and analysis of supply chain operations reference (SCOR) model", Supply Chain Management: an International Journal ,vol:9 (1), pp:23-29.
[24] Huhns, M. N., & Larry, M. Stephens, (2001), "Automating Supply Chains", IEEE Internet Computing, vol. 5(5), pp. 92-95.
[25] Hussain, Matloub. Paul, R. Drake., & Dong, M. Lee., (2007), "Quantifying the Impact of a Supply Chain's Design Parameters on the Bullwhip Effect", 7th Global IEE Conference on Business & Economics, Italy, pp. 210 – 213.
[26] Javier Esparza, & Mogens Nielsen, (1994), "Decidability Issues for Petri Nets", Basic Research in Computer Science, BRICS Report Series RS948
[27] Jensen, K., (1992), "Colored Petri Nets – Basic Concepts, Analysis Methods and Practical Use: Basic Concepts", Springer-Verlag, London, 1996, vol. 1 pp. 234,
[28] Kopczak, L.,& Johnson, M., (2003), "The Supply Chain Management Effect", MIT Sloan management Review, vol. 44(3)
[29] Krogstie John., (2001), "Using a semiotic framework to evaluate UML for the development of models of high quality", Idea Group Publishing, Chapter 5, (pp.:89-106)
[30] Lewlyn, L., Rodrigues, R., Hebbar, Sunith., & Herle, Ramdev., (2011), "Bullwhip Effect Mitigation in Trading System: A System Dynamics Approach", Proceedings of the World Congress on Engineering, UK, Vol. I, ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)
[31] Nagaraju. (2009), "Indian e-Commerce Industry, e-Commerce Information Growth of Retail e-Commerce in India & World Wide", Smart e-commerce: e-Business Information, A Blog from Embitel Technologies India Pvt. Ltd., source: techcrunchies.com, pp. 102–105.
[32] Obeidat, R., & Zaatreh, Z., (2010), "Motivation for Integrating Supply Chains Using Service Oriented Architecture Approach", Proceedings of NGMAST. 2010, 102-105.
[33] Power, Damien., (2005), "Supply Chain management integration and implementation, a literature review", Supply Chain Management An International Journal, vol:10(4), pp. 252–263.
[34] Puigjaner, L. and Lainez, J.M., (2008), "Capturing dynamics in integrated supply chain management", Computers and Chemical Engineering, vol:32 (11), 2582–2605
[35] Ralph, F. Wilson.,& Pettijohn, James. B., (2006), "E-Mail Marketing Software, The State of the Art", Journal of Digital Business, vol. 1(1), pp.75-94.
[36] Sarode, A. D., Sunnapwar, V. K.,& Khodke, P. M., (2008), "A Literature Review for Identification of Performance Measures for Establishing a framework for Performance Measurement in Supply Chains", International Journal of Applied Management and Technology (IJAMT), vol:6(3),pp:241-273
[37] Shukla, V., (2009), "Bullwhip and backlash in supply pipelines", International Journal of Production Research, vol:47(23),pp:6477– 6497
[38] Sterman, John. D., (1989), "Modeling Managerial Behaviour: Misperceptions of Feedback in Dynamic Decision Making Experiment", Joutnal of Management Science, vol:35(3),pp:321-339
[39] Susan, A. Sherer., (2010), "Enterprise Applications for Supply Chain Management", International Journal of Information Systems and Supply Chain Management (IJISSCM), vol:3(3), pp:18-28
[40] Tado Murata, "Petri Nets : Properties, Analysis and Applications", Proceedings of the IEEE, Volume 77, No .4, April 1989, Page:541-580, ISBN: 0-387-13723
[41] Top 10 Most Recommended Systems, Supply Chain Management Software, http://www.softwareadvice.com/
[42] Wang, Eric., (2006), "A Virtual Integration Theory of Improved Supply Chain Performance", Journal of Management Information Systems, vol:23(2),pp:41-66
[43] Wang, Jiacun, (1998), "Timed Petri Nets: Theory and application, chapter 1 & 2", Kluwer Academic Publishers, vol:3(32),pp:1-134
[44] Wei, Shi., (2011), "Dynamic Consumer Decision Making Process in E-Commerce", Digital Repository at the University of Maryland, http://hdl.handle.net/1903/11944
[45] WANG, Weipeng. (2011), "Analysis of Bullwhip effects in Perishable Product Supply Chain Based on System Dynamics Model", IEEE, International Conference on Intelligent Computation Technology and Automation, vol: 1, pp: 1018 – 1021
[46] Yan, Hong., (2003), "Strategic Model for Supply Chain Design with Logical Constraints: Formulation and Solution", Computers and Operations Research, Elsevier Science, vol:30(14), pp:2135-2155
[47] Zhuoqun, LI., (2011), "A Multi-stages and Tree-shape Supply Chain Model Based Simulation System for Bullwhip Effect", Journal of Computational Information Systems, vol:7(1),pp: 281-289.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1153–1156

The postulates of consensus determining in financial decision support systems

Jadwiga Sobieska-Karpińska
Wrocław University of Economics ul. Komandorska 118/120, 53-345 Wrocław, Poland
Email: jadwiga.sobieska-karpinska@ue.wroc.pl

Marcin Hernes
Wrocław University of Economics
ul. Komandorska 118/120, 53-345 Wrocław, Poland
Email: marcin.hernes@ue.wroc.pl

***Abstract*—This article presents the problem of consensus determining postulates defining in financial decision support systems. The consensus determining methods and function is characterized in the first part. Next the general postulates for consensus estimation and their characteristics are presented. The final part of article suggest new postulates pertaining to financial decisions, and the possibility of their use in practical solutions. The application of these postulates, as a consequence, can lead to the process of making financial decisions will be more flexible, and the risk involved in financial decisions will be significantly reduced.**

I. Introduction

MAKING decisions in financial matters has become a key component of any business activity. Problems in this area are typically associated with highly volatile character of financial market [9]. Decisions must be made virtually in real time, since only prompt and accurate reaction to changing market conditions provides tangible benefits, for example high return rate. Another important determinant is the high level of risk involved in financial decisions. Since analysis of information and drawing valid conclusions is a time-consuming process, and since real-time computing is beyond human processing capabilities, the process of making financial decisions is typically supported by computer software, employing a range of computing methods, such as artificial intelligence systems, capable of identifying relevant information and drawing conclusions based on input data. Important area of development in recent years is the use of agent and multi-agent systems [10,18] – these, unlike other AI systems, offer the capability of unaided operation and unaided decision-making, i.e. without user input and irrespective of any external factors.

At present, financial decision support systems (DSS) are typically distributed [2]. These systems offer the potential of fast processing of large amount of data. However, in most cases, distributed systems used for support of financial decision-making processes tend to generate multiple variants of solutions, which may result in knowledge conflict within the system. For example, in multi-agent systems, each individual agent may utilize a different method of decision support and, consequently, arrive at a different solution. Users expect a unified variant – or, to put it in simple terms, they want a single decision. Decision-making process is followed by implementation, and only one decision can be implemented at any given time – ideally, one that will bring tangible benefit to the user, while simultaneously limiting the level of risk involved. If a decision support system generates multiple variants, users face the problem of selecting the best possible variant – the ultimate decision. Since the task of selecting the best possible variant, as already mentioned, should ideally be realized in (or close to) real time, it is expected that the DSS will automatically present a single variant that offers best possible results for the user, thus solving the knowledge conflict. Professional literature presents a wealth of methods that can be used to this effect, such as negotiation methods [4] and deductive computing methods [1]. Negotiation methods allow for determining a solution that best suits all parties involved, based on compromise, but it is burdened with the problem of mass exchange of information between system components, which makes the postulate of real-time computing particularly difficult to achieve – or even impossible. On the other hand, deductive methods of computing (such as those based on game theory, classical mechanics and selection methods) offer high computing power, but do not easily satisfy the requirement of identifying best possible variant with simultaneously limiting the risk of inadequate selection.

It seems that the above inconveniences (and the resulting knowledge conflict) can be resolved with the use of consensus methods [8,15]. Consensus methods offer the benefit of determining a single best variant (or a single decision, in this context) out of multiple possible variants. It must be noted that the single decision determined using consensus methods will not necessarily belong to the domain of variants generated by the system in the first place. This is because consensus methods take into consideration all conflicting parties and interests. The ultimate decision is endorsed by all modules (parties) and the decision represents interests of all parties to a degree that satisfies all conflicting parties.

Sobieska-Karpińska and Hernes [10, 5] argue that using consensus methods for the purpose of identifying and presenting a target solution to the user will offer a reduction of decision-making time, since users are not burdened with the task of analysing and selecting the best possible variant. It also reduces the risk involved in the process, since variants identified and selected by the user may fail to bring the expected benefit, or even result in a loss.

978-83-60810-53-8/$25.00

It must be noted, however, that consensus algorithms used in DSS systems, including systems for financial decision support, must satisfy certain consensus postulates. These postulates represent conditions to be met by consensus-calculating functions. Only proper definition of these conditions will ensure that decisions made with the help of consensus algorithms will bring tangible benefit to the user.

Professional literature (see, e.g. [1,13,14]) does provide some general (universal) postulates for consensus estimation, but those assumptions fail to take into account some important aspects of financial decision-making, such as the risk and uncertainty involved. Therefore, it seems necessary to broaden the list of postulated parameters.

The purpose of this paper is to present the general postulates for consensus estimation (that need to be included, regardless of the problem they are meant to address via consensus calculation) and their characteristics, as well as suggest new postulates pertaining to financial decisions. This will allow for more accurate construction of consensus algorithms and, consequently, development of IT solutions able to calculate consensus results automatically, based on a set of solutions generated by the system. In this approach, the system will present the user with one ultimate decision that may be implemented to best effect. Consequently, the process of making financial decisions will be more flexible, since the system will suggest the most appropriate solution in (or close to) real-time. In addition, the risk involved in financial decisions will be significantly reduced, since users will not be able to manually select a decision that may be burdened with such risk at implementation phase.

II. General Postulates Of Consensus Determining

Purpose of introduction the postulates is determination on their bases classes of functions of consensus or otherwise saying, different methods of a consensus determining.

In addition, because the postulates are conditions which are expected to meet on consensus function, you can get it to justify the use of these functions in practice.

In farthest part of article we will use following symbols:

$\Gamma(U)$ - set of all don't empty subsets of universe U (e.g. set of objects-financial instruments),

$\Gamma'(U)$ - set of all don't empty subsets with repetitions of universe U,

$\cup'$ – sum of set with repetitions.

Let $X, X_1, X_2 \in \Gamma'(U)$, $x \in U$. In farthest part of article we will use next parameters:

$o(x,X)=\sum_{y\in X} o(x,y)$,

$o^n(x,X)=\sum_{y\in X}[o(x,y)]^n$ for $n\in \mathrm{N}$.

Let's notice, that parameter $o(x,X)$ represents sum of distance from element x belongs to universe U for elements of profiles X , but largeness $o^n(x,X)$ represents sum of n-powers it distance. This value can be interpreted as measure of evenness of distance from element x for elements of profiles (eg. det of financial decisions) X. if value n is greatest memorial then n, distances are more even.

In work [15] consensus function is defined next:

Definition 1.

Consensus function at space (U,o) we call optional functions of forms:

$$c: \Gamma'(U) \rightarrow \Gamma(U). \quad (1)$$

For profile $X\in\Gamma'(U)$ each of elements set $c(X)$ we call his *consensus,* however all set $c(X)$ we call *representation of profile X.* . Let C is set of all consensus functions in a space *(U, o).*

Using the overall function of the consensus you can then define the more detailed class consensus functions, relating to the various methods of its determination, including [15]:

a) Constructive methods, rely on solving problem of consensus on two levels: microstructures and macrostructures universe U. Microstructure is a structure of elements U, macrostructure is structure of universe U.
b) Optimizing methods, rely on defining function of consensus behind assistance of optimizing rules. Often in this methods functions quasi-mediane are applying, consensus is most approximated for all solutions from which be appointed, distances of consensuses are even for individual solutions simultaneously.
c) Methods taking advantage bool conclude, rely in the form encoding problem of consensus in bool formula to such manner that each first implicant this formula appoints solution of problem.

Therefore, in order to define the classes of functions relating to the above methods, it can use the postulates for consensus defined as follows (on the basis of [1, 13, 14]):

Definition 2.

Let X is optional profile we say, that consensus function $c\in C$ grants postulate:

1. *Reliability (Re), if*
$$C(X)\neq\emptyset \quad (2)$$
2. *Consistency (Co), if*
$$(x\in C(x))\Rightarrow(x\in c(X\cup' \{x\})) \quad (3)$$
3. *Quasi-unanimous (Qu), if*
$$(x\notin C(x))\Rightarrow((\exists n\in \mathrm{N})x\in c(X\cup' \{n*x\})) \quad (4)$$
4. *Proportional(Pr), if*
$$(X_1\subseteq X_2\wedge x\in c(X_1)\wedge y\in c(X_2))\Rightarrow(o(x,X_1)\leq o(y,X_2) \quad (5)$$
5. *1-Optymality (O_1), if*
$$(x\in C(x))\Rightarrow(o(x,X)=\min_{y\in U} o(y,X)) \quad (6)$$
6. *2-Optymality (O_2), if*
$$(x\in C(x))\Rightarrow(o^2(x,X)=\min_{y\in U} o^2(y,X)). \quad (7)$$

These postulates for function of consensus express primary condition define different method consensus. First postulate (*reliability*) sets up, that it is possible to appoint consensus for each profile always. It answers optimistic attitude each conflict give solve. Reliability is known criterion in theory of choice [3].

Postulate *consistency* requires implementation of condition, that if some element x is consensus for profile X , then after expansion this profile about x ($X\cup' \{x\}$), this element should be consensus for new profile. Consistency is important ownership of consensus, because it allows users to forecast behavior of rule of appointment of consensus, when premises of independent choices are jointed.

According to postulate *quasi-unanimous,* if certain element x is not consensus for profile X, that it will be consensus for profile X^1 inclusive X and n protrude element x for certain n. In other words, each of elements of universe U should be chosen as consensus for such profile, if number of its pronouncement is sufficiently big.

Proportionality postulate is natural ownership enough, because if profile is greatest memorial then difference between its elements and consensus is greatest.

Last two postulates are very particular. First of it, postulate *1-Optymality* require that consensus is nearest (most similar) to elements of profile. This postulate, in literature very well known, it defines concrete function class, called medians. Instead postulate *2-Optymality*, on the other hand, requires, in order to sum of square of distance from consensus for elements of profiles was smallest. Cause of introduction of this postulate results from (also very natural) following condition concerning determination function consensus: consensus have to be „fair"; it means, that its distance for elements of profiles should be the most even. Let's notice, that number $o^n(x,X)$ defined earlier, can be treated as measure of evenness of distance between certain object x and elements of profiles X. Therefore, above-mentioned condition requires, in order to value o^n (consensus, X) be minimal. In work [6, 7] show, that functions granting postulate *2-optymality* are better than function granting postulate *1-optymality*, by the reason of greatest evenness, but they differ from other function of consensus greatest similarity for elements of profiles. From it result, that postulate *2-optymality* is good criterion of appointment of consensus.

Let us note that the first three postulates, namely *Re*, *Co* and *Qu*, are independent of the structure of universe U, represented by a distance function o (used to establish consensus function class in methods based on Boolean reasoning), while the last three postulates (*Pr*, O_1 and O_2) are formulated on the basis of o function (these postulates are employed in optimization methods). Postulates *Re*, *Co* and *Qu* are also used in cases when distance function (or, in more general terms – the macro-structure) for universe U cannot be specified. For financial decisions, function of distance can always be reliably defined, therefore all postulates can be employed, allowing for the use of both constructive and optimization methods of consensus estimation.

The above general postulates of consensus estimation, as already mentioned, are not sufficient for financial purposes. For this reason, this author puts forward two additional postulates to supplement the above list.

III. The Proposal To Extend The List Of Postulates In Terms Of Making Financial Decisions

A good approach in estimating best possible decisions in financial matters, i.e. when dealing with problems typically burdened with risk and uncertainty, is to employ evenly distributed consensus – that is, one that takes into account all possible solutions, with each solution estimated at equal measure. This helps minimize the risk of ultimate decision, since the potential of putting more weight to an incorrect decision is eliminated. Therefore, if *2-Optimality* postulate offers more even distribution that *1-Optimality* postulate, then a postulate of *n-Optimality* should be defined, to offer even smoother distribution than *2-Optimality* for n>2. Consequently, definition for such new postulate will take the following form:

Definition 3.

The consensus function $c \in C$ grants an n-Optimality postulate (O_n), if

$$(x \in C(x)) \Rightarrow (o^n(x,X) = \min_{y \in U} o^n(y,X)). \quad (8)$$

This postulate is a generalization postulates *1-Optimality* and *2-Optimality*.

Another extended postulate on consensus determining in financial decision support systems is a inconsistency of knowledge postulate:

Definition 4.

Consensus function $c \in C$ grants an inconsistency of knowledge postulate (Uk), if

$$(x \in C(x)) \Rightarrow (o^n(x,X) > \min_{y \in U} o^n(\{X \setminus y\},X)). \quad (9)$$

The above postulate allows for determination of an element, for which the distance to consensus is larger than the sum of consensus distances of all the remaining elements (in other words, one of the profile elements is markedly more distant from the consensus than the others). Such situation may result from inadequate knowledge on the part of one of the conflicting parties (for example, a software agent). If this is the case, then decisions generated by the defaulting party should not be taken into account. This problem may be solved by adopting a multi-stage process of consensus estimation, as suggested in [16]. Such method is under implementing in a-Trader multiagent system for FOREX platform [11]. These systems consist of the large number (hundreds) of processing agents, which on the basis of the FOREX signals, take the specified decision on buy/sale. The paper [12] presents using the consensus methods to reduce the level of the investment risk, as a strategy of Supervisor Agent in a-Trader System.

Building consensus algorithms on the above postulates is not necessarily a guarantee of success in arriving at best possible solution. For example, the algorithms may find a consensus solution for which a given element of the decision-making process is at the same time adopted and rejected, leading to a contradiction. Some authors also take into account the profile's consensus susceptibility [8]. If a profile (a set of decisions) is not prone to consensus, methods of satisfying its susceptibility may be adopted, such as inclusion of decisions generated by new parties (e.g. new software agents).

However, it must be noted that consensus estimation functions used for the purpose of supporting financial decisions must meet both general consensus postulates and the expanded postulates (these postulates may also be used in order to the other, than only financial problems). Otherwise, the estimated consensus may not warrant tangible benefit to the user, for example – by placing more weight on an inappropriate decision contained in the profile.

IV. Conclusion

Making decisions in financial matters is a complicated process, particularly in the face of high risk and uncertainty associated with this form of activity, since it may lead to unpredictable results. Improper decisions may detriment the functioning of a whole organisation. Distributed systems offering support for financial decisions are a viable solution, provided that they are able to generate a single, reliable recommendation. However, if individual nodes of the system (such as software agents) generate multiple instances of solutions, the overall reliability of the system is considerably lower. Therefore, proper care should be taken to ensure that the user receives the best possible solution generated automatically by the system, so that he or she can make a correct decision that will result in benefit for the organization. Use of consensus methods provides the potential of arriving at a single best decision – one that not necessarily belongs to the original domain of decisions generated by individual nodes, but adequately similar. Consequently, the level of risk involved is considerably lower. If users were to perform own analyses and manually select from decisions generated by the system under time pressure, their choices would be potentially burdened with error – the more so if we take into account the time pressure involved. In addition, consensus methods allow for considerable reduction of decision time, since the system presents the user with a single best solution determined automatically on the basis of variants generated by individual nodes.

For obvious reasons, consensus methods do not warrant absolute accuracy of resultant decision, but they do warrant some degree of satisfaction. Some of the individual variants generated by system nodes may prove more appropriate than the automatic suggestion determined using consensus methods, but one can never be certain that the user would have selected such best variant (if he were to analyse and select it manually). In such a case, selecting the worst possible variant is also possible, which can only increase the risk involved.

However, correct algorithms for consensus estimation should incorporate and take into account all the postulates presented above. Negligence in this respect may result in ill-advised suggestions with negative consequences.

Proper implementation of consensus postulates is, therefore, a prerequisite for correct design of consensus algorithms to be used in financial DSS (decision support systems).

References

[1] Barthlemy J.P., Janowitz M.F., "A formal theory of consensus", in: Siam J., *Discrete math* 4, 1991.

[2] Coulouris G., Dollimore J., Kindberg T., *Distributed system. Concept and Design*, WNT, 1998.

[3] Daniłowicz C., Nguyen N.T. , *Metody wyboru reprezentacji podziałów i pokryć uporządkowanych*, Oficyna Wydawnicza PWr, Wrocław 1992.

[4] Dyk P., Lenar M., "Applying negotiation methods to resolve conflicts in multi-agent environments", in: *Multimedia and Network Information systems*, MISSI 2006 Zgrzywa A. (red.), Oficyna Wydawnicza PWr, Wrocław 2006

[5] Hernes M., "Weryfikacja metod consensusu w wieloagentowym systemie wspomagania decyzji finansowych", *Business Informatics* 22-2011 Prace Naukowe UE we Wrocławiu, Wydawnictwo UE we Wrocławiu Wrocław 2011.

[6] Hernes M., Nguyen N.T., *Deriving Consensus for Incomplete Ordered Partitions*. in: Nguyen N. T.(ed.), *Intelligent Technologies for Inconsistent Knowledge Processing*. Advanced Knowledge International, Australia 2004.

[7] Hernes M., Nguyen N.T., "Deriving Consensus for Hierarchical Incomplete Ordered Partitions and Coverings", *Journal of Universal Computer Science* 13(2), 317-328, 2007.

[8] Hernes M., Sobieska-Karpińska J., "Susceptibility to consensus of conflict situation in intelligent multi-agent decision support system", in: Kubiak B.F., Korowicki A. (ed.), *Information Management*, Gdansk University Press, Gdańsk 2009.

[9] Jajuga K., *Podstawy inwestowania na Giełdzie Papierów Wartosciowych*, Giełda Papierów Wartościowych S.A., Warszawa 2007.

[10] Korczak J., Lipiński P. "Agents systems in capital market decision support", in: Stanek S. , Sroka H. , Paprzycki M., Ganzha M. (ed.), *Evolution multiagent information systems in socialy-economic environment*, Placet Press, Warszawa 2008.

[11] Korczak J., Bac M., Drelczuk K., Fafuła A., "A-Trader – Consulting Agent Platform for Stock Exchange Gamblers", in: *Proceedings of Federated Conference Computer Science and Information Systems* (FedCSIS), Wrocław, 2012.

[12] Korczak J., Hernes M., Bac M., "Risk avoiding strategy in multi-agent trading system", in: *Proceedings of Federated Conference Computer Science and Information Systems* (FedCSIS), Kraków, 2013 (in press).

[13] McMorris F.R., Mulder H.M., Powers R.C., "The median function on median graphs and semilattices", *Discrete Applied Mathematics* 101, 2000.

[14] Nguyen N.T. , *Metody wyboru consensusu i ich zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych*, Oficyna Wydawnicza Politechniki Wrocławskiej, 2002.

[15] Sobieska-Karpińska J., Hernes M., "Determining consensus in distributed computer decision support system", in: Dziechciarz J. (red.), *Ekonometria. Zastosowania metod ilościowych*, nr 31, Wydawnictwo UE we Wrocławiu, Wrocław 2011.

[16] Sobieska-Karpińska J., Hernes M., "Consensus determining algorithm in multiagent decision support system with taking into consideration improving agent's knowledge", *Proceedings of Federated Conference Computer Science and Information Systems* (FedCSIS), Wrocław 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1157–1164

The DDMKCC Decision Support Architecture in the Light of Case Studies

Stanislaw Stanek
General Tadeusz Kosciuszko Military Academy of Land Forces 51-150 Wroclaw ul. Czajkowskiego 109 POLAND
s.stanek@wso.wroc.pl

Jolanta Wartini Twardowska
Uniwersytet Ekonomiczny 42-287 Katowice ul. 1 Maja 50 POLAND
j.wartini_twardowska@ue.katowice.pl

Zbigniew Twardowski
CONSORG S.A. 41-506 Chorzów Al. Bojowników o Wolność i Demokrację 38 POLAND
ztwardowski@consorg.pl

***Abstract*—What makes the development of decision support systems (DSS) particularly challenging is the change dynamics of the design space, the instability of initial specifications, and the lack of an adequate model of the decision making process. Facing these, one can appreciate a methodology that can drive the designer's creative effort within a particular decision context. The paper aims to outline the origin and the evolution of research on the DSS architecture commenced by Sprague and Carlson and carried on under the auspices of the International Federation for Information Processing (IFIP) and the International Society for Decision Support Systems (ISDSS)[1]. In particular, the paper presents insights, findings, recommendations and conclusions derived from case studies conducted in domestic medium-sized and large enterprises.**

I. Introduction

GROUND-breaking studies that gave rise to the computerized decision support strand appeared around mid-20th century[2]. By mid-1980s, a new science discipline emerged, concentrating a vast research potential. The term "decision support system" (DSS) has a number of connotations. Most scholars have accepted the term to mean "an interactive computer based system that helps decision-makers use data and models to solve ill-structured, unstructured or semi-structured problems" [1], although some argue that this definition is too narrow, pointing out that a DSS should be able to support ill-structured decisions as well as structured tasks. This leads to a more general definition ([4], [2]) as an interactive computer-based system or sub-system intended to help decision makers use communication technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks and make decisions. In his seminal doctoral thesis, Steve Alter [3] put down the following three axioms of the unfolding paradigm which were approved by most of his fellow researchers: (1) that DSS are designed specifically to facilitate decision processes; (2) that DSS should support rather than automate decision making; and (3) that DSS should be able to respond quickly to the changing needs of decision makers.

However, neither theory-oriented research approaching decision support from the management science perspective nor the few experimental studies have been able to lay solid foundations for DSS designers to build upon. The following research outcomes have proved useful to developers:

- Recommendations concerning the architecture of computer decision support based on the Data – Dialog – Modeling paradigm. The principal idea behind the paradigm represents that the designer's responsibility in designing a decision support system is to build the data, dialog and modeling components and to ensure interaction among these. The idea underpins the theoretical work referenced above, e.g. in adopting a DSS definition. The paradigm has made it possible for Sprague and Carlson [1] to articulate an influential architectural model and helped their followers ([5], [4], [22], [12]) further advance it and extend it.
- The guidelines of "meta-design" methodology pioneered by Moore and Chang [6]: "(…) the classic MIS development life-cycle approach is insufficient as a prescriptive guide for building DSS [since it] (…) does not lay out a step-by-step procedure or even an exhaustive list of topics (…). We synthesize ideas from existing DSS design frameworks to produce a meta-design methodology from which individual DSS designers can develop their own design frameworks, appropriate to their particular needs." The meta-design approach had its advocates who undertook to further develop it (cf. [7], [8], [10]). Due to the volatility and change dynamics of the design space, frequently coupled with a need for organizational change entailed by IT

[1] Now known as Association for Information Systems Special Interest Group on DSS.

[2] One of those was definitely Michael S. Scott-Morton's Ph.D. dissertation written in 1964-1967 at Harvard Business School and published in book form by HBS Press in 1971 under the title "Management Decision Systems." Recalling his work on the dissertation in *DSSResources*, Scott-Morton mentions collaboration with such prominent scholars as H. Simon, J. McKeney, or F. Carr. Other famous contributors to the decision support concept include H. Leavit, T. Gorry, D. Ness, J. Little, T. Gerrity, P. Keen and C. Stabell. As regards institutional involvement, one could agree with P. Keen when he states that the concept originates in the theoretical studies of organizational decision making done at Carnegie Institute of Technology during the late 1950s and early 1960s and the technical work on interactive computer systems, mainly carried out at Massachusetts Institute of Technology in the 1960s.

978-83-60810-53-8/$25.00

deployment, we see ongoing integration of current trends in system design and organizational change [11].

The paper seeks to address both these research areas. The following discussion draws on views that have been voiced in prior publications:

- "The traditional view of DSS components remains useful because it identifies commonalities between different types of DSS, but it provides only an initial perspective for understanding DSS architectures." [5]
- "The architectural design should set a common level of understanding among technical, non-technical and management participants." [9]
- "DSS (…) ideas and concepts were developing in the early 1970s, the technology became widely available at reasonable cost in the 1980s, and in 2008 are rarely used effectively. When they are, they are huge beneficial impacts; indeed some firms could not exist without them. (…) The general unresolved issue I see is one of understanding the management of change." [17]

A thorough analysis of refutations and cracks flawing the theoretical foundations of DSS has led us to propose the extended DDMKCC paradigm: Data – Dialog – Modeling – Knowledge – Communication – Creativity [7]. Further in the paper, a theoretical underpinning is provided alongside a description of architectural recommendations, resulting from our recent research as well as from a wealth of practical experience with the proposed paradigm. The research involved a group of 10 business companies selected from among some 200 in which DSS implementation projects were completed by Consorg S.A. between 2000 and 2012. Each implementation has been examined for the degree to which each of the DDMKCC architectural components is used in making decisions at (1) operational, (2) tactical and (3) strategic level.

I. Traditional DSS Architectures Based on the DDM Paradigm

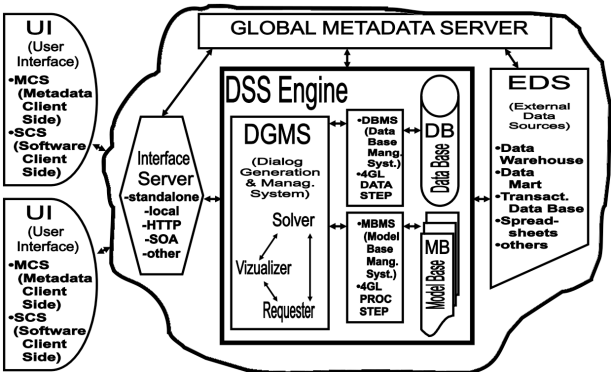


Fig. 1 A traditional DSS architecture in an IT development context

A traditional architecture, such as the one shown in Fig. 1, afforded a possibility to exploit available information technologies in implementing early computer decision support concepts founded on the DDM paradigm and embracing the reflection on interactions between architectural elements such as Data Base, Model Base, Visualizer, Generator, and Solver within a network environment. Looking back at the expansion of specific technologies and the evolution of theory, one easily recognizes the advances in technical solutions identified e.g. with multi-layer architectures, grid computing, data analysis and presentation using data warehouses, or Business Intelligence environments. One of the remaining challenges is to link the DSS model base with models existing in organizations. A lot of such models are in the form of hardly scalable spreadsheets developed by painstaking or enthusiastic users. Their flat meta-data structure is just another important weakness. By introducing further levels, we may be able to implement many of the proposals stemming from IFIP research on the problems of context within DSS ([16], [13]).

II. Origin of the DDMKCC Paradigm

We have already remembered the DSS design paradigm framed by Sprague and Carlson, demanding that DSS consist of three sets of capabilities belonging in the areas of dialog, data and modeling. Many researchers insist that a good DSS should retain balance among the three capabilities. It should be easy to use, too, allowing non-technical decision makers to interact freely with the system. It should be able to access a wide variety of data and provide ample analytical and modeling capabilities [18]. However, observation clearly demonstrates that practice does not fully raise to the promises of theory. Sprague and Watson [19], for instance, contend that many early systems would adopt the name DSS when they were strong in one area and weak in the other. Having analyzed 56 DSS cases, within the two main groups Alter distinguished seven sub-groups based on "the degree to which the system's output can directly determine the decision" [3]:

- Data-oriented DSS: (1) File Drawer Systems, whose purpose is to automate certain manual processes and provide access to data items; (2) Data Analysis Systems, which facilitate the analysis of current and historical data, in order to produce reports for managers; (3) Analysis Information Systems, which provide access to a multitude of support data bases for the decisional process, as well as a series of simple models in order to supply information necessary for solving particular decisional situations.
- Model-oriented DSS: (4) systems oriented on accounting and financial models. The models employed are of "what-if" and "goal-seeking" types and they are frequently utilized in producing profitability estimates for new products, estimative balances, etc; (5) systems oriented on representational models, which use simulation models to estimate consequences; they are used extensively in risk analysis, in production simulation etc.; (6) systems oriented on optimization models which help produce optimal solutions for different activities; (7) systems oriented on suggestion models which carry out the logical process leading to a suggested decision for activities with

a certain degree of structuring (such as determining the frequency of insurance renewal, models for the optimization of bond supply, etc.).

Further studies on existing decision support systems confirm the growing dynamics, diversity, complexity and diffusion of the DSS area. The outcomes become contradictory, the foundations of DSS are crumbling, and special cases are increasingly often reported that supersede or undermine prior research findings. Investigating the "cracks" enables (or compels) researchers to lay broad and solid foundations on which to build up the knowledge [20]. "Many software vendors, information systems consultants, and even some academic researchers are periodically tempted to create a revised vocabulary for existing concepts. Synonyms are variants on accepted concepts which can sometimes aid in understanding, but they can also lead to conceptual confusion. The globalization of discourse on topics like decision support has added to the challenge of communicating meaningfully about our research; more terms increase the difficulty. The academic community needs to control the word labels that are used in our research and discourse for important concepts and constructs. This task is important if researches want to manage and evolve a stream of systematic research on decision support." [21].

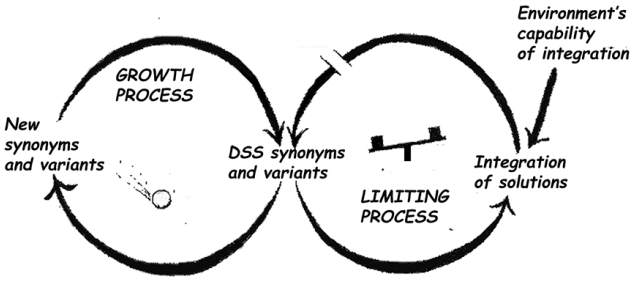


Fig. 2 A Casual loop diagram for a solution to the explosive growth of DSS synonyms and varieties in line with the Limits to Growth Systems Archetype

Theory-oriented research indicates the emergence of the birth effect: a growing number of DSS synonyms and variants generates a larger number of new concepts. For outcomes to be comparable with one another, integration of research sub-areas is required [22], which – according to the Limits to Growth Systems Archetype – involves a negative balancing loop (Fig. 2).

The extended DDMKCC (Data – Dialog – Modeling – Knowledge – Communication – Creativity) arises from the need for integration (cf. Fig. 3), in the context of existing "cracks" in the traditional DSS architecture, drawing on research on the deployment of information technology in organizations [26] and acknowledging the incorporation of soft systems methodology into DSS research. In subsequent publications, further architectural details are added or elaborated.

At the same time, Power [5] develops an integrated DSS classification aligned with the idea of pivotal component proposed by Sprague and Watson (cf. Table I).

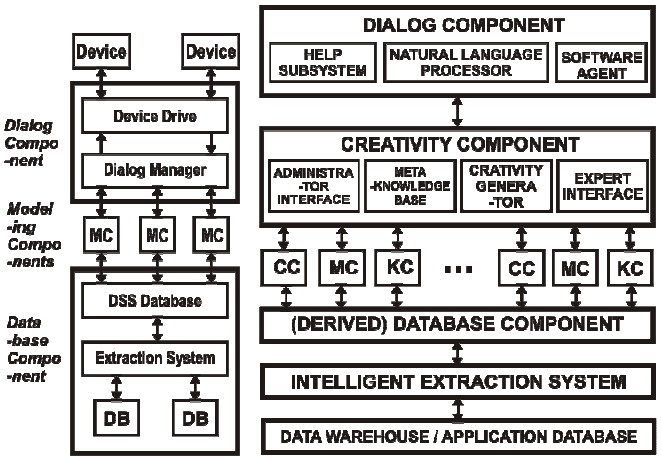


Fig. 3 The DDM architecture – the tower (left) and the DDMKCC. Source: [1], [7]

TABLE I.
A NEW DSS FRAMEWORK

Dominant DSS Component	User groups: Internal, External	Purpose: General, Specific	Enabling Technology
Communications **Communications-Driven DSS**	Internal teams, now expanding	Conduct a meeting Bulletin Board Help users collaborate	Web or Client / Server
Database **Data-Driven DSS**	Managers, staff, now suppliers	Query a Data Warehouse	Main Frame, Client/Server, Web
Document base **Document-Driven DSS**	Specialists and user group is expanding	Search Web pages Find documents	Web
Knowledge base **Knowledge-Driven DSS**	Internal users, now customers	Management Advice Choose products	Client/Server, Web
Models **Model-Driven DSS**	Managers and staff, now customers	Crew Scheduling Decision Analysis	Stand-alone PC

Source: [5]

In this paper, the central research question concerns the software architecture for a decision support system. Theoretical insights and prior validation efforts have led to an extension of the classical Sprague-Carlson DDM proposal toward the DDMKCC paradigm. The paradigm not only implies the building blocks of a decision support system but also provides a basis for addressing and analyzing a broad array of cases. In the following chapters, the findings of research on decision support system design are presented that substantiate the above propositions and underpin practical recommendations.

III. A STUDY OF THE PERFORMANCE OF DDMKCC MODEL COMPONENTS

We selected 10 out of 200 implementation projects run in 2000-2012 by Consorg S.A. In this way, we arrived at a group of 10 business organizations from both the production and the services sectors which we deemed the most representative for our analysis and appraisal of the

proposed approach – viz. its performance and practical effects. The sample included large enterprises as well as capital groups. Some of them are listed on the Warsaw Stock Exchange, while 3 of them (being international capital groups) have parent companies based outside Poland. Most of the companies operate in production industries. The average number of DSS users ranges from 5 to 10 advanced users and 20 to 30 novices or occasional users (see Table II).

In all of the organizations support is centered on decision making processes in the area of financial control. The solution was implemented in an effort to support decisions at all (operational, tactical, and strategic) management levels. At each level, a different set of analytical tools was offered, following the classification of models proposed by Turban and Aronson [26] (see Table III).

TABLE II.
THE 10 SELECTED BUSINESS ORGANIZATIONS FROM THE MANUFACTURING AND SERVICES SECTORS

	Customer	Industry	Organization structure	DSS users	Project duration	Headquarters
1	**Tauron Wytwarzanie** (formerly Południowy Koncern Energetyczny S.A. [Southern Power Corporation]) Parent quoted on Warsaw Stock Exchange	power generation	one-layer capital group	50 primary users 150 other users	2000–2004	Katowice, Poland *www.pke.pl*
2	**Vattenfall Heat Poland S.A.** (Elekrociepłownie Warszawskie S.A.)	power generation	multi-layer capital group	25 primary users 20 other users	2002–2004 upgrade: 2009–2010	Warszawa, Poland (subsidiary) *www.ewsa.com.pl*
3	**Odra Trans S.A.**	inland navigation	multi-layer capital group, 20 subsidiaries (including a Germany based sub-group)	5 primary users 35 other users	implementation period: 2006–2007	Szczecin, Poland
4	**Black Red White S.A.**	furniture manufacturing	one-layer capital group, 30 subsidiaries	5 primary users 30 other users	2007–2008	Biłgoraj, Poland
5	**Pradyż / Ceramika Paradyż sp. z o.o.**	white ware	no capital group	10 primary users 30 other users	2006–2008	Opoczno, Poland
6	**Cersanit S.A.** Parent company quoted on Warsaw Stock Exchange	white ware	multi-layer capital group, 40 subsidiaries (including a Russia based sub-group)	10 primary users 25 other users	2009–2010	Kielce, Poland
7	**EC Będzin S.A.** Listed on Warsaw Stock Exchange	heat generation	member (subsidiary) of RWE capital group	12 primary users	2009–2010	Będzin, Poland (subsidiary)
8	**Kamis S.A.** Listed on Warsaw Stock Exchange	food industry	no capital group	12 primary users 92 other users	2010 –2011	Lubliniec, Poland
9	**Lentex S.A.** Parent company quoted on Warsaw Stock Exchange	chemical industry	one-layer capital group, 1 subsidiary		2010–2011	Poland
10	**Paccor S.A. / Veriplast S.A.**	food packaging	multi-layer capital group, 25 subsidiaries	5 primary users 45 other users	2010–2011	Luxembourg

TABLE III.
DECISIONS, TOOLS AND DECISION MAKING MODELS USED WITHIN THE DDMKCC MODEL

	DECISIONS — TOOLS and MODELS	
	Capital groups	**Other**
1	• Consolidation of financial statements for external reporting (IFRS) • Financial monitoring and budgeting • Corporate supervisory activities • Cash flow planning and monitoring under cash-pooling models	• Operating budget planning and financial analysis • Cash flow analysis models • Pricing models
2	• Simulations and financial monitoring using expert system reports • Benchmarking of functions and processes (diagnosing the causes of deviations in key performance indicators within capital group's strategy performance monitoring) • Models of asset allocation throughout the capital group	• Simulations and financial monitoring using expert systems reports • Benchmarking of the functions and processes of operating budget planning within the corporation's production units (diagnosing causes of deviations in key performance indicators) • Key investment project analysis models

3	• Multi-dimensional simulations of capital sub-group structures; strategic and financial monitoring for management purposes • Value management models for capital groups	• EVA corporate value management models • BSC strategic management models • Strategic resource planning models

where:
1. Operational decisions
2. Tactical decisions
3. Strategic decisions

Besides, the distinct nature of capital group management was reflected in specially tailored business models enabling decision support to be addressed at the parent company level [27]–[30].

Data Sources and Data Models

In the following discussion of the DDMKCC model's "data" component and its usage statistics, data sources and data models will receive separate treatment.

- Data sources

The taxonomy of data sources (cf. Table IV) was based on the classification proposed by R. Sprague and H. Watson [28]. The findings of our observations and analyses are consistent the widely known fact that, although data come from diverse sources, strategic decisions will involve greater use of external sources and less reliance on internal ones (e.g. ERP/MRP systems). The way corporate knowledge bases are exploited in making tactical decisions is notable, too. It is easy to see that relevant data are most commonly sourced from investment project analysis cases stored in archives, since such cases often provide valuable insights into how similar projects were evaluated in the past and offer analogies which can be instrumental in assessing the risk of new investments.

TABLE IV.
DATA SOURCES FOR THE DDMKCC MODEL

DECISIONS	1	2	3	4
Operational	xxx	x	-	-
Tactical	xx	xxx	xx	x
Strategic	x	xx	xxx	xxx

where:
1. Traditional ERP/MRP
2. Text processing and document processing systems; corporate knowledge bases
3. Open access data bases
4. Business information libraries; economic intelligence agencies

- Data models

Likewise, usage analysis of data models led us to believe that data warehouses were used the most when making strategic decisions (cf. Table V), and they were central to capital group management, particularly in groups that have not implemented a single transactional system. For them, a data warehouse can become a integrating DSS component, unifying data and processes across the group. This corresponds to employing a data warehouse to support both operational and strategic decisions.

Interesting observations can be made in examining the use of multi-dimensional OLAP data structures in business decision making processes. When a well designed OLAP cube is combined with an ergonomic viewer, all of the reporting process can be handled via multi-dimensional structures, regardless of the type of decision to be made. It does not matter at all whether OLAP is embedded in a data warehouse or the multi-dimensional repository accesses transactional data directly. Obviously, at the operational level the application of OLAP technologies is reduced to relatively simple (and repetitious) reporting. Advanced functionalities, on the other hand, such as those of data mining and hypothesis validation, are employed at the other decision levels. By surveying the businesses from our sample we were able to ascertain that wherever OLAP technology has been successfully implemented, reporting almost exclusively hinges on data supplied in this way, while other methods have been nearly abandoned.

TABLE V.
DATA MODELS EMPLOYED WITHIN THE DDMKCC PARADIGM

DECISIONS	1	2	3	4
Operational	xx	xx	xxx	x
Tactical	x	xx	xxx	x
Strategic	-	xxx	xxx	xxx

1. Relational data bases
2. Relational data bases – data warehouses
3. Multi-dimensional OLAP data base models
4. File system / document repository

Dialog and Communication Components

The discussion of the dialog component's functionality will be broken down in a pattern proposed by Bennett for the assessment of DSS user interface [4]: (1) knowledge base, conceived as a set of users' essential skills (knowledge) enabling them to work with the system, (2) command language – the way in which users operate the system, and (3) presentation language, i.e. the way in which output is represented [4].

- Knowledge base

Nearly every user highly appreciates the availability of complete system documentation including operating instructions (cf. Table VI). Few, however, actually use it and, as a result, most of them require individual training. As

long as it is fairly sufficient for users at the operational management level, those having to cope with less structured decision problems will normally need to have a good understanding of the problem solving process and to know the applicable techniques. Without this know-how, users situated beyond the operational level might not be able to use the system resources efficiently: even if an expert system is activated to provide them with support in choosing the most suitable tools (models) for their problem, the choice has to be ultimately made by the user. Observation reveals that the most common reason why some systems are not used in tactical or strategic problem solving is not the technology itself but the relatively high demand they put on users' competence (knowledge base).

- Command language

No matter what kind of problems are solved, the ability to communicate with the system via standard and context menus is usually taken for granted or seen as a minimum requirement concerning functionality. Individuals solving tactical problems in companies where an enterprise data warehouse module has been implemented also demanded the option of similar case finding in knowledge bases.

TABLE VI.
THE KNOWLEDGE BASE OF THE DSS USER INTERFACE WITHIN THE DDMKCC MODEL

DECISIONS	1	2	3	4	5
Operational	xxx	x	xxx	x	x
Tactical	xxx	xxx	xx	xxx	x
Strategic	xxx	xx	x	x	xxx

1. Interactive operation manual
2. Examples suited to user skill level
3. Support functions for system navigation
4. Problem solving skills training facility
5. Support options for user learning

Users engaged in solving tactical and strategic problems will rather expect the system to become a "partner in problem solving." Interestingly enough, we found that the lowest skill levels are associated with the highest expectations from the system, including a proactive attitude in assisting the user. Conversely, the expectations of most advanced and creative problem solvers are limited to being offered an efficient technology and a rich collection of presentation tools.

TABLE VII.
USE OF COMMAND LANGUAGE FUNCTIONS IN THE DDMKCC MODEL

DECISIONS	1	2	3	4	5	6
Operational	xxx	xxx	xxx	-	-	-
Tactical	xxx	xxx	x	-	xxx	xx
Strategic	xxx	xxx	-	-	-	xxx

1. Standard system menu
2. Context menu
3. Data base query languages
4. Communication based on natural language processing
5. Knowledge base search for similar cases
6. Ability to guide user dialog toward problem resolution

- Presentation language

Our survey demonstrated that users' expectations concerning the presentation language are closely tied with the mental model of the decision maker being the end user of information output by the system. For example, financial analysts working for top executives expect the presentation language to be as rich as possible and hence capable of satisfying the needs of any user further enhancements to the system. The extent to which specific functions of the presentation language are used will vary largely depending on who uses the output information (e.g. corporate board members will have other preferences than line managers).

TABLE VIII.
FUNCTIONS OF THE PRESENTATION LANGUAGE IN THE DDMKCC MODEL

DECISIONS	1	2	3
Operational	xxx	x	x
Tactical	xxx	xxx	xx
Strategic	xxx	xxx	xxx

1. Data and report presentation in a variety of forms – tables, text, presentation graphics
2. Report definition in terms of detail level and format of delivery (PDF, HTML, Word DOC, etc.)
3. Parallel work with multiple data sections, presentation in multiple forms using multi-window technology

Knowledge and Modeling Components

Within the DDMKCC model, the knowledge component is defined as a resource comprising mathematical models and algorithms designed to transform data into information (deep knowledge) alongside heuristics used to support the decision making process (shallow knowledge) – rules, constraints, boundary conditions or any other information which may be generated within the DSS or acquired during the system's productive operation [5, p. 16]. This approach allowed us to perform usage analysis of specific knowledge components vis-á-vis the type of decision problem. The findings provide important insights that can inform further evolution of the DDMKCC paradigm.

First of all, addressing support to decision making processes at the operational management level does not involve any major modifications to the pre-defined decision making models. These are typically simple cause-effect models focused on explaining deviations of actual performance from plan. It is vital, nevertheless, that simulation and prediction functions be implemented in this class of models to enable "what if" and "what else" analysis (cf. Table IX).

TABLE IX.
FUNCTIONS OF THE PRESENTATION LANGUAGE WITHIN THE DDMKCC MODEL

DECISIONS	1	2	3

Operational	xxx	x	-
Tactical	x	xxx	xx
Strategic	x	xx	xxx

1. Pre-defined models embedded in the DSS
2. Expandable pre-defined models
3. Custom model development and integration tools

The conclusions will be very different as soon as we look at how the system supports strategic decision making processes. What is required of the system in such circumstances is, in the first place, adaptability and expandability by appending new decision models. The DSS not only has to offer the requisite tools to freely build decision models but also needs be able to instantly integrate (owing to two-way data interchange) with dedicated external systems addressing specific business problems. (This would be necessary, for example, in a situation where an investment bank will not agree to open a long-term credit facility unless project performance is assessed and monitored using a model preferred by the bank.).

Secondly, the DDMKCC model includes a special resource containing knowledge on business processes utilized in decision making (decision workflows). Identifying the key business processes and analyzing the decision making processes intrinsic to them makes it possible to accumulate knowledge needed to discover and assess relationships between decisions and their outcomes. This appears critical, in the light of our research, for decision analysis at all levels – operational, tactical, and strategic (cf. Table X).

TABLE X.
FUNCTIONS OF THE PRESENTATION LANGUAGE WITHIN THE DDMKCC MODEL

DECISIONS	**1**	**2**	**3**	**4**
Operational	xxx	xxx	x	x
Tactical	xxx	xxx	xx	xxx
Strategic	x	xxx	xxx	x

1. Mathematical models and algorithms
2. Workflow procedures
3. Heuristics based on expert knowledge
4. Algorithms founded on fuzzy expert rules

Thirdly, our observations suggest that other than deterministic models are used relatively rarely. The most common approach is that founded on deterministic scenario building techniques where the best- and the worst-case scenario are identified. Where probabilistic models are used, preference is given to approaches based on subjective probability.

TABLE XI.
FUNCTIONS OF THE PRESENTATION LANGUAGE WITHIN THE DDMKCC MODEL

DECISIONS	**1**	**2**	**3**
Operational	xxx	x	-
Tactical	xxx	xx	-
Strategic	xxx	x	x

1. Deterministic
2. Fuzzy
3. Probabilistic

Creativity Component

Our survey indicates that the most frequently used creative problem solving tools include: (1) context-sensitive help along with access to historical data and similar cases, (2) a multi-dimensional OLAP data base viewer for convenient hypothesis testing during the creative problem solving process, (3) group work support tools, such as dedicated discussion forums or widely popular instant messengers, (4) SWOT analysis support tools, (5) tools and models for multi-criteria "what if" analyses, and (6) context-oriented reports recapitulating the user's work outcomes; to deliver these outcomes, such reports make use of e.g. expert systems, presentation graphics, tabular views and layouts [6].

The use of each type of tool was examined by observing the subsequent stages of budget planning and budget control processes (monitoring deviations from targets) in capital groups – cf. Table XII.

TABLE XII.
CREATIVITY SUPPORT TOOLS MOST HEAVILY USED BY CAPITAL GROUPS WITHIN BUDGET PLANNING AND CONTROL PROCESSES TO MONITOR DEVIATIONS FROM PLAN

	PROCESS	**1**	**2**	**3**	**4**	**5**	**6**
1	setting budget targets for subsidiary companies	xxx	-	x	-	x	x
2	budget modeling in daughter companies	x	xxx	x	-	xxx	xx
3	management-led consolidation of financial budgets	x	xx	-	-	-	x
4	analysis of threats and opportunities to performance of consolidated group budget	x	xx	x	xxx	xxx	xxx
5	analysis of strengths and weaknesses of subsidiary companies' financial budgets	x	xxx	x	xxx	xxx	xxx
6	budget negotiations	-	xx	xxx	-	xxx	x
7	identifying KPIs/CSFs for subsidiary companies' and group's budgets	xx	x	x	-	xxx	xxx
8	monitoring deviations from plan, early warning of potential threats	xx	xx	xx	-	xxx	xxx
9	validation and control of financial budgets	xx	xxx	xx	xx	xxx	xxx

1. Intelligent context-based assistance for problem solving
2. Multi-dimensional OLAP data base viewer
3. Group work support tools
4. SWOT analysis support tools
5. "What if" analytical models

Importantly, we perceived the necessity to make DSS capable of fast and easy integration with specialized external solutions designed to support certain creative problem solving techniques (e.g. brainstorming or morphology analysis).

IV. Conclusions

Researchers dealing with computerized decision support exhibit growing interest in integrating individual and domain-specific insights and building common theoretical, methodological and applicational frameworks that can sustain systemic thinking.

In the long run, thinking in terms of software architectures facilitates DSS development and maintenance. A holistic view fosters diverse applications, iterative development and, in particular, this distinctive approach, perhaps unique to DSS, whereby systems are developed in response to changes in the decision space. Many DSS have evolved from a data-oriented system through modeling a specific domain, e.g. financial control, which became a starting point, then arousing broader interest in the system itself and inspiring innovative efforts at large. Next, there arises a need for group work and creativity support.

By investigating, across multiple aspects, the ways in which specific DDMKCC model components are used in the practice of making business decisions, we have identified the key determinants of an effective development context for computerized decision support systems. The paper presents research findings which encourage a belief that further development of context-dependent DSS design meta-methodology should be approached from the system designer's perspective.

References

[1] R. H. Sprague and E. Carlson, *Building Effective Decision Support Systems*. Upper Saddle River, NJ: Prentice Hall, 1982.

[2] D. J. Power, *Decision Support, Analytics, and Business Intelligence*. New York: Business Expert Press, 2013.

[3] S. Alter, *Decision Support Systems: Current Practices and Continuing Challenges*. Reading, MA: Addison-Wesley, 1980.

[4] http://dssresources.com

[5] D. J. Power, *Decision Support Systems: Concepts and Resources for Managers*. Westport, Connecticut: Quorum Books, 2002.

[6] J. H. Moore and M. G. Chang, "Meta Design Considerations in Building DSS," in *Building Decision Support Systems*, J. L. Bennett, Ed. Reading, MA: Addison Wesley, 1983, pp. 173-204.

[7] S. Stanek, *Metodologia budowy komputerowych systemów wspomagania organizacji*. Series: *Prace Naukowe AE Katowice*. Katowice: Wydawnictwo Uczelniane AE Katowice, 1999.

[8] A.R. Hevner and S. Hatterjee, *Design Research in Information Systems*. Series: *Integrated Series in Information Systems*, Vol. 22, Springer 2010.

[9] R. Lambert, "Data Warehousing Fundamentals: What You Need to Know to Succeed". *Data Management Review*, March 1996.

[10] O. E. El-Gayar, A. V. Deokar and J. Tao, "DSS-CMM: A Capability Maturity Model for DSS Development Processes," in *Engineering Effective Decision Support Technologies. New Models and Applications*, D. Power, Ed. Hershey, PA: IGI Global, 2013.

[11] T. Fry, *Design Futuring: Sustainability, Ethics and New Practice*. Oxford: Berg, 2009.

[12] S. Liu, A. H. B. Duffy and I. M. Boyle, "Integration decision support systems to improve decision support performance," *Knowledge Information Systems*, March 2010, Vol. 22, Issue 3, pp. 261–286.

[13] *Fusing Decision Support Systems into the Fabric of Context*, A. Respicio and F. Burstain, Eds. Series: *Frontiers in Artificial Intelligence and Applications*, Vol. 238, IOS Press, 2012.

[14] *Bridging the Socio-technical Gap in Decision Support Systems*, A. Respicio, F. Adam, G. Phillips-Wren, C. Teixeira and J. Telhada, Eds. Series: *Frontiers in Artificial Intelligence and Applications*, Vol. 212, IOS Press, 2010.

[15] *Collaborative Decision Making: Perspectives and Challenges*, P. Zarate, J. P. Belaud, G. Camilleri and F. Ravat, Eds. Series: *Frontiers in Artificial Intelligence and Applications*, Vol. 176, IOS Press, 2008.

[16] *Context Sensitive Decision Support Systems*, D. Berkeley, G. Widmeyer, P. Brezillon and V. Rajkovic, Eds. Dordrecht: Chapman & Hall / Kluwer Academic Publishers, 1998.

[17] M. S. Scott-Morton, *Reflections of Decision Support Pioneers*. http://dssresources.com/reflections/scottmorton/scottmorton9282007.html

[18] U. Averweg, "Decision Support Systems and Decision-Making Processes," in *Encyclopedia of Decision Making and Decision Support Technologies*, F. Adam and P. Humphreys, Eds. Hershey–New York: Information Science Reference, 2008, pp. 218-224.

[19] R. H. Sprague and H. J. Watson, "Bit by Bit: Toward Decision Support Systems," *California Management Review*, Vol. 22, No. 1, pp. 60–68, Fall 1979.

[20] S. Stanek, "Two Poles: towards Integration of Research Results and a New Strategic Information Technology Management," in *Proc. of the 4th Conference of the International Society for Decision Support Systems*, Lausanne, Switzerland, July 21–22, 1997.

[21] D. J. Power, "Defining Decision Support Constructs," in *DSS in the Uncertainty of the Internet Age*, T. Bui, H. Sroka, S. Stanek and J. Gołuchowski, Eds. Katowice: AE Katowice, 2003, pp. 51–61.

[22] *Decision Support Systems for Sustainable Development. A Resource Book of Methods and Appliations*, G. E. Kersten, Z. Mikolajuk and A. Gar-On Yeh, Eds. Norwell, MA: Kluwer Academic Publisher, 1999.

[23] M. Chen, "A Model Driven Approach to Accessing Managerial Information: The Development of a Repository-Based Executive Information System," *Journal of Management Information Systems*, Springer 1995, Vol. 11, No. 4, pp. 33–63.

[24] N. M. Duffy, "EIS in Context," in *Context Sensitive Decision Support Systems*, D. Berkeley, G. Widmeyer, P. Brezillon and V. Rajkovic, Eds. Dordrecht: Chapman & Hall / Kluwer Academic Publishers, 1998.

[25]http://dssresources.com/reflections/inmon/inmon06072007.html

[26] E. Turban and J. E. Aronson, *Decision Support Systems and Intelligent Systems*. Upper Saddle River, NJ: Prentice Hall, 2001.

[27] Z. Twardowski, J. Wartini-Twardowska and S. Stanek, "A Decision Support System based on DDMCC paradigm for strategic management of capital groups," in *Advanced Information Technologies for Management AITM 2011 – Intelligent Technologies and Applications.* Research Papers of Wrocław University of Economics, No. 206, J. Korczak, H. Dudycz, M. Dyczkowski, Eds. Wrocław: Publishing House of the Wrocław University of Economics, 2011.

[28] R. H. Sprague and H. J. Watson, *Decision Support for Management*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[29] G. M. Marakas, *Decision Support Systems in the 21th Century*. Upper Saddle River, NJ: Prentice Hall / Pearson Education, 2003.

[30] J. Wartini-Twardowska and Z. Twardowski, "Proces konsolidacji budżetów finansowych w strategicznym zarządzaniu grupą kapitałową," *Zeszyty Naukowe*, No. 534. Series: *Finanse, Rynki finansowe, Ubezpieczenia*, Vol. 17 (2009). Szczecin: Uniwersytet Szczeciński, 2009, pp. 613-627.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1165–1172

The Structure of Agility from Different Perspectives

Roy Wendler
Technische Universität Dresden
Faculty of Business Management and Economics
Chair of Information Systems, esp. IS in Manufacturing and Commerce
Dresden, Germany
Email: roy.wendler@tu-dresden.de

***Abstract*—Agility as a term is widely known today. However, a common understanding of what agility means and what it consists of is missing. Until today, a lot of frameworks have been developed, but they are very heterogeneous regarding content and structure. This paper approaches that issue by conducting a systematic comparison of 28 available agility frameworks out of the domains of agile manufacturing, agile software development, agile organization, and agile workforce. Altogether, 33 concepts related to agility were identified. The results of the comparison show that even within the examined specific domains a lack of consensus is obvious. In addition, the utilized concepts are very ambigious and overlapping. So, the interdependencies between the identified concepts were analyzed in detail. This revealed five recurring "clusters" that each combine several concepts with similar content, but despite the amount of available frameworks, none of it reflects these clusters directly. Hence, the study shows that the factors beyond the construct of agility are not fully uncovered yet.**

I. Introduction

FOR several years, businesses and organizations have faced a more and more volatile environment, marked with challenges such as increased competition, globalized markets, and individualized customer requirements accompanied with many changes in every organizational field. Such scenarios were already described in the 90s, for instance by Goldman et al. [1] or the Iacocca Institute [2]. As a response, different concepts emerged that should enable organizations to master these challenges. The most recent is the concept of agility, but others like flexibility and leanness are mentioned often, too.

A lot of research activities about agility and its related concepts have been conducted in the meantime. However, until today there exists no common understanding of what constitutes agility. Although many frameworks and models describe agility and its characteristics, they often heavily differ in content and structure. This makes it difficult for both, researcher and practitioner audiences, to build upon the insights obtained until today. On the one hand, researchers are missing a well-founded basis to develop the topic further, on the other hand, practitioners cannot easily uncover what parts of their organizations have to be changed and to what respect they have to be changed to respond to new market challenges.

This is particularly of interest for organizations in the software and information technology (IT) industry. With the appearance of agile software developing methodologies, or in a broader sense agile values and principles (see for instance [3]–[5]), in the early 2000s, the advantages of these new approaches became visible. However, it turned out to be difficult to transfer the experienced benefits beyond the team level [6]–[8]. But this step is necessary so that the whole organization can benefit from agility.

The idea for this paper arose from the attempt to select a suitable agility framework that represents the structure and components of agility in an organization for a further empirical study. Unfortunately, it turned out that due to the aforementioned problem of a lack of consensus, a selection of one framework seemed unsatisfactory. Some were unsuitable to describe the organization as a whole, others specialized on a specific aspect only. Generally, the literature was confusing and inconsistent. Therefore, it seemed necessary to get through the literature and systematically compare available frameworks. The aim of this work is to analyze the frameworks regarding common ground and differences and to search for recurring concepts. Finally, this will create a basis for further work on a common understanding of agility.

A review about agility is already given by Sherehiy et al. [9] which serves as an important starting point for this study, too. However, they mainly included work of the agile manufacturing domain, because publications about the agile organization as a whole were scarce at this time. The aim of the authors was to deduce a summarized framework describing the agile organization. Interestingly later published frameworks again differ heavily from the one developed by Sherehiy et al. That shows that their work was still not sufficient and a further investigation is necessary.

The remainder of the paper is structured as follows: In section II the concept of agility and its history are shortly described and its connections to the principles of lean and flexible are mentioned. Section III introduces the agility frameworks that are analyzed in this paper. The systematic comparison of these frameworks and the discussion of the results are given in section IV. The paper ends with a conclusion and an outlook to further research currently conducted in section V.

II. The Concept of Agility

Looking up the term agile in a dictionary delivers "having the faculty of quick motion; nimble, active, ready" [10, p. 255], whereby agility is the "quality of being agile" [10, p. 256]. Given this explanation as a basis, a huge variety of definitions emerged today, heavily influenced by context and application domain. A discussion of all available definitions is

978-83-60810-53-8/$25.00

beyond the scope of this paper. Different authors already list various definitions of agility, for instance [11], [12]. Another comprehensive collection is given in the appendix of [13].

An extensive definition, which fits well to the context of this work, was developed by Ganguly et al. [14] based on the work of Dove [15], [16]. They define agility as "an effective integration of response ability and knowledge management in order to rapidly, efficiently and accurately adapt to any unexpected (or unpredictable) change in both proactive and reactive business / customer needs and opportunities without compromising with the cost or the quality of the product / process" [14, p. 411]. The handling of change as a fundamental prerequisite for agility is confirmed by Conboy, who named creation of change, proaction in advance of change, reaction to change, and learning from change as components of agility [17].

The concept of agility is nothing new. Early works are already found within social sciences and date back to the 1950s [18]. However, agility gained significantly more attention in the 1990s, especially after the so called "Lehigh report" [2] was published explaining a new idea of manufacturing strategies. This development was accompanied by the increasing emphasis on customer orientation and proactivity instead of reactivity. Later on, mainly after the year 2000, process orientation was focused on additionally and led to an examination of agility from an organizational perspective [13]. Simultaneously, agility became well known within the software industry, whereby the "Agile Manifesto" [3] triggered a lot of research in this field.

While research about agility progresses continously, there are two other closely connected and underlying concepts: flexibility and leanness. Although both share some common ground with agility, they are not the same and should be distinguished. A detailed discussion is given in [17], which is shortly summarized here. First, flexibility is very similar to agility. The main differences of flexibility lie in issues like the lack of speed and rapid action, continual change instead of a one-off change, a missing inclusion of knowledge and learning, and the application as single practices in specific parts of the company instead of an organization-wide view. The difference of leanness, however, is much more straight forward. In contrast to agility, leanness is unsuitable to variability and uncertainty and emphasizes simple cost reduction over value-related issues, mainly value for the customer [17].

III. Agility Frameworks

A review of the literature revealed a variety of frameworks and models describing the concepts that determine agility or at least proposed different items to measure agility. Finally, 28 frameworks or similar concepts could be identified that can be categorized into four domains and will shortly be introduced in the following:

- Agile Manufacturing,
- Agile Software Development,
- Agile Organization/Agile Enterprise, and
- Agile Workforce.

A. Agility Frameworks Focusing on Agile Manufacturing

As explained in section II, the concept of agility mainly originates from the manufacturing domain. Hence, ten of the identified frameworks focus on agile manufacturing [11], [19]–[27].

One of the earlier frameworks was developed in 1999 by Sharifi & Zhang [19]. The core idea is the distinction between Agility Drivers, Agility Capabilities, and Agility Providers. Drivers are mainly changes in the environment. Capabilities like responsiveness, competency, flexibility, and speed are the the required abilities of the company to respond to these changes. Providers are the means to achieve these capabilities in the areas of organization, technology, people, and innovation [19]. This framework was refined and extended later by Sharifi et al. [20], however, the main structure remained stable. Finally, it led to a theoretical approach to develop an agile manufacturing strategy [21].

A similar structure was chosen by Vázquez-Bustelo et al. [22], by grouping the elements of their conceptual model into Agility Drivers, Agility Enablers (or Practices), and Outcomes. The core concept is the Agile Enablers, which are similar to the Capabilities mentioned above, but are further detailed into Human Resources, Value Chain Integration, Concurrent Engineering, Technologies, and Knowledge Management [22].

Two other early frameworks were developed by Gunasekaran [23] and Yusuf et al. [24], whereby both identify four major dimensions affecting the agile manufacturing system. Gunasekaran mentions Strategies, Technologies, People, and Systems [23]. Yusuf et al., however, state Core Competence Management, a Capability for Reconfiguration, a Knowledge-driven Enterprise, and the formation of Virtual Enterprises as core concepts. They furthermore detail them into 32 attributes [24]. In 2002, Gunasekaran & Yusuf published another framework of agile manufacturing strategies and techniques that implemented concepts of both predecessors [11].

The remaining three frameworks show different approaches. Meredith & Francis propose a so called "Agile Wheel" structuring agility into Strategy, Processes, Linkages, and People each with four sub-practices [25]. Agarwal et al. focus on the agile supply chain by stating four main characteristics dealing with Market, Information Integration, Process Integration, and Planning [26]. Addditionally, Kisperska-Moron & Swierczek conducted an exploratory factor analysis with Polish companies and obtained a framework built of the four factors Relation with Customers, Relation with Suppliers, Relation with Competitors, and Intensity of IT Use [27].

B. Agility Frameworks Focusing on Software Development

Research about agile software development is much younger. As described in section II, the Agile Manifesto [3] can be seen as a trigger for further studies. The 17 initiators postulate four key values of agile software development with an emphasis on individuals and interactions, working software, customer collaboration, and response to change. These values are further detailed into 12 principles [3]. Afterwards, in the years 2008 and 2009, five frameworks dealing with the topic of

agile software development were identified [28]–[32], whereby they often focus only on specific issues within the domain.

Two of the more general frameworks dealing with success factors of agile development practices are given by Chow & Cao [28] and Misra et al. [29]. Both publications show comprehensive lists of success factors grouped in different dimensions. Chow & Cao use organization, people, process, technical, and project factors [28], whereby Misra et al. only distinguish between organizational and people factors [29]. However, both narrow down these lists to six (delivery strategy, proper agile software engineering techniques, high team capabilities, good agile project management process, agile-friendly team environment, and strong customer involvement) [28] and nine (customer satisfaction, customer collaboration, customer commitment, decision time, corporate culture, control, personal characteristics, societal culture, and training and learning) [29] critical success factors via emprical investigations, respectively.

In contrast, Chan & Thong [30] ask what affects the acceptance of agile methodolgies. In this context, they build a conceptual framework where acceptance is dependent from the characteristics of the agile methodologies and knowledge management-related activities like creation, retention, and transfer of knowledge and experience. They furthermore identify three concepts affecting knowledge management: factors related to abilities, motivation, and opportunities [30].

Agility in the specific domain of distributed development teams was analyzed by Sarker & Sarker [31]. They distinguish three different dimensions of agility. First, Resource Agility that mainly consists of people and technology. Second, Process Agility that includes aspects like methodology, environmental awareness, or bridging time zones. And finally, Linkage Agility that is based on cultural and communicative issues [31].

Furthermore, Kettunen [32] did not develop a framework in a stricter sense, but compared practices of agile manufacturing to those of agile software development. For this purpose, he used a comparison matrix covering five concepts: Organization, Process, Product, Operation, and People. He concludes that issues of all manufacturing concepts are covered in different amounts by agile software development models, too [32].

C. Agility Frameworks Focusing on Agile Organization/Agile Enterprise

Research about the agile organization as a whole unit started contemporarily to research about agile manufacturing in the 90s. However, a concentration can be seen at the time the interest in agile software development grew. Additionally, the newest publications (from 2010 and 2011) of all analyzed frameworks belong to this group. This might be an indicator that it becomes more important to understand the effects of agility to the overall organization beyond single development teams or the manufacturing domain. All together, 11 frameworks were identified covering the topic of the agile organization [1], [9], [33]–[41].

One of the first and a well-known publications is the book of Goldman et al. [1]. They label agility as "A Framework for Mastering Change" and define four dimensions to stay competitive: enriching the customer, cooperating to enhance competetiveness, organizing to master change and uncertainty, and leveraging the impact of people and information [1].

Besides this, different approaches dealing with organizational agilty have been developed. A part of the literature focuses on measurement tools. Ren et al. [33], for instance, propose a measurement system utilizing the Analytical Hierarchy Process (AHP) based on the four dimensions of Goldman et al. [1], [33].

Other authors utilize fuzzy logic as a measurement tool. Tsourveloudis & Valavanis [34] name a set of parameters to measure agility by assessing the infrastructure of production, market, people, and information [34] whereas Lin et al. [35] closely connect their fuzzy logic model to the concepts of agile manufacturing with agility enablers, capabilities, and drivers (see section III-A) [35]. Later on, Tseng & Lin use this model to introduce an agility development method [36].

The use of agile manufacturing concepts can also be observed in other publications. Eshlagy et al. [37] again use the distinction in agility enablers, capabilities, and drivers in their research. They finally identified 12 factors that have an effect on organizational agility by applying path analysis. Interestingly, the most significant are leadership, organization commitment, and job satisfaction while typical manufacturing issues like supply chains and the like play a less important role [37]. In a similar way, Bottani [38] uses the framework of Yusuf et al. [24] to conduct an empirical study with the aim of analyzing what profile agile companies have and which tools they use [38].

A comprehensive work to develop a measurement scale with qualitative and quantitative studies can be found in Charbonnier-Voirin [39]. The given scale consists of four factors that can be seen as a framework for agility, too. The factors are somewhat similar to the dimensions of Goldman et al. [1]. They are named practices directed towards mastering change, practices valuing human resources, cooperative practices, and practices of value creation for customers [39].

Similar to section III-B, there also exist some publications dealing with very specific topics. Tallon & Pinsonneault investigate the impact of strategic IT alignment on agility [40]. Zelbst et al. show that the utilization of RFID technology enhances agility [41]. Both additionally identify a positive effect of agility on the performance of the firm [40], [41].

Finally, a review of concepts related to enterprise agility is given by Sherehiy et al. [9]. They reviewed a number of frameworks, models, and measurement tools of agility and extracted a list of characteristics of the agile enterprise. They distinguished into characteristics related to global strategies including customer, cooperation, organizational learning, and culture of change as well as characteristics related to organization and workforce including authority, rules and procedures, coordination, structure, human resource management, proactivity, adaptivity, and resiliency [9].

D. Agility Frameworks Focusing on Agile Workforce

Within the domain of agile workforce, only one publication could be identified [42]. However, in its content specializing on people without referring to manufacturing or software development, it forms a unique sub-domain of agility. Breu et al. identify ten key attributes of an agile workforce that are grouped into the five capabilities intelligence, competencies, collaboration, culture, and information systems [42].

IV. Systematic Comparison of Agility Frameworks

A. Procedure

To achieve a systematic comparison of the frameworks introduced in section III, the following procedure has been applied: First, the core concepts (for instance "customer," "processes," "change," etc.) of the first framework were listed. Then, the core concepts of the next frameworks were assigned to appropriate existing ones or they were added to the list, if they were new. If there were only different labels, but the the same content (for instance "people" vs. "workforce" vs. "teams" vs. "employees") these concepts were treated as one. This step was repeated for every framework. At the end, this resulted in a list of 33 concepts of agility.

As mentioned in section III, the concepts sometimes were detailed into further indicators, attributes, etc. This information was used afterwards to assess whether or not two or more concepts were linked to each other content wise. As a result, a network could be drawn showing the interdependencies between the concepts.

B. Mapping of the frameworks

Figure 1 shows the complete mapping of the analyzed frameworks to the extracted concepts of agility. The numbers on the right side show how often a concept was used over all frameworks. It becomes clear that the concepts and frameworks are very ambigious. In none of the domains is there a more or less stable structure. This indicates that despite the ongoing research, there is still no common understanding of what constitutes agility.

A few of the concepts are prevalent in every domain. Most used was the concept "Workforce/Teams." That seems obvious, because workforce plays an important role when talking about agility. Closely connected are the "Organizational Competences/Abilities," which are also often used and nearly equally distributed over all domains. In addition, two other concepts are interesting: "Cooperation" and "Technology" are among the most used concepts. However, they are both only once named within the domain of agile software development.

Figure 2 summarizes the mapping per domain. The numbers in the cells represent the number of frameworks that use the corresponding concept. Figure 2 reveals that the domain of the agile organization is the most comprehensive one by covering 30 of the 33 identified concepts. But, as mentioned in section III-C, many of the frameworks in this domain utilize structures of agile manufacturing. This is also visible by the fact that every concept of the manufacturing domain is used at least once within the domain of the agile organization. However, Eshlaghy et al. showed that pure manufacturing related concepts had the least significant effects on agility from an organizational perspective [37] (see section III-C). So, one should ask, if it is useful to simply transfer the concepts of agile manufacturing to the agile organization.

Another interesting fact lies in the domain of the agile workforce. Of course, the number of concepts is the lowest, because only one framework was identified. However, two concepts, namely "Intelligence" and "Collaboration," are only present in this domain. This is surprising, because it could be assumed that these workforce-related concepts are important in every domain.

At this point it becomes clear that the inherent ambiguity makes it difficult to compare the frameworks in more detail. Of course, concepts that occur only once may also be covered by differently named concepts. For instance, "Adaptivity," "Resiliency," "Collaboration," or "Intelligence" may also be covered by "Organizational Culture" or others. Also the fact that for instance "Workforce/Teams" is not used in every framework may be an indicator that it is also covered by other concepts. Hence, as described in section IV-A, the links between the concepts are analyzed further.

C. Interdependencies of agility concepts

After looking into the details of every concept, it was possible to determine connections between them. Some are on higher abstraction levels, so that they include others. In other situations, two concepts overlap in some parts (for instance "Customer" and "Market," or "Education" and "Intelligence"). Yet, they could not be merged, because both also covered unique content. Generally spoken, a connection between two concepts means that they are linked to each other content wise, but without a further semantic specification. After identification of every connection, a network was drawn visualizing the interdependencies. This network was created with the open-source tool Gephi [43] using the layout algorithm ForceAtlas 2 with the concepts as nodes and the connections between them as unweighted edges. The resulting graph is given in figure 3.

The first noticeable issue is the high number of linkages between the single concepts. This again underlines the fact that a common understanding of agility is missing. However, by arranging the network with the layout algorithm mentioned above, some "clusters" that have connections to a lot of other concepts become visible. These are illustrated as colored ellipses in figure 3 and are namely:

- Organizational Culture,
- Workforce,
- Customer,
- Organizational Abilities, and
- Technology.

Comparing the analyzed frameworks with this new structure, it turns out that only seven cover concepts of all five clusters [20], [21], [24], [33], [35], [37], [38]. These seven frameworks are out of the domains of agile manufacturing or the agile organization. In contrast, nine frameworks only cover concepts from three out of the five clusters: four frameworks in the

	Agarwal et al. (2007) [26]	Gunasekaran (1999) [23]	Gunasekaran & Yusuf (2002) [11]	Kisperska-Moron & Swierczek (2009) [27]	Meredith & Francis (2000) [25]	Sharifi & Zhang (1999) [19]	Sharifi et al. (2001) [20]	Vázquez-Bustelo et al. (2007) [22]	Yusuf et al. (1999) [24]	Zhang & Sharifi (2007) [21]	Breu et al. (2001) [42]	Agile Manifesto (2001) [3]	Chan & Thong (2009) [30]	Chow & Cao (2008) [28]	Kettunen (2009) [32]	Misra et al. (2009) [29]	Sarker & Sarker (2009) [31]	Bottani (2010) [38]	Charbonnier-Voirin (2011) [39]	Eshlaghy et al. (2010) [37]	Goldman et al. (1995) [1]	Lin et al. (2006) [35]	Ren et al. (2000) [33]	Sherehiy et al. (2007) [9]	Tallon & Pinsonneault (2011) [40]	Tseng & Lin (2011) [36]	Tsourveloudis & Valavanis (2002) [34]	Zelbst et al. (2011) [41]	
Customer				X						X		X				X			X	X	X			X	X	X			10
Market	X		X						X									X				X	X		X	X	X	X	10
Product			X												X										X			X	4
Quality									X			X			X			X					X	X				X	7
Cooperation	X		X	X	X			X	X	X		X						X	X	X	X	X	X	X	X	X	X	X	18
Organizational Culture							X			X	X	X		X	X	X	X			X	X			X					11
Structure																				X				X					2
Coordination																	X							X					2
Authority																								X					1
Change									X			X						X	X	X		X	X	X					8
Integration	X							X	X	X								X				X	X			X			7
Organizational Learning								X					X											X					3
HRM Practices								X												X				X					3
Processes	X				X			X				X		X	X		X									X		X	9
Innovation							X			X																X			3
Strategy		X			X																							X	3
Workforce / Teams		X	X		X		X		X	X		X	X	X	X	X	X	X	X	X	X	X	X				X		19
Proactivity										X														X					2
Adaptivity																								X					1
Resiliency																								X					1
Org. Abilities / Competences						X	X		X	X	X	X	X			X		X		X		X	X			X		X	14
Intelligence											X																		1
Collaboration											X																		1
Motivation													X							X									2
Welfare									X									X				X	X						4
Education									X									X				X	X						4
Technology		X	X	X			X	X	X	X	X						X	X		X		X	X		X	X		X	16
Systems		X	X							X																	X		4
Information							X													X						X	X		4
Project														X		X													2
Responsiveness						X	X			X										X						X			5
Flexibility						X	X			X										X						X			5
Quickness						X	X			X										X						X			5

Legend

Focus Agile Manufacturing | Focus Agile Organization / Enterprise | Focus Agile Software Development | Focus Agile Workforce

Fig. 1. Mapping of agility frameworks and agility concepts

domain of agile manufacturing [19], [23], [26], [27], two in the domain of agile software development [28], [32], and three in the domain of the agile organization [36], [39], [40].

There are also differences between which clusters are missing within the frameworks. The cluster that is covered by most frameworks is "Customer." Only three frameworks are missing any concept of this cluster. Five respectively six frameworks do not cover concepts of the clusters "Organizational Abilities" and "Workforce." The remaining two clusters are missing most within the frameworks. Eight do not share concepts of "Organizational Culture" and even ten neglect "Technology."

Interestingly, all but one framework of agile software development are missing the latter one. The only one covering the technology aspect is [31]. The reason may be that in agile software development technologies and systems are basic prerequisites and therefore not seen as factors affecting agility. Also the gaps in "Organizational Abilities" are prevalent in frameworks of the software development domain [28], [31], [32]. There are studies reporting a lot of problems when adopting agile methods beyond the development team (see, for instance, [6]–[8]). The gaps in the analyzed frameworks regarding organizational abilities may be a cause of these problems.

A similar accumulation of gaps can be observed for the cluster "Organizational Culture." Four frameworks within agile manufacturing do not cover concepts of this cluster [11], [20],

	Agile Manufacturing	Agile Workforce	Agile Software Development	Agile Enterprise / Organization
Customer	2	-	2	6
Market	3	-	-	7
Product	1	-	1	2
Quality	1	-	2	4
Cooperation	7	-	1	10
Organizational Culture	2	1	5	3
Structure	-	-	-	2
Coordination	-	-	1	1
Authority	-	-	-	1
Change	1	-	1	6
Integration	3	-	-	4
Organizational Learning	1	-	1	1
HRM Practices	1	-	-	2
Processes	3	-	4	2
Innovation	2	-	-	1
Strategy	2	-	-	1
Workforce / Teams	6	-	6	7
Proactivity	1	-	-	1
Adaptivity	-	-	-	1
Resiliency	-	-	-	1
Org. Abilities / Competences	4	1	3	6
Intelligence	-	1	-	-
Collaboration	-	1	-	-
Motivation	-	-	1	1
Welfare	1	-	-	3
Education	1	-	-	3
Technology	7	1	1	7
Systems	3	-	-	1
Information	1	-	-	3
Project	-	-	2	-
Responsiveness	3	-	-	2
Flexibility	3	-	-	2
Quickness	3	-	-	2

Fig. 2. Number of frameworks regarding agility concepts

[26], [27]. Surprisingly, the other four frameworks missing any concept of organizational culture are to be found in the domain of the agile organization [34], [36], [39], [40]. However, these frameworks cover many more concepts of the clusters "Workforce" and "Organizational Abilities."

Besides, another issue draws attention. The two concepts "Processes" and "Change" have a very central position with many connections to other concepts, but it is difficult to identify new clusters around them. Change itself is often named as one of the key drivers of agility. Processes are an important internal element of every organization. Without changing processes, it will not be possible to change the way of work. Hence, their central position may be an indicator that many authors consider it relevant within other concepts. However, this issue has to be examined in more detail in future studies.

V. Conclusion and Outlook

This study identified and systematically compared 28 frameworks of agility. These covered the domains of agile manufacturing, agile software development, agile organization, and agile workforce. As the observations in section IV clearly reveal, it is difficult to draw a sharp line between the five identified clusters. Furthermore, there is absolutely no consensus of what really constitutes the construct of agility. The analyzed frameworks are very different regarding their structure and content. Even within the specific domains of agility, the frameworks vary a lot.

This has significant implications for research. Due to the lack of consensus, it is difficult to conduct empirical studies or to build upon existing frameworks. When researchers have to decide between one or another of the available frameworks as the basis for their research, they will most likely miss some concepts of agility, as shown in section IV-C.

Hence, this study may serve as a good starting point to choose one of the frameworks, because it will give the reader an overview of the covered concepts of every framework. It sharpens awareness that the frameworks have gaps and gives the reader the opportunity to close these gaps with parts of other suitable agility frameworks. However, to date there is no empirical study that delivered a comprehensive picture of agility in an exploratory way. So it remains unclear, which concepts of the frameworks are prevalent in practice and how the factors behind agility are composed.

Of course, some of the analyzed publications included exploratory studies, but they all show some limitations. Examples are the works of Kisperska-Moron & Swierczek [27] and Charbonnier-Voirin [39] that both also conducted exploratory factor anaysis. However, both are missing some important concepts (see section IV-C). Apart from that, other authors conducted empirical studies, too, but only used a specific framework that, again, does not cover all identified agility concepts. For example, Bottani [38], who used the framework of Yusuf et al. [24], or Zhang & Sharifi [21], who used their own developed framework [19]–[21].

Due to this limitation, the author of this paper currently is conducting an empirical study about the question of what constitutes an agile enterprise at an organizational level. Hence, the identified agility concepts were merged into a questionnaire with 68 items. The final aim is to conduct a factor analysis to uncover the structure that lies behind the construct of agility. The currently focused target group are both general and IT-related decision makers in companies of the software and IT-service industry. In contrast to the aforementioned studies, it contains all of the concepts given in figure 1. Therefore, it will deliver a comprehensive and, to date, not available view on agility. According to Conboy, who states that "the search for a definitive, all-encompassing concept of agility might not be found simply through an examination of

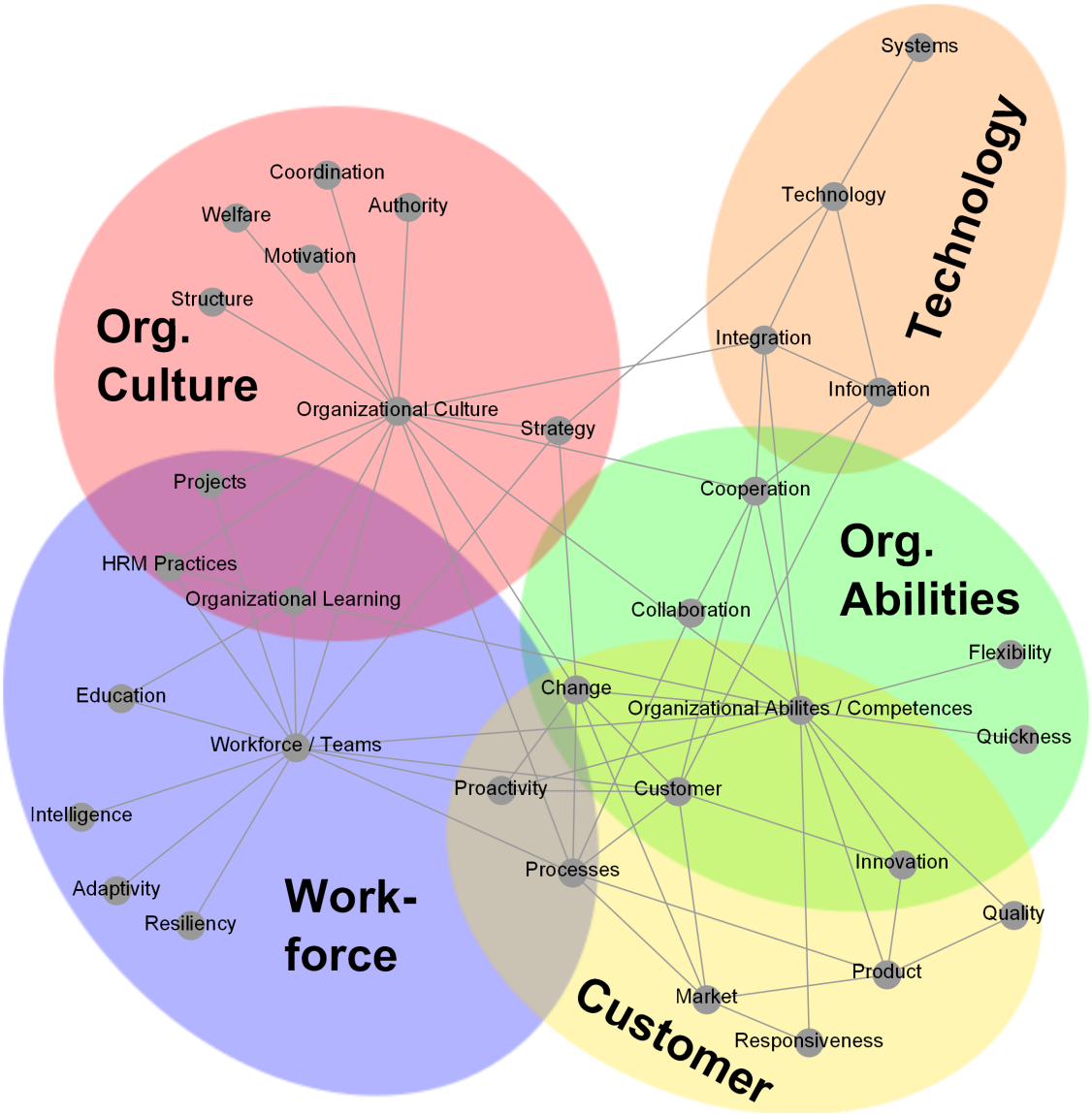


Fig. 3. Interdependencies of agility concepts

agility in other fields" [17, p. 334], this ongoing research will ideally solve the contradictions identified within this paper and contribute to an increasing consensus of what constitutes agility.

REFERENCES

[1] S. L. Goldman, R. N. Nagel, and K. Preiss, *Agile Competitors and Virtual Organizations: strategies for enriching the customer*. New York: Van Nostrand Reinhold, 1995.

[2] Iacocca Institute, *21st Century Manufacturing Enterprise Strategy: An Industry-Led View*. Bethlehem, PA: Iacocca Institute, Lehigh University, 1991.

[3] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. R. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas, "Manifesto for Agile Software Development," 2001. [Online]. Available: http://agilemanifesto.org/

[4] A. Cockburn, *Agile Software Development: The Cooperative Game*, 2nd ed. Boston, MA: Pearson Education, 2007.

[5] J. Highsmith, *Agile Software Development Ecosystems*. Boston, MA: Pearson Education, 2002.

[6] P. Abrahamsson, K. Conboy, and X. Wang, "Lots done, more to do: the current state of agile systems development research," *European Journal of Information Systems*, vol. 18, no. 4, pp. 281–284, Aug. 2009. [Online]. Available: http://www.palgrave-journals.com/doifinder/10.1057/ejis.2009.27

[7] P. J. Agerfalk, B. Fitzgerald, and S. a. Slaughter, "Introduction to the Special Issue–Flexible and Distributed Information Systems Development: State of the Art and Research Challenges," *Information Systems Research*, vol. 20, no. 3, pp. 317–328, Sep. 2009. [Online]. Available: http://isr.journal.informs.org/cgi/doi/10.1287/isre.1090.0244

[8] R. Wendler and A. Gräning, "How Agile Are You Thinking? An Exploratory Case Study," in *Proceedings of the 10th International Conference on Wirtschaftsinformatik, WI 2.011*, no. February, Zurich, Switzerland, 2011, pp. 818–827.

[9] B. Sherehiy, W. Karwowski, and J. K. Layer, "A review of enterprise agility: Concepts, frameworks, and attributes," *International Journal of Industrial Ergonomics*, vol. 37, no. 5, pp. 445–460, May 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/

pii/S0169814107000236
[10] J. A. Simpson and E. S. C. Weiner, *The Oxford English Dictionary*, 2nd ed. Oxford: Oxford University Press, 1989.
[11] A. Gunasekaran and Y. Y. Yusuf, "Agile manufacturing: A taxonomy of strategic and technological imperatives," *International Journal of Production Research*, vol. 40, no. 6, pp. 1357–1385, Jan. 2002. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00207540110118370
[12] E. S. Bernardes and M. D. Hanna, "A theoretical review of flexibility, agility and responsiveness in the operations management literature: Toward a conceptual definition of customer responsiveness," *International Journal of Operations & Production Management*, vol. 29, no. 1, pp. 30–53, 2009. [Online]. Available: http://www.emeraldinsight.com/10.1108/01443570910925352
[13] K. Förster and R. Wendler, "Theorien und Konzepte zu Agilität in Organisationen," 2012.
[14] A. Ganguly, R. Nilchiani, and J. V. Farr, "Evaluating agility in corporate enterprises," *International Journal of Production Economics*, vol. 118, no. 2, pp. 410–423, Apr. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S092552730800385X
[15] R. Dove, "Knowledge management, response ability, and the agile enterprise," *Journal of Knowledge Management1*, vol. 3, no. 1, pp. 18–35, 1999.
[16] ——, *Response Ability: The Language, Structure, and Culture of the Agile Enterprise*. Hoboken, NJ: John Wiley & Sons, 2001.
[17] K. Conboy, "Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development," *Information Systems Research*, vol. 20, no. 3, pp. 329–354, Aug. 2009. [Online]. Available: http://isr.journal.informs.org/cgi/doi/10.1287/isre.1090.0236
[18] T. Parsons, R. Bales, and E. Shils, *Working Papers of the Theory of Action*. Berlin: Free Press, 1953.
[19] H. Sharifi and Z. Zhang, "A methodology for achieving agility in manufacturing organisations: An introduction," *International Journal of Production Economics*, vol. 62, pp. 7–22, 1999.
[20] H. Sharifi, G. Colquhoun, I. Barclay, and Z. Dann, "Agile manufacturing: a management and operational framework," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 215, no. 6, pp. 857–869, Jan. 2001. [Online]. Available: http://pib.sagepub.com/lookup/doi/10.1243/0954405011518647
[21] Z. Zhang and H. Sharifi, "Towards Theory Building in Agile Manufacturing Strategy - A Taxonomical Approach," *IEEE Transactions on Engineering Management*, vol. 54, no. 2, pp. 351–370, 2007.
[22] D. Vázquez-Bustelo, L. Avella, and E. Fernández, "Agility drivers, enablers and outcomes: Empirical test of an integrated agile manufacturing model," *International Journal of Operations & Production Management*, vol. 27, no. 12, pp. 1303–1332, 2007. [Online]. Available: http://www.emeraldinsight.com/10.1108/01443570710835633
[23] A. Gunasekaran, "Agile manufacturing: A framework for research and development," *International Journal of Production Economics*, vol. 62, no. 1-2, pp. 87–105, May 1999. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0925527398002229
[24] Y. Y. Yusuf, M. Sarhadi, and A. Gunasekaran, "Agile manufacturing: The drivers, concepts and attributes," *International Journal of Production Economics*, vol. 62, pp. 33–43, 1999.
[25] S. Meredith and D. Francis, "Journey towards agility: the agile wheel explored," *The TQM Magazine*, vol. 12, no. 2, pp. 137–143, 2000.
[26] A. Agarwal, R. Shankar, and M. Tiwari, "Modeling agility of supply chain," *Industrial Marketing Management*, vol. 36, no. 4, pp. 443–457, May 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0019850106000022
[27] D. Kisperska-Moron and A. Swierczek, "The agile capabilities of Polish companies in the supply chain: An empirical study," *International Journal of Production Economics*, vol. 118, no. 1, pp. 217–224, Mar. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0925527308002673
[28] T. Chow and D.-B. Cao, "A survey study of critical success factors in agile software projects," *Journal of Systems and Software*, vol. 81, no. 6, pp. 961–971, Jun. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0164121207002208
[29] S. C. Misra, V. Kumar, and U. Kumar, "Identifying some important success factors in adopting agile software development practices," *Journal of Systems and Software*, vol. 82, no. 11, pp. 1869–1890, Nov. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S016412120900123X
[30] F. K. Chan and J. Y. Thong, "Acceptance of agile methodologies: A critical review and conceptual framework," *Decision Support Systems*, vol. 46, no. 4, pp. 803–814, Mar. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167923608002133
[31] S. Sarker, "Exploring Agility in Distributed Information Systems Development Teams: An Interpretive Study in an Offshoring Context," *Information Systems Research*, vol. 20, no. 3, pp. 440–461, Aug. 2009. [Online]. Available: http://isr.journal.informs.org/cgi/doi/10.1287/isre.1090.0241
[32] P. Kettunen, "Adopting key lessons from agile manufacturing to agile software product developmentA comparative study," *Technovation*, vol. 29, no. 6-7, pp. 408–422, Jun. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0166497208001302
[33] J. Ren, Y. Y. Yusuf, and N. D. Burns, "A prototype of measurement system for Agile enterprise," in *Proceedings of the 3rd. International Conference on Quality, Reliability and Maintenance*, G. J. McNulty, Ed. Oxford: University of Oxford, 2000, pp. 247–251.
[34] N. C. Tsourveloudis and K. P. Valavanis, "On the Measurement of Enterprise Agility," *Journal of Intelligent and Robotic Systems*, vol. 33, pp. 329–342, 2002.
[35] C.-T. Lin, H. Chiu, and Y.-H. Tseng, "Agility evaluation using fuzzy logic," *International Journal of Production Economics*, vol. 101, no. 2, pp. 353–368, Jun. 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0925527305000514
[36] Y.-H. Tseng and C.-T. Lin, "Enhancing enterprise agility by deploying agile drivers, capabilities and providers," *Information Sciences*, vol. 181, no. 17, pp. 3693–3708, Sep. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0020025511002088
[37] A. T. Eshlaghy, A. N. Mashayekhi, A. Rajabzadeh, and M. M. Razavian, "Applying path analysis method in defining effective factors in organisation agility," *International Journal of Production Research*, vol. 48, no. 6, pp. 1765–1786, Mar. 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00207540802566410
[38] E. Bottani, "Profile and enablers of agile companies: An empirical investigation," *International Journal of Production Economics*, vol. 125, no. 2, pp. 251–261, Jun. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S092552731000068X
[39] A. Charbonnier-Voirin, "The development and partial testing of the psychometric properties of a measurement scale of organizational agility," *M@n@gement*, vol. 14, no. 2, pp. 120–155, 2011.
[40] P. P. Tallon and A. Pinsonneault, "Competing Perspectives on the Link Between Strategic Information Technology Alignment and Organizational Agility: Insights from a Mediation Model," *MIS Quarterly*, vol. 35, no. 2, pp. 463–486, 2011.
[41] P. J. Zelbst, V. E. Sower, K. W. Green Jr., and R. D. Abshire, "Radio Frequency Identification Technology Utilization and Organizational Agility," *Journal of Computer Information Systems*, vol. 52, no. 1, pp. 24–33, 2011.
[42] K. Breu, C. J. Hemingway, M. Strathern, and D. Bridger, "Workforce agility: the new employee strategy for the knowledge economy," *Journal of Information Technology*, vol. 17, no. 1, pp. 21–31, Mar. 2001. [Online]. Available: http://www.palgrave-journals.com/doifinder/10.1080/02683960110132070
[43] Gephi, "Gephi: The Open Graph Viz Platform." [Online]. Available: https://gephi.org/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1173–1180

Measuring the information society in Poland – dilemmas and a quantified image

Ewa Ziemba
University of Economics
ul. 1 Maja 50, 40-287 Katowice,
Poland
Email: ewa.ziemba@ue.katowice.pl

Rafał Żelazny
University of Economics
ul. 1 Maja 50, 40-287 Katowice,
Poland
Email: rafal.zelazny@ue.katowice.pl

***Abstract*—This paper focuses on measurement of information society in Poland. The aim of this paper is twofold. The first objective is to present a coherent picture of measurement methods for information society. The second aim of the paper is to show measurement findings of information society in Poland. Firstly, the paper presents available methods of information society measurement and a core set of internationally agreed information society indicators. Secondly, the measurement of information society in Poland has been performed with the application of two methods – measuring the influence of ICT on GDP and measuring ICT Development Index. Finally, a discussion has been undertaken in order to establish a framework for development of information society quantitative measurement methods in Poland.**

I. Introduction

INCREASING role of information, knowledge, information and communication technology (ICT) determines the complexity and variability of a social system and its sub-systems, and especially the economic one. Transitions of these systems have been reflected in many research concepts. According to the assumptions being put forward, the basic factors of socio-economic development are information and its derivative – knowledge.

The pioneering work in this field was undertaken by Bell who first used the term postindustrial society in Salzburg in 1959. By its means he denominated a society which transitioned from the stage of foods production to the stage of service society [1]. These studies were further developed by Bell in the direction of identifying the position of knowledge in social development [2]. The concepts of knowledge economy, knowledge industry, types of entities managing knowledge and types of knowledge were introduced to economic research by Machlup [3]. In parallel, e.g. at the beginning of 1960s the term information society came out in the Japanese social science [4]. At the end of the 1970s Drucker stressed the significance of transition to the so-called post-capitalist society, based on knowledge and knowledge economy [5]. He developed this idea in his further work by introducing the notion of knowledge economics [6]. On the basis of Bell's, Machlup's and Drucker's approaches Porat stemmed his research devoted to the information economy and information industry [7]. In the 1980s Toffler presented the idea of "the third wave" – post-industrial civilization where the basic resources are: information and ICT [8]. The informational manner of development of contemporary capitalist societies (network societies) based on ICT expansion, which creates the ground for a complete change of conditions and style of social life was studied by Castells [9], [10], [11]. Issues concerning information society and knowledge based society have become widely discussed in publications in Poland [12], [13], [14], [15], [16], [17], [18], [19], [20]. Economies and societies using information and knowledge, to extend unprecedented ever before, are denominated in various ways e.g. as based on knowledge, digital, post-industrial, new or information.

The researchers face many cognitive and empirical challenges referring to information society (IS). The cognitive challenges refer to terminology describing information society, identification of phenomena, processes and success factors of this society and also the methodology of information society measurement. The empirical challenges are mainly connected with building information society and its measurement. Research of this scope is conducted in the academic environment [20], [21], as well as among practitioners [22].

The measurement is an important issue in the debate about the information society and the role it plays in economic and social development [21], [20], [23], especially in transition and emerging economies. This paper focuses on measurement of information society in Poland. The aim of this paper is twofold. The first objective is to present a coherent picture of measurement methods for information society. The second aim of the paper is to show measurement findings of information society in Poland.

To achieve those aims, the paper takes the following structure. Firstly, the paper presents available methods of information society measurement and a core set of internationally agreed information society indicators. Secondly, the measurement of information society in Poland has been performed with the application of two methods – measuring the influence of ICT on GDP and measuring ICT Development Index. Finally, a discussion has been undertaken in order to establish a framework for development of information society quantitative measurement methods in Poland.

Hopefully, the achieved research findings can become useful in diagnosing information society, planning for information society undertakings as well as monitoring and evaluating the conducted undertakings.

978-83-60810-53-8/$25.00

II. Research Methodology

The primary objectives of the research required commencing work of theoretical and empirical characteristics. Various research methods were applied here. In order to present the methods of information society measurement, a critical analysis of foreign and Polish subject literature has been carried out as well as reports prepared by international organizations. The Internet statistical databases were explored at the 72 industry level for all non growth accounting variables, i.e. EU KLEMS [24]. Additionally, data from the International Telecommunication Union and European Statistical Office (Eurostat) were used for the measurement of the information society in Poland. The calculations, figures and tables were prepared in the Microsoft Excel program.

III. Theoretical Background – Information Society As Measurement Subject

To date there has not been in operation a commonly accepted definition of information society [2], [8], [4], [25], [14], [26], [18]. Lack of consensus with regard to the definition of information society is undoubtedly a derivative of complexity of processes taking place in a social system, characteristics of information as a resource, and the dynamics of ICT changes. This brings specific consequences for the undertaken attempts for measuring phenomena within the frame of a category, which might be and is understood in various ways.

Nonetheless, despite the conceptual limitations there are attempts taken to describe the information society quantitatively. These ideas were presented in the following sources:

- scientific monographies and papers, i.a. Machlup [3], Porat [7], Timmer, Inklaar, O'Mahony, van Ark [27], Dziuba [28], Goliński [20], Oleński [29], Batorski [30], Żelazny [21];
- reports and studies prepared by international organizations, i.a. International Telecommunication Union – ITU [31], [32], Organisation for Economic Cooperation and Development – OECD [33], United Nations – UN [34], European Union – EU [35], [36], World Bank [37], World Information Technology and Services Alliance – WITSA [38];
- reports of commercial organizations, i.a. World Economic Forum – WEF [39], International Data Corporation – IDC [40], Economist Intelligence Unit – EIU [41]; and
- monographies of national and trans-national services of public statistics and authorities, i.a. Central Statistical Office – GUS [42], Statistical Office of the European Union – Eurostat [43], Office of Electronic Communication – UKE [44], Ministry of Administration and Digitization of Poland – MAC [45], [46].

Generally speaking there are two approaches to the quantitative description of information society. The first one comprises the preparation of the list of indicators characterizing information society. The other is connected with compiling the so-called composite indexes which are aggregate measures. It should be stressed that the composite index is based on the previously chosen set of indicators. Some significant constraints can be pinpointed in both approaches. The arbitrariness of the choice of indicators, disorderliness of gathering source data, lack of standardization and time-space comparability, substantive errors in assigning indicators to specified information society dimensions and errors in constructing a given index – those are some of the significant drawbacks and constraints.

The above mentioned constraints gave rise to taking efforts on the international scale to institutionalize the methodology of information society quantification. Work in this field was commenced by OECD. In 1997 the OECD established the Working Party on Indicators for the Information Society, which main objective was development of index-based description of information society. One of its major achievements was identifying ICT sector. In 1998 an ICT sector definition was provided basing on the so-called International Standard Industry Classification (ISIC Rev. 3), according to which [33]:

- for manufacturing industries, (1) the products must be intended to fulfill the function of information processing and communication including transmission and display, and (2) the products must be use electronic processing to detect, measure and/or record physical phenomena or control a physical process; and
- for services industries, the products must be intended to enable the function of information processing and communication by electronic means.

Taking into account the above approach, the following were regarded as ICT industries: manufacture of office, accounting and computing machinery (3000), manufacture of insulated wire and cable (3130), manufacture of electronic valves and tubes and other electronic components (3210), manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy (3220), manufacture of television and radio receivers, sound or video recording or reproducing apparatus, and associated goods (3230), manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes, industrial process control equipment (3312 – 3313), wholesale of machinery, equipment and supplies (5150), renting of office machinery and equipment (including computers) (7123), telecommunications (6420) as well as computer and related activities (7200). The OECD's activity-based definition of ICT was slightly reviewed in 2002 (ISIC Rev. 3.1). The entry 5150 was replaced then by its components i.e.: wholesale of computers, computer peripheral equipment and software (5151), wholesale of electronic and telecommunications parts and equipment (5152).

One important feature of the ICT sector definition by OECD is that it breaks the traditional ISIC dichotomy between manufacturing and services activities. Activities producing or distributing ICT products can be found everywhere in the economy. Moreover, by identifying the key sectors whose main activity is producing or distributing ICT products, this definition constitutes a first order approxima-

tion of the "ICT producing sector". Hence, ICT producing sector means both ICT manufacturing industries (items: 3000, 3130, 3210, 3220, 3230, 3312, 3313) and ICT services industries (items: 5151, 5152, 7123, 6420, 7200) [33].

The following modifications of definitions of ICT sector resulted from the review of ISIC rev. 4 and ended in 2007. The presented above definition of ICT sector was narrowed in the part referring to manufacturing industries accounting only for activity and products which fulfill the function of information processing and communication including transmission and display [23], [47]. The present complete set of ICT sector is shown in [23].

An identical view of ICT sector can be found in the statistical classification of business activities in the European Union – Nomenclature statistique des Activités économiques dans la Communauté Européenne (NACE rev. 2), being in force from January 2008. To the ICT sector were included the following types of business activities: 261, 262, 263, 264, 268 (ICT manufacturing) and 465, 582, 61, 62, 631, 951 (ICT services) [48].

Another milestone in the development of information society statistics, after defining ICT sector, was the establishment of the Partnership on Measuring ICT for Development [49]. The participants of this forum became the following organizations and their agencies: ITU, OECD, Eurostat, United Nations Conference on Trade and Development (UNCTAD), UNESCO Institute for Statistics (UIS), World Bank, United Nations Department of Economic and Social Affairs (UNDESA), United Nations Economic Commission for Africa (ECA), United Nations Economic Commission for Latin America and the Caribbean (ECLAC), United Nations Economic and Social Commission for Asia and the Pacific (ESCAP), United Nations Economic and Social Commission for Western Asia (ESCWA), and United Nations Environment Programme/Secretariat of the Basel Convention (UNEP/SBC). As a result of the activities taken up by the Partnership on Measuring ICT for Development, a core list of ICT indicators was developed. The core list of ICT indicators is composed of over 50 indicators in the following areas:

- ICT infrastructure and access (A – 10 indicators);
- ICT access and use by households and individuals (HH – 12 indicators);
- ICT access and use by enterprises (B – 12 indicators);
- ICT sector and trade in ICT goods (ICT – 4 indicators);
- ICT in education (ED – 8 indicators); and
- ICT in government (EG – 7 indicators).

The list, which is revised regularly (the last time in 2012), was identified to help guide countries in measuring the information society. The full list of basic ICT indicators is available in [50].

The above indicators endorsed by the UN Statistical Commission are recommended as a measurement standard of information society on the international scale. As it has already been indicated, this set is a subject to supplementation and modification in response to the dynamic processes occurring in the economic and social environment. Such a consensual composition of a common set of indicators by major international institutions should be evaluated positively. A divergent issue stays the scope of implementation of this proposal in the statistical practice of the states, especially developing ones.

Indicatory description of the information society can be found in works of many organizations, both the members of the Partnership on Measuring ICT for Development, e.g. Eurostat, ITU, OECD or World Bank, and those remaining outside (WEF, IDC, EIU). These organizations collect and publish statistical data monitoring information society in various dimensions. Hence their proposals of composite indexes are an important element of their activities. As Goliński [20] argues the increasing popularity of composite indexes is connected with, among the others:

- ease of their interpretation and creation of prices on their basis;
- media attractiveness of composite indexes in relation to the necessity of conducting complex analyses based on single indicators;
- ICT development expediting the acquisition of statistical data, their processing and presentation; and
- demand for attractive tools expediting the evaluation of new socio-economic challenges.

Currently the most popular composite indexes measuring the information society are – ICT Development Index (IDI) of the authorship of the International Telecommunication Union and Networked Readiness Index (NRI) of the authorship of the World Economic Forum.

Considering the significance of works undertaken by the world oldest international organization – ITU on research and measurement of IS, and its active membership in the Partnership on Measuring ICT for Development, the further analysis was conducted on IDI. ITU experience in works on information society measurement was taken into account in the methodology for compiling this indicator. The theoretical framework for this indicator was based on the three-stage model for information society development, i.e. readiness, intensity and impact [31], [23]. The first stage – readiness – reflects the level of networked infrastructure and access to ICT. The second stage – intensity – reflects the level of use of ICTs in the society. The third stage – impact – reflects the result of efficient and effective ICT use. Therefore, the construction of IDI is based on three sub-indexes – access sub-index, use sub-index and skills sub-index. Relevant statistical dependence is presented in Table I.

The IDI was computed applying the following steps – preparation of the complete data set, normalization of data, rescaling of data and weighting of indicators and sub-indices. The IDI is currently calculated for 155 countries.

IV. Research findings – measurement of information society in Poland

A. Share of ICT producing sector in GDP in Poland

Evaluating the share of ICT sector in GDP the most current available data were used from Eurostat referring to 2009 [52], and the database of EU KLEMS [24] referring to the period of 1995-2006.

The value added at factor cost in the ICT sector as percentage of total value added at factor cost of the selected EU countries in 2009 is presented in Figure 1. The value added at factor cost is defined as gross value added (at basic prices) minus other taxes less other subsidies on production.

The lowest share of ICT in GDP (3.15%) was found in Poland among the researched countries (Figure 2). In the group of the Central and East European countries the best result was achieved by Hungary (5.93%). An interesting fact is that in the majority of the countries there was a drop in the share of ICT in GDP in the period of 2000-2009. The increase was only noted in case of Hungary (from 5.91% in 2000 to 5.93% in 2009) and in Bulgaria – from 4.63% in 2000 to 5.36% in 2008. A significant decrease took place in Finland – from 10.16% to 5.31%. [52]

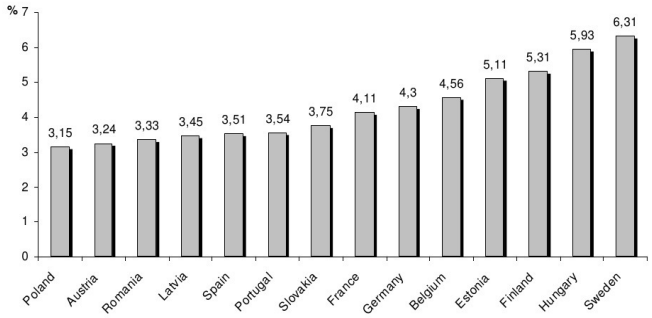


Fig. 1. Percentage of the ICT sector on GDP in selected EU countries with available data in 2009
Source: own study based on Eurostat data [48]

Accounting for the components of the ICT sector i.e. manufacturing industries and service industries, the major

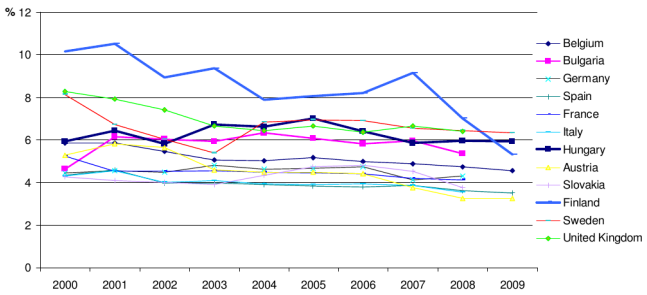


Fig. 2. Percentage of the ICT sector on GDP in selected EU countries with available data in 2000-2009
Source: own study based on Eurostat data [48]

role of business activities based on services needs to be emphasized in all countries. The only exception was Finland in the period of 2000-2007, when the share of ICT manufacturing industries in GDP was higher than the share of ICT service industries. In the Eurostat database there is a lack of data referring to the share of ICT manufacturing industries, as well as ICT service industries in Poland's GDP. In Poland the presented share of ICT sector in GDP at the level of 3.15% in Figure 1 took place in 2009 and was composed respectively of 0.35% manufacturing industries and 2.8% service industries. The share of net ICT sector revenues from sales in the total net sector revenues from sales was about 4.8% in 2009, 5.3% in 2010 and 5.1% in 2011 in Poland [42].

Manufacturing goods and providing ICT services directly influence the increase of the value added generated in the economy. The ICT influence on economic growth is calculated as a product of a nominal ICT producing sector share in GDP and a real output growth and provision of services by this sector. In order to estimate the ICT producing sector share in GDP one should: (1) select the period for an analysis, (2) on the basis of a chosen classification (in this paper ISIC Rev. 3) estimate the share of ICT producing sector in GDP, and (3) calculate the product of ICT producing sector share in GDP and the real growth rate of ICT producing sector. The result of using this algorithm is the value of ICT producing sector share in the GDP growth rate in percentage

TABLE I.
ICT DEVELOPMENT INDEX (IDI) – SUB-INDEXES, INDICATORS AND WEIGHTS

Sub-index	Weights (sub-indexes)	Indicators	Weights (indicators)	Reference value
ICT access	40%	Fixed-telephone lines per 100 inhabitants	20%	60
		Mobile-cellular telephone subscriptions per 100 inhabitants	20%	180
		International Internet bandwidth (bit/s) per Internet user	20%	408'813
		Percentage of households with a computer	20%	100
		Percentage of households with Internet access	20%	100
ICT use	40%	Percentage of individuals using the Internet	30%	100
		Fixed (wired)-broadband Internet subscriptions per 100 inhabitants	30%	60
		Active mobile-broadband subscriptions per 100 inhabitants	30%	100
ICT skills	20%	Adult literacy rate	30%	100
		Secondary gross enrolment ratio	30%	100
		Tertiary gross enrolment ratio	30%	100

Source: [31].

points. A suitable data and original calculations in this respect are presented in Table II.

B. ICT Development Index in Poland

As it has been mentioned earlier, the ICT Development Index (IDI) is very often used in order to measure the information society. The values of ICT Development Index, sub-indexes and individual indicators for the years 2011 and 2010 are presented in Table III.

C. Discussion of research findings

The measurement of information society in Poland has been conducted by applying two diagnostics approaches. The influence of ICT sector on GDP has been measured and the composite index – IDI has been presented.

The performed calculations and statistical data analysis proved a small share of ICT sector in Poland's GDP. The average share of ICT producing sector in GDP in the period of 1995–2006 in Poland constituted only about 17% of the total GDP growth rate, i.e. 0.62% out of 3.7%. The percentage of ICT producing sector value added (ICT manufacturing industries and ICT services industries) in GDP in the period of 1995-2006 equaled in real value 4.17% average. The ICT services industries decidedly dominated over the ICT manufacturing industries – the annual average in the period of 1995-2006 was at 2.9% and 1.3%. In 2009 it was respectively 2.8% and 0.35% with the total share of 3.15%. It proves a relatively weaker position of Poland in producing ICT (like hardware) in comparison to other countries. At the same time, significant difficulties were identified in getting to current data allowing for making appropriate calculations and international comparisons. Generally speaking, the attempts to study the ICT sector in Poland (even though they embrace the business entities with workforce over 10 persons) by the Central Statistical Office should be evaluated positively [42]. The access to data with regard to the number of enterprises and employees of the ICT sector, the size and structure of net revenues from sales, labor efficiency, operating costs of ICT sector, profitability of sales or import and export of ICT goods are essential, all the same it should be complemented by the measurement of this sector influence channels over economic growth, also at the regional level.

Taking into account the IDI in 2011, Poland occupied the 31st position out of 155 studied countries. With the value of the IDI equals 6.19 it took the 21st position among the studied European countries, and 17th among the EU countries. The theoretical maximum value of the indicator can amount to 10. In comparison to 2010 the result improved by 0.1, however in the global ranking Poland fell by one position. It is the result of faster development of the countries close to Poland with regard to information society development. South Korea opens the ranking with the IDI value equals 8.56, on the second position is ranked Sweden (8.34), and the third Denmark (8.29). Assuming for the particular sub-indexes (the maximum possible result–10), Poland achieved the best result in the field of skills, next access and in order – use. According to this method, the level of IS development in Poland, taking into account the group of developed countries, is moderate.

V. CONCLUSIONS AND FUTURE WORKS

This research can be useful for researchers and practitioners who are interested in measuring information society. It suggests important issues for measuring information society. The replication of this study in emerging and developing countries will be useful to improve their knowledge related to information society, its measurement and its monitoring.

Both diagnostic approaches to the information society measurement have benefits and drawbacks. Manufacturing goods and providing ICT services directly increase the value added generated by an economy. However, the calculation of ICT service industries and ICT manufacturing industries share in GDP is mainly based on hardly accessible historical data on the international scale. Apart from that there is the necessity of accounting for the qualitative dynamic changes and using deflators allowing for these changes. Their use allows for calculating prices proportionate to the changes in ICT products and services quality. The ICT producing sectors identification by itself and on the regular basis accounting for changes in the methodology of calculations are the steps in the right direction, heading to diligent measurement of information society. They allow for conducting comprehensive estimates of the values of the sector in particular countries and conducting trans-national comparisons.

Despite the advantage of IDI over other proposed composite indexes (e.g. NRI) with respect to methodological correctness it cannot be used for the complex evaluation of information society in a given country. It is worth to notice that in the construction of IDI just few indicators from the core list of ICT indicators were used. The compatibility of some indicators to the description of IDI sub-index seems to be disputable, e.g. the percentage of households with a computer indicator to the ICT access characteristics, or the adult literacy rate indicator to the ICT skills. The weighting of selected sub-indexes for the IDI calculation also pose some doubts. Lower weighting for ICT skills is explained by the adoption of proxy indicators with regard to the absence of more targeted indicators, such as ICT literacy. Taking into account the methodology applied to the Principal Components Analysis (PCA) such an approach seems to be controversial [32].

The methodology of information society measurement showed in this research should be explored in greater depth. In the opinion of this paper authors', in works on the measurement of information society, the critical success factors for implementing information society in a given country or region should be accounted for. For every identified factor, an indicator or indicators should be pointed which will allow for its quantitative description. Surely, such an approach will provide for reflecting on current issues of information society implementation. Simultaneously, it may turn out to be helpful in modification of the existing methods of the information society measurement. Such research is conducted by the authors.

TABLE II.
CALCULATING OF ICT PRODUCING SECTOR IN GDP GROWTH IN POLAND IN 1995–2006

		POLAND (gross value addend in constant prices from 1995 in mln PLN											
Code as per ISIC rev. 3	**Business activity**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
30	Office, accounting and computing machinery	195	244	244	339	592	882	562	501	599	721	1 268	918
313	Insulated wire	441	547	569	644	673	751	747	739	1,018	1,088	1,157	1,726
321	Electronic valves and tubes	139	178	223	215	206	202	223	161	162	268	282	398
322	Telecommunication equipment	423	466	577	674	811	753	867	660	538	713	700	688
323	Radio and television receivers	572	594	719	827	956	874	1,088	985	1,127	1,325	1,247	1,754
331	Scientific instruments	889	1,179	1,372	1,582	2,041	1,851	1,755	1,522	1,555	1,804	2,066	2 341
64	Post and telecommunications	5,048	5,221	6,399	7,197	7,445	7,485	9,040	10,891	10,964	13,155	12,466	13,377
72	Computer and related activities	825	1,209	1,104	1,419	1,600	1,809	2,173	2,441	2,743	3,133	2,731	3,372
∑ ICT producing value added		8,531	9,637	11,207	12,896	14,324	14,607	16,455	17,899	18,706	22,207	21,918	24,574
∑ gross value added		297,702	314,547	334,663	350,346	364,925	378,448	382,113	387,138	400,580	418,880	432,043	457,294
nominal share of ICT producing sector in GDP (%)		2.87	3.06	3.35	3.68	3.93	3.86	4.31	4.62	4.67	5.30	5.07	5.37
GDP growth rate (w %)		–	5.66	6.40	4.69	4.16	3.71	0.97	1.31	3.47	4.57	3.14	5.84
average share of ICT producing sector in GDP (%)		**4.17**											
ICT producing sector annual growth rate (%) – Törnqvist index		14.80											
real GDP annual growth rate (%)		3.70											
average share of ICT producing sector in GDP growth (in percentage points)		**0.62**											

Source: own study based on EU KLEMS database [24]

TABLE III.
VALUES OF PARTICULAR IDI COMPONENTS FOR POLAND IN 2011 AND 2010

IDI /position in 2011	Sub-index	Sub-indexes /position in 2011	Indicators	2011	2010
6.19/31	ICT access	6.46/43	Fixed-telephone lines per 100 inhabitants	18.1	20
			Mobile-cellular telephone subscriptions per 100 inhabitants	128.5	122.7
			International Internet bandwidth (bit/s) per Internet user	40'244	37'729
			Percentage of households with a computer	73	69
			Percentage of households with Internet access	66.6	63.4
	ICT use	4.57/32	Percentage of individuals using the Internet	64.9	62.3
			Fixed (wired)-broadband Internet subscriptions per 100 inhabitants	14.4	13
			Active mobile-broadband subscriptions per 100 inhabitants	48.4	50
	ICT skills	8.89/17	Adult literacy rate	99.5	99.5
			Secondary gross enrolment ratio	97	97
			Tertiary gross enrolment ratio	70.5	70.5

Source: own study based on statistical data from [31]

ACKNOWLEDGEMENTS

This research has been supported by a grant entitled "Designing a system approach to sustainable development of the information society – on the example of Poland" from the National Science Centre in Poland, 2011/01/B/HS4/00974, 2011-2014.

REFERENCES

[1] M. A. Rose, *The post-modern and the post-industrial: a critical analysis*. Cambridge: Cambridge University Press, 1991, pp. 170.
[2] D. Bell, *The coming of post-industrial society: A venture in social forecasting*. New York: Basic Books, 1973.
[3] F. Machlup, *The production and distribution of knowledge in the United States*. Princeton: Princeton University Press, 1962.
[4] L. Z. Karvalics, *Information society – what is it exactly?* Budapest: Network for Teaching Information Society, 2007.
[5] P. F. Drucker, *The age of discontinuity: guidelines to our changing society*. New York: Harper and Row, 1968.
[6] P. F. Drucker, *Post-capitalist society*. New York: Harper Business, 1993.
[7] M. U. Porat, *The information economy*, vol. 1. Washington DC: Office of Telecommunications, US Department of Commerce, 1977, pp. 44.
[8] A. Toffler, *The third wave*. New York: Bantam Books, 1980.
[9] M. Castells, *The information age: economy, society and culture. The rise of network society*, vol. 1. Oxford: Blackwell Publishers, 1996.
[10] M. Castells, *The information age: economy, society and culture. The rise of network society*, vol. 2. Oxford: Blackwell Publishers, 1997.
[11] M. Castells, *The information age: economy, society and culture. The rise of network society*, vol. 3. Oxford: Blackwell Publishers, 1998.
[12] A. Szewczyk, Ed. *Społeczeństwo informacyjne. Problemy rozwoju*. Warszawa: Difin, 2007.
[13] J. Papińska-Kaceperek, Ed. *Społeczeństwo informacyjne*. Warszawa: PWN, 2008.
[14] C. M. Olszak, and E. Ziemba, Eds. *Kierunki rozwoju społeczeństwa informacyjnego i gospodarki opartej na wiedzy w świetle śląskich uwarunkowań regionalnych*. Katowice: Akademia Ekonomiczna, 2010.
[15] P. Sienkiewicz, and J. S. Nowak, *Społeczeństwo informacyjne. Krok naprzód, dwa kroki wstecz*. Katowice: Polskie Towarzystwo Informatyczne, 2009.
[16] R. Żelazy, "Determinanty rozwoju gospodarczego Polski w aspekcie koncepcji gospodarki opartej na wiedzy," in *GOW – wyzwanie dla Polski*, J. Kotowicz-Jawor, Ed. Warszawa: Polskie Towarzystwo Ekonomiczne, 2009.
[17] E. Ziemba, T. Papaj, and R. Żelazny, "New perspectives on information society: The maturity of research on a sustainable information society," *Online Journal of Applied Knowledge Management*, vol. 1, issue 1, pp. 52–71, 2013.
[18] E. Ziemba, "The holistic and systems approach to the sustainable information society," *Journal of Computer Information Systems*, to be published.
[19] E. Ziemba, and C. M. Olszak, "Building a regional structure of an information society on the basis of e-administration," *Issues in Informing Science and Information Technology*, vol. 9, pp. 277–295, 2012.
[20] M. Goliński, *Społeczeństwo informacyjne. Geneza koncepcji i problematyka pomiaru*. Warszawa: Oficyna Wydawnicza Szkoły Głównej Handlowej, 2011.
[21] R. Żelazny, "Wybrane mierniki rozwoju społeczeństwa informacyjnego i gospodarki opartej na wiedzy. Problemy pomiaru na poziomie regionalnym," in *Kierunki rozwoju społeczeństwa informacyjnego i gospodarki opartej na wiedzy w świetle śląskich uwarunkowań regionalnych*, C. M. Olszak, and E. Ziemba, Eds. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego, pp. 48-57, 2010.
[22] *Raport monitoringowy Strategii Rozwoju Społeczeństwa Informacyjnego Województwa Śląskiego do roku 2015*. Katowice: Śląskie Centrum Społeczeństwa Informacyjnego, 2013, retreive from: http://www.e-slask.pl/files/zalaczniki/2013/04/09/1276770448/1365509299.pdf, 2013.
[23] *OECD Guide to Measuring Information Society 2011*. Paris: OECD, 2011.
[24] retrieve from: http://euklems.net/, 2012.
[25] R. Mansel, *The information society. Critical concepts in sociology*. London: Routledge, 2009.
[26] D. R. Raban, A. Gordon, and D. Geifman, "The information society. The development of a scientific specialty," *Information, Communication & Society*, vol. 14, issue 3, pp. 375–399, 2011.
[27] M. P. Timmer, R. Inklaar, M. O'Mahony, and B. van Ark, *Economic growth in Europe. A Comparative Industry Perspective*. Cambridge: Cambridge University Press, 2010.
[28] D. T. Dziuba, *Sektor informacyjny w badaniach ekonomicznych*. Warszawa: Difin, 2010.
[29] J. Oleński, *Ekonomika informacji. Podstawy*. Warszawa: PWE, 2001.
[30] D. Batorski, "Korzystanie z technologii informacyjno-komunikacyjnych," in *Diagnoza społeczna 2011. Warunki i jakość życia Polaków*, J. Czapiński, and T. Panek, Eds. Warszawa: Rada Monitoringu Społecznego, 2011.
[31] *Measuring the Information Society 2012*. Geneva: International Telecommunication Union, 2012.

[32] *Measuring the Information Society. The ICT Development Index*. Geneva: International Telecommunication Union, 2009.
[33] *Measuring the Information Economy*. Paris: OECD, 2002.
[34] *The Global Information Society: a statistical view*. Santiago-Chile: Partnership on measuring ICT for development, United Nations, 2008.
[35] *Benchmarking Digital Europe 2011-2015 a conceptual framework*. *i2010 High Level Group*. European Union, 2009.
[36] *BISER eEurope Regions Benchmarking Report*. European Community, 2004, retrieve from: http://www.biser-eu.com, 2009.
[37] *Knowledge Assessment Methodology*. World Bank, retrieve from: http://www.worldbank.org/kam, 2013.
[38] *Digital Planet 2010. The Global Information Economy*. WITSA, 2010, retrieve from: http://www.witsa.org/v2/media_center/pdf/DP2010_ExecSumm_Final_LoRes.pdf, 2012.
[39] S. Dutta, and B. Bilbao-Osorio, Eds. *The Global Information Technology Report 2012. Living in a Hyperconnected World*. Geneva: World Economic Forum, 2012.
[40] *Information Society Index*. International Data Corporation, retrieve from: http://www.idc.com/groups/isi/main.html, 2012.
[41] *Digital economy rankings 2010. Beyond e-readiness*. London, New York, Hong Kong, Geneva: Economist Intelligence Unit, IBM Institute for Business Value, 2010.
[42] *Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2008-2012*. Warszawa: GUS, Urząd Statystyczny w Szczecinie, 2012.
[43] *Information society statistics – Statistics Explained*. Brussels: Eurostat, 2013, retrieve from: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Information_society_statistics, 2013.
[44] *Raport pokrycia terytorium Rzeczypospolitej Polskiej istniejącą infrastrukturą telekomunikacyjną, zrealizowanymi w 2011 r. i planowanymi w 2012 r. inwestycjami oraz budynkami umożliwiającymi kolokację*. Warszawa: Urząd Komunikacji Elektronicznej, 2012.
[45] *Społeczeństwo informacyjne w liczbach*. Warszawa: Ministerstwo Administracji i Cyfryzacji, 2012.
[46] *Badanie wpływu informatyzacji na działanie urzędów administracji publicznej w Polsce w 2011 roku*. Warszawa: ARC Rynek i Opinia na zlecenie MSWiA, sierpień 2011.
[47] *Information Economy – Sector definitions based on the International Standard Industry Classification (ISIC 4)*. Paris: OCED, Working Party on Indicators for the Information Society, 2006, retrieve from: *www.oecd.org/dataoecd/49/17/38217340.pdf* , 2012.
[48] retrieve from: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tin00074&plugin=1, 2012.
[49] *Partnership on Measuring ICT for Development. 2004 Project Document*. Retrieve from: http:// http://www.itu.int/en/ITU-D/Statistics/Documents/partnership/Partnership_Project_June2004.pdf, 2013.
[50] retrieve from: http://www.itu.int/ITU-D/ict/coreindicators/ index.html, 2012.
[51] *Report of the Partnership on Measuring Information and Communication Technology for Development*. United Nations Economic and Social Council, December 2011.
[52] retrieve from: http://appsso.eurostat.ec.europa.eu/nui/show.do ?dataset=isoc_bde15a g&lang=en, 2013.
[53] retrieve from: http://epp.eurostat.ec.europa.eu/tgm/refreshTableAction.do?tab=table&plugin=1&pcode=tin00074&language=en

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1181–1187

The outcomes of the research in areas of application and impact of software agents societies to organizations so far. Examples of implementation in Polish companies.

Mariusz Żytniewski, Andrzej Sołtysik, Radosław Kowal
University of Economics in Katowice ul. 1 Maja 50 , 40-287 Katowice, Poland
Email: {zyto, soltys, radek}ue.katowice.pl

***Abstract*–The development of information management systems stimulates the search for new forms of supporting business processes which take place in the organization. One of the solutions that can be applied here is software agents that enable support activities to the employee and the customer promoting information and knowledge about the organization. Such solutions are commercially available for several years in the form of interface agents, but there is insufficient research on the modeling, the applications and the impact on the organization and its environment. The purpose of this paper is to present theoretical research in this area regarding companies providing such solutions on Polish territory.**

I. Introduction

THE development of information management systems stimulates the search for new forms of supporting business processes. In addition to the systems aimed at processing and distribution of data and information, organizations are increasingly looking for solutions that support the processing of the knowledge held by their staff and its environment.

Recent research in this area indicates that such solutions should be built in the current concept of WEB 3.0, based on methods of semantic knowledge representation [10] and support the various stages of the business process. One of such solutions may be software agents. Knowledge driven multiagent systems are one of the current trends in the development of software agent technologies [1].

The primary goal of this scientific research is the creation of new trends in the development of the concept of society of software agents in terms of ubiquitous communication for the purpose of supporting and improving business processes in knowledge-based organizations involving semantic solutions.

Such solutions can support the process of disseminating information and knowledge about the organization, contributing to the improvement of processes occurring in them by improving customer satisfaction and affecting him, not only during the implementation of the business process such as sales, but can strengthen ties with customers by presenting information and transferring intra-organizational knowledge also not related to the process. These solutions, built on elements of artificial intelligence, need to equip them with adequately prepared and codified knowledge, which allows the user to communicate with the system.

The structure of the knowledge of the system and its features goes beyond the existing expert systems and other information systems which means that, in the case of its construction, it is necessary to search for new methods of their modeling and implementation.

One aspect of the research conducted by the authors was to demonstrate the approach of Polish companies regarding the construction of such solutions. For this purpose, a series of interviews with organizations based on the knowledge, that are dealing with programming agents during performing their everyday operations, had been conducted. This study consists of a diagnosis of key areas of application software agents in knowledge-based organizations. As an introduction to this research, organizations were divided into two groups: manufacturers or distributors and users of the software. At the same time the assumption was made that the manufacturers may also be the users, using this type of solution in the course of their business. The first stage of the qualitative research was the analysis of areas in which organizations creating or distributing this type of solutions support business processes and the problems faced by such organizations in the process of doing so. At the moment, we have conducted structured interviews with five larger manufacturers of this type of solutions in Poland. Some companies did not agree for an interview.

Interviews with representatives of those organizations were supposed to help in answering the following questions:

- To what extent software agents society can find its application in knowledge-based organizations?
- In what areas software agents society can support the business processes?
- What factors influence the limited applicability of the concept of software agents society?
- How is it possible to use intra-organizational knowledge in the process of achieving the objectives of software agents society?
- How to augment the organization's knowledge resources with the participation of agent technologies?

978-83-60810-53-8/$25.00

In the first part of the article the basic typologies of agent systems will be shown in the context of their intra-organizational use. The second part will present partial results of research on the issue of modeling agent solutions in the context of companies implementing such software in organizations that observe the need for computer support for the process of identification, codification, distribution and dissemination of knowledge in the area of business process execution.

II. Agent Technologies in organizations

Due to the characteristics of agents, i.e. autonomy, ability to recognize the context, pro activity and reactivity, and the ability to communicate and interact, the agents are mainly used in tasks that require distributed processing and exchange of information. The most frequently mentioned examples of agent-based approaches may include supporting the circulation of electronic documents, distributed problem solving, e-business and the use of agents on the Internet. However, they are not sufficient, as shown by the current study conducted by the authors, regarding application of software agents in knowledge-based organizations [6].

It is difficult to discern a consensus about definitions and a similar position among researchers and authors of various works. In terms of the life cycle of knowledge management systems, the approach presented here refers to the stage of dissemination and promotion of intra-organizational knowledge. One of the most interesting classification of agent systems applications are shown by Paprzycki in his work [5]. Referring to the use of software agents, he proposes the division of a sphere of existing applications of agents into three classes.

The first group includes the use of agents playing a role of components of distributed systems, in particular, those which are associated with Internet use. In this group, one can specify searching agents which mission is to find a well-defined information based on a specific sequence of keywords, or on the basis of questions posed in natural language.

The second group of agents mentioned in this scheme is used as a tool for modeling complex systems. Agents are seen as mechanisms that allow for mapping and modeling of real world phenomena. In particular, in this group business process modeling should be distinguished.

The third group are agents used to manage and personalize information supporting the user by means of, for example, animated characters. Now, thanks to continuous development of IT tools using this kind of technology becoming increasingly common, especially in systems using sites or web portals, which have become the natural environment of operation for interface agents or chatter bots. Agents are often inspired by their living counterparts. The animation uses movie clips and images of people who become a prototype of the agent. This type of user agents support services such as banking, corporate web pages, and wide range of online shops. Typically, they are built on the basis of aiml language i.e. [7].

This type of agent-based solutions can be divided into the following groups [2]:

Purchase agents (shopping agents) represent particular interests of customers by searching for the best offer for the customer and facilitating the process of making a purchase in the online store. Agents in this role shall make, on the client's behalf, a selection of a commercial information available via the Internet. Often it is possible to use them to formulate orders and finalization of commercial transaction without direct user intervention.

Selling agents represent the interests of the sellers and are used to streamline the sales process. Frequently such agents are supported by animated characters which role is to ensure the communication with customers in a natural language or close to natural. Agents participate in the whole business transaction, or in some stages. It is possible to conduct negotiations between purchase agents and selling agents. Selling agents have the ability to tailor their offerings to the needs posed by potential customers – the diagnosis takes place through dialogue with the customer, or personalized user profiles.

Marketing agents gather, through dialogue or searching the web, all available customer information and analyze it using statistical and econometric methods to optimize and allow preparing marketing campaigns targeting specific customers. Also they are used to adjust a supply to the strategy and the expectations of the market.

Virtual assistants support the user in the search for a specific item, or while visiting a particular site. Agents of this type can: represent the company by answering questions asked by clients, give visitors advise about how to navigate through the website, or help with the choice of a particular product. Agents have mechanisms which help them in searching for information/messages useful for the customer. Assistants support the promotion of new products as well. Software agents should support natural language processing and generate explanations for customers [8].

Specified typologies, related to the perception of agents in the context of the solutions that are part of information systems, can be extended on the basis of the research on agent-based support for knowledge-based organizations and are the easiest type of solutions in the context of building software agents society [10].

The classifications discussed here are related to modeling software agents and indicate that we should consider them in terms of solutions supporting and perfecting processes within the organization. These solutions allow the distribution of codified intra-organizational knowledge in the course of and besides the implemented business processes. They are able to become not only a part of the information system, but also a knowledge management system.

As indicated by L. Dignum and V. Abecker [4], from the point of view of the organization, there is an inherent dichotomy between the goals of the business processes and the objectives of knowledge management processes – employees always try in the first place to achieve business goals,

because they are to be held accountable for its execution. Taking care of organizational knowledge, or adding to the common pool of knowledge, its organization or even browsing is perceived as the goal of much less importance then effective performance of business-related tasks. Therefore, for effective implementation of knowledge management it is necessary to emphasize its importance for achieving the business objectives of the organization.

Knowledge management addresses issues that change over time – all of the elements of an organization, affecting the knowledge management system, is a subject to change [4]. Therefore, it is difficult to design and implement a system that is at the same time generic, equally useful for all departments and members of the organization and will be evolving without losing its usefulness. Another disadvantage is the fact that very rare knowledge management system is being implemented at one time across the whole organization – in majority of cases an incremental approach is used, when system is at first implemented within a single department, and then made available to the others.

Taking into account these assumptions about requirements and challenges for the design of modern knowledge management systems it must be noted that such systems should have a built-in ability to adapt to a changing environment. At the same time, these changes can affect both IT infrastructure – especially considering the trends associated with the growing use of mobile devices for accessing the shared resources of knowledge, as well as the changing needs of users whose demand for knowledge continues to grow along with the participation of knowledge in the creation of value added. This trend, involving the use of mobile devices also affects a second important aspect related to the design of modern knowledge management systems. It is the growing importance of the social nature of knowledge which in recent years has been becoming increasingly important, mainly due to the prevalence of social media [3].

Considering the increasing demand for the interaction between humans results in an increased interest on the role of information systems, including knowledge management systems, in building relationships. On the other hand, global competition and the increasing rate of information processing necessitates the development efforts related to the enrichment of knowledge management systems by adding user context, what is supposed to result in their personalization – see the above-mentioned demand for knowledge management systems adaptation to the specific needs of users, related to their organizational roles. Always present is also a need for systems that will not only provide adequate information, to the right people at the right time, but also will work proactively towards enhancing audience's creativity. The development trends of ubiquitous processing and geolocation indicate the need for the study on the integration of knowledge management systems into the human's environment, so that they become an integral and, from the point of view of the user, invisible part.

As pointed out, an interesting and increasingly wide spreading use of software agents is supporting activities related to the overall communication with the customer. Software agents, in the form of virtual advisors, appear on the websites of companies, taking on the role of a seller who is able to enter into dialogue with the customer, assess and identify customer needs and propose the solution in form of a particular product. Of course, the virtual advisor is also able to properly advertise the product. Similarly, though in a slightly different fragment of marketing communications, agents providing after-sales service work. Their job is to answer customer questions related either to the product or service and receiving complaints and / or propose a solution to the problem. As a result these solutions are used not only to provide knowledge regarding business processes during their execution, but they can help users who are looking for answers to questions that go beyond the core business of the organization. Thus, in terms of the construction of modern knowledge management systems, software agents can be a coherent part of such systems, supporting the different stages of its life cycle and actively supporting the knowledge-based organizations.

The issues related to the software agents usage indicate that modeling software agents society is not a trivial task and requires reference not only to the knowledge about the architecture of information systems, but also to the model of knowledge for such system. This knowledge model will not only serve as a knowledge structure for an agent, but will provide links to the knowledge embedded into other information systems of an organization, which is of great importance especially in the context of building heterogeneous software agents society [10].

III. The Outcomes of the Research So Far

The advancement in the concept of software agents society and a variety of approaches to its architecture and methods of construction, caused that the research has been focused on providing the theoretical framework for building the multiagent solutions that offer support for the organizations, in particular, knowledge-based organizations. Such agent societies support the processing and distribution of information and knowledge using the mechanisms of semantic knowledge representation and operate in context of ubiquitous communication. Four of the indicated assumptions limited our research to the application of agent-based solutions in supporting the human – computer interaction and in the context of their possible use as part of knowledge management systems.

As indicated earlier, these solutions can be considered in the context of interface agents, each with its own codified knowledge base, associated with the information systems of an organization and actively participating in the ongoing business processes. Analysis of commercially available solutions pointed out that the vast majority of applications of agents in organizations are "Virtual advisors". All surveyed organizations offer such solutions. Various solutions, how-

ever, differ from each other significantly. Most of the solutions with primary goal in the form of finding and providing information to users, use a limited set of questions, that they are able to recognize, so their functionality is very limited. As a mean to overcome this limitation some solutions use tips in the form of 'choose-from-list' mechanism. One of the analyzed solutions uses natural language analysis speech that is used in interpreting user's statements, coupled with the speech recognition and biometric methods, which makes the solution more user-friendly, efficient, and significantly expands the range of other possible applications and makes them more versatile. One of the producers also highlights the special feature of their solution, which is the ability to focus on the essence of a "conversation" and the ability to return and resume the interrupted thread. Another solution integrates seamlessly with the entire service platform designed for supporting the public sector. Among other features automatic handling of callers to the call center can be distinguished. This solution, as the only one, uses a greater number of agents at the same time. Despite the focus on the same topic, they don't cooperate, nor create a society. However, it is possible to examine the software agents acting individually.

A. The benefits of using the software agents in an organization.

Organizations using software agents take advantage of a wide variety of benefits, among which the most frequently mentioned is an improvement for internal and external communication channels. The main functionality of virtual advisors is to provide the end customer with instant access to information without having to call the hotline or search for a specific Web page. This is undoubtedly the improvement of the process of communication and provides new channels of communication for external users. Virtual advisor also facilitates the usage of internal applications for employees. It provides a kind of help desk, which is the first line of support, and eliminates the need for traditional help files. An employee who is not proficient in terms of the usage of the software may simply ask a question related to the functionality of the software, and agent gives him or her the expected answer. Given this it surely improves the internal communication and speeds up access to information and knowledge for internal stakeholders. The use of software agents in a call center to answer the most common customers' questions brings measurable benefits for the organization, by reducing the cost of its service and by increasing customer satisfaction.

B. Software agents as a tool for perfecting business processes in the organization.

With regard to perfecting business processes in the organization of various types of software agents are utilized in many different ways. The use of agents typically require clarification and formalization of the knowledge processing or even the introduction of knowledge management given the fact that the virtual advisor requires actual knowledge. The most obvious seems to be looking for improvement in the distribution of knowledge, e.g. the provision of information, and the customer service process, especially in those organizations where existing procedures are the most formal and well documented.

Agent interface representing intelligent search mechanism, streamlines the process of handling an applicant in the office, which, due to its use, becomes more efficient and faster. Similarly, software agent can support internal users which execute formal, well-defined and well-described processes, by reducing the time and improving access to information necessary for the resolution of the problems. Thus consulting the agent will optimize the business processes.

Organizations recognize another opportunity to improve business processes in running the trainings via the Internet supported by software agent, what makes it possible to carry out an individual training program, which depends on the needs and expectations of participants. Despite the fact that the lecture is run by a machine the participant may, in the course of his or hers study, receive help in the form of answers to his or hers questions. According to a representative of one of the companies participating in the survey it is a significant improvement in the business process. Software agents are also reported to bring benefits in the case of supporting the process of taking orders in an online shop. The survey results include a description of sales service process improvement in which information software agents are used to replace the traditional communication channels like inquiries, E-mail, or instant messaging like Skype or MSN Messenger, in situations where it is necessary to identify and/or specify the technical parameters of equipment sold or to clarify some definitions. The customer can get the answer on-line from the agent. In addition to the obvious benefits of cost savings and speeding up the process, companies indicated a higher attractiveness of this form of assistance.

C. Factors affecting the limited applicability of the concept of software agents society in the organization

Organizations share the same view when it comes to the identification of key factors influencing the limited applicability of the concept of software agents society in the organization. Typically, respondents indicated three main factors.

As a main reason social factor was indicated, i.e. lack of awareness of the existence of such solutions and potential benefits of its application, or perceiving them as a toy, amusing marketing gadget on the site, and not the real support. Both organizations and consumers are not yet ready for the introduction of software agents. Usually it is related to a misunderstanding of its purpose, as stated above, and a fear that stems from believing that they are a tool for spying on the users. Users are also not convinced about the reasonableness of the use of such mechanisms. Agent systems producers count on the fact that the situation will change over time, what will result in enhancing the intelligence of these solutions, which will enable them to assertively respond to user's behavior and intelligently adapt to his way of thinking.

Manufacturers also point to the problems that affect the older part of the population which is related to using the keyboard to communicate with the agents. Vendors try to develop new interfaces that eliminate the issue of typing. Probably the most interesting thing happening nowadays, regarding that matter, is the introduction of speech recognition and touch screens that release older people from necessity of using a keyboard. With speech recognition fluent conversation with the agent will be possible, rather than using written text.

Important factors limiting the implementation of software agents are organizational problems. We identified several of such problems that are the most common organizational factors hindering the use of agent-based solutions:

- lack of preparation of the organization for the introduction of agent systems, arising from organizational culture,
- lack of appropriate procedures,
- lack of separated and unambiguous definitions of business processes.

Manufacturers assume that a wider promotional campaign carried out by both the companies offering these solutions, as well as institutions that procure them, would result in their wider adoption in the organizations.

Problems related to knowledge management in organizations which plan to implement software agents are also indicated. There are potential users, organizations that would gladly implement such solutions, which have the necessary budget, are aware of the interest of the customers, but they realize that they are unable to adequately manage the knowledge, deliver it on time or in the correct amount that could be used by an virtual adviser.

An important factor limiting the applicability of agent solutions are technical and technological conditions. We distinguish two groups of such conditions. The first one is the constraints in the network infrastructure, making it difficult to make service call. For example, "virtual advisors" are the solutions which depend heavily on the access to the Internet connections. In Poland, the network infrastructure is still not sufficiently developed, what makes it difficult for everyone to gain access to and communicate with the virtual consultant. Residents of highly urbanized areas, usually the big cities, do not have this problem, while those living in poor urban areas, country side and terrain such as mountains may experience difficulties in accessing the Internet, which may be one of the main technical barriers.

Although informational technology is a subject for constant development, despite the significant progress and enormous achievements in the field of artificial intelligence, from the point of view of the technology that is used, we still are not able to create solutions which "thinks" like a human being and is able to "come up with" something to talk about on its own. Agent may conclude faster and much "better" than a man, it can make use of an analytical knowledge it has accumulated, but it is not in a position to come up with the subject of conversation. It can only base on pre-defined themes.

Most manufacturers indicate inadequacy or lack of mechanisms that would allow for a complete "mapping" features of the human brain that would allow for their repetition by the machine. Surely a computer, the machine, the computer program will neither act and "think" in a way the human being does, nor operate on the basis of associations. Most manufacturers also point to inadequacy or lack of tools for effective speech recognition, which, in their view, is blocking the applicability of the available solutions. Those that are currently available on the Polish market are still not mature enough and do not allow free communication and unambiguous speech recognition. Most of these tools allow to communicate in English, and users of agent systems, especially the virtual advisors, rely mostly on Polish, which is country's official language. It is possible, as shown by some manufacturers, to unblock the communication channel with use of software agents, which in turn should greatly improve their adaptation.

Another blocking factor, indicated by the respondents, is economic in its nature – namely agent solutions are not cheap. On the other hand the survey revealed that it is possible to overcome this one by applying for different types of subsidies that offer financing the implementation of such innovative solutions.

D. Use of intra-organizational knowledge in the process of achieving the objectives of software agents.

Key activities in the utilization of intra-organizational knowledge recognized by respondents comprise any action performed in order to collect and organize the knowledge base, for example creating a repository of knowledge, which can be used in the future. Filling the agent system with a domain knowledge requires the collection of this knowledge in different places and in different ways. Accumulated knowledge should be structured, but in different organizations, such structure will be achieved differently.

Agent systems vendors often report a lack of proper knowledge and preparation on the customer's side and difficulties in persuading the client to systematize the knowledge necessary for the functioning of the system.

Another vital activity appearing in customer service area, both internal and external, is the acquisition and distribution of knowledge, which should be strongly tailored to the individual needs of the user. As the survey shows it is always easier to implement such behavior in the case of a client who is aware of knowledge management and accumulation.

The organization aware of the needs of the knowledge management is more eager to collect the knowledge needed by the implemented solution. However, there are many organizations that lack the awareness of having such knowledge structured and stored. It is then necessary to establish close cooperation with persons designated and responsible for ordering and systematizing knowledge and implementation of specific procedures. This results in increased awareness of

the need for the process of collecting, organizing the knowledge and preparing appropriate procedures for structuring and organizing the knowledge within the organization.

E. Agent technologies in enrichment of organizational knowledge

Basically, the enrichment of organizational knowledge with the use of software agents is possible mainly by their usage in the acquisition of knowledge which is necessary to effectively carry out business processes of the organization. The agent system has to be taught. In the context of agent systems it means that agent have to be able to learn, i.e. gain knowledge from different organizational sources. There is also a body of knowledge outside of an organization which can be used by agents indirectly, for example through a solution of "browser of quotations" or through the integration with different types of external knowledge bases.

Agent-based solutions may be able to query different databases. Usually it is not a problem to teach agent the knowledge in a particular field. It can be done through connecting him to the existing knowledge base but to do so we need a help of a knowledge engineer.

The required knowledge can be sought out through a semantic search engine that collects information from various websites. Another way to supply resources of organizational knowledge may be communication with the user in natural language. Respondents also indicated the possibility of using surveys, interviews, questionnaires to gather knowledge in the specific area of interest.

Also, the knowledge needed for the operation of software agents enhances organizational knowledge resources. From the point of view of knowledge engineer the resources of knowledge used by the agent should be structured in the machine-readable form, so by definition such vocabulary can be entered into the agent's knowledge base. Such knowledge is frequently extending scenarios of usage that contain threads related to the specific knowledge pools. Threads are grouped into scenarios. We can say that knowledge that is used by the agents is two-folded: a part of it takes a form of vocabulary, and the other part is a knowledge about the scenarios of usage.

IV. CONCLUSION

Conducted partial studies indicate that such solutions are now heavily used by organizations and require further consideration in terms of the methodology of agent construction solutions. The studies have shown that companies implementing such solutions do not sufficiently take a methodical approach to the problem of the design and deployment of agents. This is due to the fact that the construction of such solutions requires, on the one hand, to address the issues of software development methods, and on the other theories of knowledge engineering that are required in the context of modeling the knowledge base of agents. This necessitates the search for the new methodologies for the design and construction of such solutions. In particular, in the context of semantic methods of agents' knowledge representation, which should be linked to the information systems of the organization. These issues were highlighted by the participants of this study.

Despite the differences in the details of the methodologies used in the design and the tools used, companies generally agree on the steps of the implementation of agent-based solutions. In simple terms it can be assumed that implementation is based on four main processes. The first one is to analyze and collect information from the user, based on which the knowledge of an agent will be formulated. The information comes from many sources, such as individual interviews, the results of searching through the paper-based and electronic documents. Because such knowledge usually is not codified in clear and understandable manner, it is necessary to systematize the acquired information. Thus the next step is to design a model of the knowledge which will be used to its structuring. Once the information is structured it can be used by an agent to identify the thread and give you the right answers to users' queries. The next step is the implementation of the agent system. The last stage is usually testing the system by the user and getting feedback on its operation. Feedback information allows designers to assess whether the knowledge that has been introduced to the agent is correct, if something has been missed and can be supplemented, if the scope or substance of knowledge has been changed in some way. Then it is necessary to update the knowledge. Despite the use of their own methodologies agent systems' vendors base on proven, UML-based tools. Such tools facilitate modeling of knowledge structures for knowledge bases, allow the ordering of knowledge structures and to describe some of relations in the knowledge base. They are also used when it comes to designing the architecture of agent systems. We can name several of them, both universal, such as Enterprise Architect, Power Designer and Eclipse, the Semantic Works or more specific like Protege for building ontologies. Companies recognize the benefits of using tools like CASE even when they can confirm that their usage has to be preceded by the learning process which poses some difficulties for its users. Consistency in the use of such solutions provides good documentation of its architecture and guarantees the appropriate level of maintainability of the agent system.

The process of building agents using the semantic mechanisms of knowledge representation requires that, at the design stage of the system, it is necessary to determine how the knowledge of agents will be sourced from within the organization and how it will be updated and managed. This requires that the organization, in which this solution is built, has been focused on knowledge management processes, which can support the use of such solutions. Respondents also pointed out that finished implementations allowed to disclose previously unknown places in the organization, in which the domain knowledge is stored. As a result, the process of implementing such a solution should not end with its completion, but requires further improvement of knowl-

edge bases of agents. Although it requires posting staff members to supervise the knowledge base of agents, but it is feasible from the point of view of the organization efficiency because agents can handle multiple clients simultaneously using the resource of codified knowledge. This aspect becomes a vital part of a research in the context of agent solutions, since contemporary methods of their construction are mainly focused on designing their architecture and, in a small percentage, indicate the possibility of modeling the knowledge of the system. The issue of methodological approaches to the modeling of knowledge structures for software agent societies, using the semantic mechanisms of knowledge representation, will be a main focus of authors' further research.

ACKNOWLEDGMENT

The issues presented constitute a preliminary stage of the authors' research into the aspect of modeling software agent societies in knowledge-based organizations. The project was financed from the funds of National Science Centre 2011/03/D/HS4/00782. At the same time the authors want to thank the following companies (alphabetical order): 2CONN Tech Sp. z o.o, Instytut Technik Innowacyjnych EMAG, PIRIOS S.A., Stanusch Technologies S.A., Sztuczna Inteligencja Sp. z o.o. for their cooperation and willing participation in the study.

REFERENCES

[1] Ivanovic M., Budimac Z. (2012) Software Agents: State-of-the-Art and Possible Applications *International Conference on Computer Systems and Technologies* - CompSysTech'12, pp.11-22.

[2] Kuligowska, K. (2007). Zastosowanie inteligentnych agentów w aplikacjach handlu elektronicznego: wirtualni asystenci, *Ekonomia,* Warszawa, pp. 99-112.

[3] Liebovitz, J. (Ed.). (2012). *Knowledge Management Handbook: Collaboration and Social Networking.* Boca Raton: CRC Press.

[4] Van Elst, L., Dignum, V., Abecker, A. (2004). Towards Agent-Mediated Knowledge Management. In *A. Van Elst, L. Dignum, V. Abecker (Ed.), Agent-Mediated Knowledge Management.* Springer Verlag.

[5] Paprzycki, M. (2009). Agenci programowi jako metodologia tworzenia oprogramowania, url: http://www.e-informatyka.pl/wiki/*Agenci_programowi_jako_metodologia_tworzenia_oprogramowania,* 2013-05-20 Computer Science Department, Oklahoma State University, Tulsa, OK 74106 USA.

[6] Sołtysik-Piorunkiewicz A., Żytniewski M. (2013) "Software Agent Societies for Process Management in Knowledge-Based Organization" *European Conference on Knowledge Management - ECKM 2013* (positively reviewed article)

[7] Too Chuan Tan J., Inamura T. (2012) Extending Chatterbot System into Multimodal Interaction Framework with Embodied Contextual Understanding, HRI '12: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, 2012, pp. 251 – 252.

[8] Zimmerman J., Forluzzi J., Mancuso V. Kwak S. (2007) How interface agents affect interaction between humans and computers, *DPPI '07 Proceedings of the 2007 conference on Designing pleasurable products and interfaces* pp. 209-221.

[9] Żytniewski, M. (2010) Ontologies in multiagent systems. in: *Decision Support Systems Conference SWO'2010.* edited by T. Porębska-Miąc and H. Sroka, University of Economics in Katowice, Katowice.

[10] Żytniewski, M. (2013) Development of the conception of software agent societies, *Conference materials "European space of Electronic Communication",* Copenhagen 2013.

2nd Workshop on Information Technologies for Logistics

The main purpose of the workshop is to provide a forum for researchers and practitioners to present and discuss current issues concerning use of ICT in logistic applications (hardware and software). There will be also an opportunity for hardware integrators, software developers and logistics companies to demonstrate their solutions, as well as achievements, in different logistic systems.

Topics

The topics of interest include but are not limited to:

- Innovations in information systems supporting logistics and its management (WMS, SCM, TMS, LIS, VMI, CRP, PLM, and others)
- Innovative technologies in warehouse management: RFID, Voice Picking, Image Recognition, Pick Radar, etc.
- Logistics process modeling, including influence of warehouse automatic
- Optimization of logistics processes:
  - optimal vehicle routing and management, boundary conditions
  - optimal picking routing (global optimization, fast search, collision prediction and prevention)
  - shared mobility systems
  - day-to-day dynamic traffic assignment models
  - effective methods of picking (multi picking, batch picking ect.)
  - relationships between picking efficiency and products decomposition in warehouse area
- Environmental protection (for example carbon-aware transportation)
- Artificial intelligence systems and decision support systems in logistics
- BI, data mining and process mining in logistics
- Quality management algorithms and methods
- Material Flow Theory and applications

Event Chairs

Gontar, Beata, Uniwersity of Łódz, Poland
Wierzbicki, Marek, BinCode, Poland

Program Committee

Banaszek, Zbigniew, Warsaw University of Technology, Poland
Bobkowska, Anna, Gdansk University of Technology, Poland
Bruzda, Jaonna, Nicolaus Copernicus University, Poland
Celebi, Dilay, Istanbul Technical University, Turkey
de Koster, Rene, Erasmus University, Netherlands
Duran-Grados, Vanesa, University Cadiz, Spain
Gontar, Zbigniew, University of Lodz, Poland
Han, Chadong, Towson University, United States
Jin, Tongdan, Texas State University, United States
Lent, Bogdan, University of Bern, Switzerland
Liao, Da-Yin, National Chi Nan University, Taiwan
Løkketangen, Arne, Molde University College
Minner, Stefan, Technische Universität München, Germany
Papageorgiou, Markos, Technical University of Crete, Greece
Roe, Michael, Plymouth University, United Kingdom
Sitek, Pawel, Kielce University of Technology, Poland
Spassov, Vikenti, University of Transport "T. Kableshkov", Bulgaria
Tek, Omer, Yasar University, Turkey
Tipi, Nicoleta, University of Huddersfield, United Kingdom
Zielinski, Jerzy, University of Lodz, Poland

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1191–1194

Product Swapping and Transfer Sales between Suppliers in a Balanced Network

Ikbal E. Dizbay, Omer Ozturkoglu
Yasar University, Universite cd. Selcuk Yasar Kampusu, Agacliyol
Izmir, Turkey
Email: {ece.dizbay, omer.ozturkoglu}@yasar.edu.tr

***Abstract*—In this paper we present a preliminary, deterministic mathematical model of cooperative supply chain network of suppliers and customers. We consider horizontal cooperation among suppliers such that they can swap their orders to reduce their transportation cost, and they can purchase products from each other to reduce their shortage cost. Hence, the objective is to examine the potential swap and horizontal purchasing operations between suppliers under perfect information sharing. Assuming a balanced network in a single-period, in which total capacity of suppliers is greater than or equal to the total demand of customers, we conduct an empirical analysis for six suppliers and eight customers. The analysis suggests for many suppliers the benefits of order swapping and lateral purchasing.**

I. Introduction

IN today's competitive environment, customer satisfaction is one of the most prominent performance measures for companies, especially for the ones that serve consumers. In order to increase customer satisfaction, companies might focus on increasing customer service level, responding orders quickly, shipping the right items in the right amount. To be able to achieve these, there are several classical strategies implemented by companies such as opening new depots or warehouses close to customers, increasing inventory levels at the stores including safety stocks, using fast transportation modes or less-than-truck load shipments, etc. However, these methods cause increase in logistics and supply chain cost, hence reduces competitiveness of the companies. Therefore, reference [1] discussed that implementing co-opetition strategies, which identifies the existence of competition and cooperation strategies among different companies, provide companies to maximize their individual profits. Hence, companies look for win-win scenarios by sharing information that has an important effect on competition and the success of cooperation strategies [2]. So, what might be an example of cooperation strategies for different companies or different branches of a company? Examples to these strategies in literature might be inventory sharing, inventory pooling, lateral transshipment, and order swapping and exchanging by sharing partial or full information. In this paper, we focus on order swapping and lateral transshipment.

Swapping can be defined as an agreement between businesses, which are competing or non-competing, exchanging shipping, production, assets or market position to reduce overall costs. Our main purpose and motivation of using order swapping is to provide reduction in transportation costs by shipping products to closer customers on behalf of each other. Details and assumptions of this operation is discussed in the next section.

There are very limited research about swap operations in the context of supply chain and logistics. Reference [3] is the most relevant study to our study. Reference [3] developed a multi period mathematical model that seeks for an efficient coordination of swap and exchange transactions between supply chain partners in the field of oil and petroleum industry. They assume perfect information sharing, known demand and sufficient production to meet customers' demand. Reference [4] discussed the possible benefits and risks of swapping commodities and capacity with competitors by giving real-life examples. For example, two different manufacturers in chemical industry, one located in USA the other is in Europe, agreed to swap their monomers to use in their polymer operations after verifying that the product is the same. Hence, both company saved tens of million dollars in logistics cost per year. They also mentioned that industries that produce textile, paper, iron and steel products might include potential savings by implementing swap strategies. Lateral transshipment strategy allows suppliers or retailers in the same echelon to pool their inventories in order to enhance lower inventory levels and costs while providing at least the required customer service level. [5] define two types of lateral transshipment according to the timing of transshipments: proactive and reactive. While proactive transshipment can be planned in advance, reactive transshipment is performed when needed, for example when a company stocks out or faces a risk of stock out. Almost all of the studies about lateral transshipment reviewed by reference [5] focuses on inventory problems in stock points or branches of the same company. However, we consider different, competing companies in our study. Because of this reason, we suggest readers to read detailed review on lateral transshipment in [5].

In the light of these references, we want to notice that the contribution this study is to integrate two effective cooperation strategies in a mathematical model of single echelon supply chain network. Hence, in Section II we present assumptions of our model and the model formulation. Then, we conclude our study with an empirical analysis of the developed model and discussion of the results and future research questions.

978-83-60810-53-8/$25.00

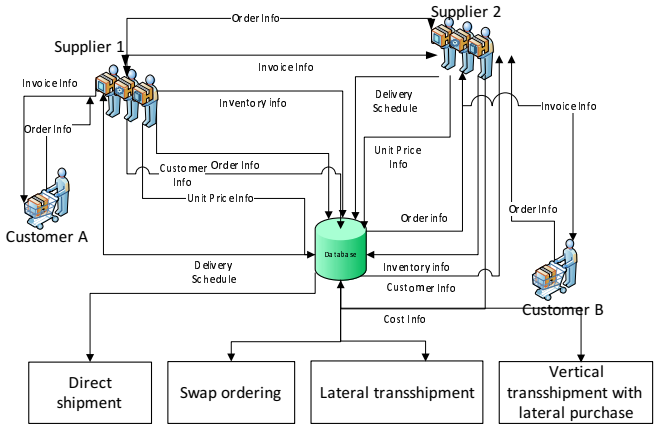


Fig. 1. Information system to generate shipping orders for suppliers in cooperation.

II. Model Assumptions and Formulation

In this study a single, commodity product network is examined which consists of a group of competing, but co-operating suppliers, where each supplier produces at their respective capacity which is known and constant. Customer demands are also known and constant. After customer sends their order requests to a supplier, the supplier share all the relevant data with other cooperating suppliers such as unit cost of purchase, unit market price, unit cost of transportation, inventory levels, location of customers and their orders. Hence, we assume pure information sharing among competing but cooperating suppliers. After all the data are processed and the mathematical model is utilized to generate the shipping orders to meet customer demands, there might appear four cases: direct shipment, swap ordering, lateral transshipment, vertical transshipment with lateral purchase (see Figure 1).

The demand at each supplier may be satisfied by shipping available on-hand inventory directly to a specific customer if there is no savings in swap ordering with another cooperating supplier. If swap decision is generated by the information systems, then the system sends swap orders to the suppliers that are going to ship the determined amount of products on behalf of each other to other's customer. Hence, the model aims to provide savings in transportation cost for both suppliers in cooperation by shipping products from closer suppliers to the customers. Here, we assume a balanced swap between two suppliers to construct equity between them such that each supplier should ship the same amount of product on behalf of each other. The appropriate documents and information flows among suppliers and customers flow as seen in Figure 1.

The suppliers share inventory and pricing information and might participate in an lateral transshipment arrangement in which every supplier must receive some benefit from the lateral transshipments. Because these cooperating suppliers are also competing each other, a lateral transshipment is realized only when a supplier that stocks out (called "dependant" hereafter) decides to purchase products from another supplier that has excess stock on hand (called "seller" hereafter). Due to lateral transshipment agreement, sellers should sell and ship the required amount to the dependants unless they face a stock out. In this preliminary model, we assume that the purchasing (or selling) price of the product between a seller and a dependant is deterministic and determined by averaging the unit market price of the dependant and the unit cost of the seller. In order to provide benefit to these suppliers, we assume that the unit market price of any supplier is greater than the unit cost of any supplier. This enhances that a dependant always buys products from other suppliers with a lesser price than its market price. While this pricing mechanism is relatively simplistic, it provides a standard policy to calculate the transfer price between suppliers. Even though this strategy aims to provide benefit to cooperating suppliers, it is still possible for shortage to occur if there is no benefit of purchasing excess inventory when cost of lost sales is less than total cost of purchasing and transportation between dependant and seller. The cost of lost sales is assumed to be the sales price of a product for that supplier. Additionally, there are no holding costs associated with excess inventory because this analysis only covers one period.

In vertical transshipment with lateral purchase agreement, if a dependant decides to purchase its need from a seller, he may require the seller to ship the product to its own customer if the customer is closer the seller than the dependant. Then, the dependant pays the cost of purchasing and the transportation cost to the seller. Hence, it can provide savings in transportation cost. In summary, the mathematical model we develop considers the sequence of cases we discussed above. First, suppliers decide if there is any benefit to executing a swap for their orders. If there is a benefit for two suppliers, a balanced swap occurs in which supplier i ships products to the supplier k's customer, and vice versa. If there is no benefit from executing a swap, the supplier ships directly to their customer. If a supplier's demand exceeds their inventory level, then that supplier seeks to purchase product from suppliers that have excess inventory. Then, he may either receive transshipments from other suppliers or the product may be shipped directly from seller to the customer on behalf of the dependant. Hence, the model parameters and variables are discussed as the followings.

The relevant parameters for the model are as follows:
c_{ij} :contracted unit transportation cost between supplier i and customer j.
D_{ij} : quantity demanded from supplier i by customer j.
I_i : inventory level on-hand at supplier i.
p_i : unit market price of supplier i.
u_i : unit cost of supplier i.
The model variables are as follows: q_{ij} : quantity shipped directly from supplier i to customer j.
b_{ij} : amount of lost sales between supplier i and customer j due to shortage.
y_{ijk} : quantity shipped from supplier i to customer j on behalf of supplier k due to order swap.
x_{ki} : quantity shipped from supplier k to supplier i due to

purchase by supplier i under lateral transshipment agreement. w_{ijk} : quantity shipped from supplier i to customer j on behalf of supplier k due to purchase by supplier k under vertical transshipment with lateral purchase.
The upper bound of the cost of the i^{th} supplier (Z_i^u), which provides the worst case, is the total transshipment cost of the i^{th} supplier that ships to only its respective customers and its total cost of lost sales, if exist. The sum of Z_i^u for all suppliers is the upper bound of the cost of supply chain network (Z^u).

$$Z_i^u = \sum_{j=1}^{n} q_{ij}c_{ij} + b_{ij}p_i \tag{1}$$

$$\text{subject to} \tag{2}$$

$$\sum_{j=1}^{n} q_{ij} \leq I_i \tag{3}$$

$$q_{ij} + b_{ij} = D_{ij}, \forall i, j \tag{4}$$

$$q_{ij}, b_{ij} \geq 0. \tag{5}$$

Hence, the objective function of the model (Z) considers order swap and the lateral purchase among cooperating but competing suppliers if an excess demand exists.

$$\begin{aligned} minZ = &\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}q_{ij} + p_i b_{ij}) \\ &+ \sum_{k=1,i\neq k}^{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}y_{ijk} + c_{kj}w_{kji}) \\ &+ \sum_{k=1,i\neq k}^{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{u_k + p_i}{2}\right)w_{kji} \\ &+ \sum_{k=1,i\neq k}^{m}\sum_{i=1}^{m}\left(t_{ki}x_{ki} + \left(\frac{u_k + p_i}{2}\right)x_{ki}\right) \end{aligned} \tag{6}$$

The first constraint is related to the inventory level for every supplier that is greater than or equal to the quantity shipped directly to their customers, the quantity shipped to other supplier's customers due to swaps, and the changes in capacity from the buying or selling of product from other suppliers.

$$\begin{aligned} &\sum_{j=1}^{n} q_{ij} + \sum_{j=1}^{n}\sum_{k=1,k\neq i}^{m}(y_{ijk} + w_{ijk}) \\ &\qquad + \sum_{k=1}^{m}(x_{ik} - x_{ki}) \leq I_i, \forall i. \end{aligned} \tag{7}$$

Demand for every supplier and customer relationship must be satisfied by shipment from supplier i, shipment from a different supplier due to swaps, the shipment of material sold to other suppliers to their customers, or lost sales.

$$q_{ij} + b_{ij} + \sum_{k=1,k\neq i}^{m}(y_{kji} + w_{kji}) = D_{ij}, \forall i, j. \tag{8}$$

Every swap must be balanced between two suppliers.

$$\begin{aligned} \sum_{j=1}^{n} y_{ijk} - \sum_{j=1}^{n} y_{kji} = 0, \quad &\forall i = 1, 2, ..., (m-1); \\ &\forall k = (i+1), ..., m; i \neq k \end{aligned} \tag{9}$$

Every supplier must benefit under the network swapping arrangement considering the costs for shipment from supplier i, shipment from a different supplier due to swaps, the cost of purchasing product from other suppliers, and cost of lost sales are less than or equal to its upper bound.

$$\begin{aligned} &\sum_{j=1}^{n}(c_{ij}q_{ij} + p_i b_{ij}) + \sum_{k=1}^{m}\sum_{j=1}^{n}(c_{ij}y_{ijk} + c_{ij}w_{ijk}) \\ &+ \sum_{k=1,i\neq k}^{m}\sum_{j=1}^{n}\left(\frac{u_k + p_i}{2}\right)w_{kji} \\ &+ \sum_{k=1,i\neq k}^{m}\left(t_{ki}x_{ki} + \left(\frac{u_k + p_i}{2}\right)x_{ki}\right) \leq Z_i^u, \forall i. \end{aligned} \tag{10}$$

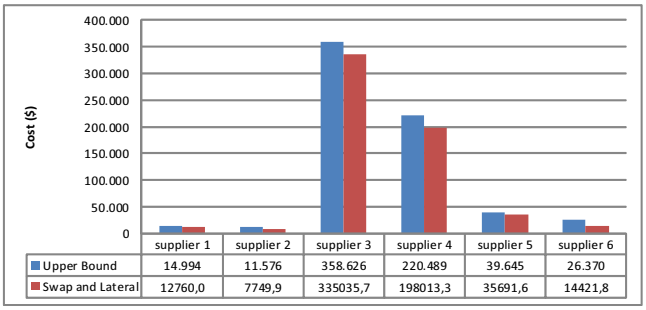
Finally, quantity shipped directly, lost sales, quantities shipped to other suppliers' customers, quantities purchased by other suppliers, and quantities transshipped between suppliers must be non-negative.

$$q_{ij}, b_{ij}, y_{ijk}, w_{ijk}, x_{ki} \geq 0. \tag{11}$$

III. Numerical Investigation and Conclusion

A demand of a customer is randomly generated number between 100 and 1000 using uniform distribution. A supplier's capacity is also assumed to distribute uniformly between 1000 and 4000 such that total stock in the network is greater than the total demand of the network. Hence, a supplier observes either a shortage or an excess inventory. Excess inventory may be sold to a supplier facing shortage at a price between the seller's unit cost and the dependant's market price. Unit cost of products was generated between 300 and 330, and then market price of a supplier is generated by multiplying its unit cost by a uniformly generated profit margin between 10% and 30%.

The upper bound and the proposed model developed in the previous section run for a network of six suppliers and eight customers. As seen in Figure 2, every supplier receives some benefit either from the swapping or lateral transshipment agreement, or from both. Hence, the model provides benefit in transportation cost by a swap agreement and reduction in cost of lost sales by allowing cooperation among suppliers for their excess demand and supply in a single period. In order to investigate the effect of the proposed model on the supply chain network cost of each individual supplier, we aim to work on multi period with holding cost and partial information sharing.



	supplier 1	supplier 2	supplier 3	supplier 4	supplier 5	supplier 6
Upper Bound	14.994	11.576	358.626	220.489	39.645	26.370
Swap and Lateral	12760,0	7749,9	335035,7	198013,3	35691,6	14421,8

Fig. 2. Total transportation and backorder costs of each supplier with and without swap and lateral transshipment.

REFERENCES

[1] A. M. Brandenburger and B. J. Nalebuff, "Co-Opetition: A Revolution Mindset that Combines Competition and Cooperation - The Game Theory Strategy That is Changing the Game of Business," *Currency*, 1997.

[2] B.R. Konsynski and F. W. McFarlan, "Information Partnerships - Shared Data, Shared Scale," *Harvard Business Review*, 1990, vol. 47:12, pp. 114-120.

[3] R.Al-Husain and T.Assavapokee and B. Khumawala, "Modelling The Supply Chain Swap Problem in The Petroleum Industry," *Int. J. of Applied Decision Sciences*, vol. 1, 2008, pp. 261-281.

[4] A. Kosansky and T. Schaeffer, "Should You Swap Commodities with Competitors?," *Supply Chain Quarterly*, vol. 2, 2010, pp. 42-47.

[5] C. Paterson and G. Kiesmüller and R. Teunter and K. Glazebrook. "Inventory models with lateral transshipments: A review," *European Journal of Operational Research*, vol. 210:2, 2011, pp. 125-136

[6] J. Narus and J. Anderson, "Rethinking Distribution-Adaptive Channels," *Harvard Business Review* 2006, pp. 112-120.

[7] N. Rudi and S. Kapur and D.F. Pyke,"A Two-Location Inventory Model With Transshipment and Local Decision Making," *Management Science,* vol. 47:12, 2001, pp. 1668-1680.

[8] K.S. Krishnan and V.K. Rao, " Technical Notes: Inventory Control in n Warehouses," *The Journal of Industrial Engineering,* vol. 1, 1965, pp. 212-215.

[9] C. Das, "Supply and Distribution Rules for Two-Location Inventory System: One Period Analysis," *Management Science,* vol. 21, 1975, pp. 765-776.

[10] G. D. Eppen, "Effects of Centralization on Expected Cost in a Multi-Location Newsboy Problem," *Management Science,* vol. 25, 1979, pp. 489-501.

[11] T.M. Simatupang and M. T., R. Sridharan, "The Collaborative Supply Chain," *The International Journal of Logistics Management,* vol. 13, 2002, pp. 15-30.

[12] L. Doug and N. Rudi, "Who benefits from transshipment? Exogenous vs. Endogenous Wholesale Prices," *Management Science,* vol. 50, 2004, pp. 936-953.

[13] H. Zhao and V. Deshpande and J.K. Ryan, "Inventory Sharing and Rationing in Decentralized Dealer Network," *Management Science,* vol. 51, 2005, pp. 531-547.

[14] X. Yan and H. Zhao, "Decentralized Inventory Sharing with Asymmetric Information," *Operations Research,* vol. 59, 2011, pp. 1528-1538.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1195–1202

Rule-based Approach For Supplier Evaluation

Andrzej Macioł, Stanisław Jędrusik, Bogdan Rębiasz
AGH University of Science and Technology,
Faculty of Management,
ul. Gramatyka 10,
30-067 Kraków, Poland
Email: {amaciol, jedrusik, brebiasz}@zarz.agh.edu.pl

***Abstract*—This paper presents a concept of use of the rule-based reasoning systems for evaluation and classification of the suppliers. The problem of suppliers selection is widely discussed in literature. Majority of the authors apply the method of multi-criteria evaluation for selection of suppliers, mainly the Analytic Hierarchy Process (AHP) algorithm and related ones, to find its solution. In this paper it has been proved that a suitably expressive system of rules management can be used as an effective tool for suppliers evaluation. In the presented work we have applied the Rebit system which was elaborated by the AGH University of Science and Technology. An example of evaluation of a supplier of primary charging materials for metal processing enterprise has been presented. It has been shown how individual evaluation criteria are grouped into sets of independent rules and how one may use tools to enhance the knowledge acquisition.**

I. Introduction

THE PROBLEM of supplier selection plays a prominent role in the modern economy. Supplier selection and evaluation is one of the most vital actions of enterprises in a supply chain. Undertaking faulty decisions in this area may be a cause of critical disturbances in execution of the fundamental tasks of the manufacturing enterprises.

Over the past several years, with the recent trend on just-in-time (JIT) manufacturing systems, there is an emphasis on strategic sourcing that establishes long-term mutually beneficial relationship with fewer but better suppliers [1].

Strategic decisions concerning supplies with raw materials are tied up with evaluation and selection of potential strategic suppliers. Selection of suppliers presents itself a complex decision-making problem which is featured with multi-criteria, of different nature of the criteria (quantitative, qualitative) and with multi-stages of the decision. In today's global and open innovation economy strategic supplier selection and evaluation decisions must not be solely based on traditional selection criteria, such as cost, quality and delivery. In strategic sourcing, many other criteria should be considered with the aim of developing a long-term supplier relationship such as quality management practices, long-term management practices, financial strength, technology and innovativeness level, suppliers' cooperative attitude, supplier's co-design capabilities, and cost reduction capabilities [1].

Both the procedure and algorithms for supplier selection cannot rely exclusively on the "historic" experience of a manager. It must reflect some strategy of the enterprise as defined by the management of the enterprise in scope of execution of its fundamental activity. This strategy determines then the purchase strategy including such factors as acceptable standard of prices, required quality, desirable conditions of the long-term cooperation etc. Thereby the selection of suppliers can be treated as an integral element of definition of the processes and business rules. Therefore, the rule-based approach which is characteristic for Business Rules Management Systems can be useful while solving the problem of selection and ranking of the suppliers. Nevertheless, there is a pre-condition that one must have at one's disposal a suitably expressive and reliable tool. And there is such a tool: The Business and Technological Rules Management - The Rebit System elaborated by the AGH University of Science and Technology.

The aim of the paper is to present the possibility of use of tools purposed for business management rules to solve the problem of supplier selection. We have presented the results of our research directed on proposing a solution which would enable one to get flexible formation of purchase strategy and current adaptation of the selection criteria for the choice of suppliers to changing market conditions.

This paper concerns problems of selection and grouping of suppliers of charging materials in respect to the quality of the services provided by them, as well as the quality and parameters of the delivered materials. Furthermore, we have analyzed the influence of factors resulting from the purchase strategy and such external factors as the destination of the acquired materials on the issue of supplier selection.

In the first part of the paper there will be formulated a problem of evaluation and classification of the suppliers and we will present a review of the relevant literature. Then we will present the Business and Technological Rules Management - the Rebit System and its specific features used to solve the problem of suppliers evaluation. An example of evaluation of a supplier of charging materials accomplished with use of the Rebit system will be also presented. The paper will be finished with a critical comparison of the proposed solution with some concepts known from the literature; some suggestions concerning further works will be also presented.

II. Supplier evaluation and selection

The problem of supplier selection can be considered from different points of view. Firstly, this is a task relying on a

This work is supported by the Innovative Economy Operational Programme EU-founded project (UDA-POIG.01.03.02-12-008/11-00)

978-83-60810-53-8/$25.00

single selection, from the list of the suppliers, such supplier who in the best way fulfills the requirements resulting from execution of the specific production order. In such case we have to do with a simple issue of multi-criteria evaluation. The set of the criteria, their inter-relation, as well as the method of their evaluation may be different in each individual case.

In business management practice much more important meaning has a task of creation and periodic updating of the list of suppliers for different groups of charging materials permanently tied up with the recipient. Depending on the nature of the realized manufacturing processes we can have to do with complex deliveries covering different kind of materials (and in some cases also services) or with special-purpose deliveries including deliveries of strictly defined sort of products. The way of solving the problem of supplier selection depends also on kind of the material needs. One shall consider the selection issue in case of materials directly consumed in the production process and otherwise in case of accessory materials. The very procedure of evaluation of the potential suppliers will be in case of any of the above mentioned situations similar and will resolve itself into establishing a ranking list of the evaluated subjects or classifying them into the previously established categories (e.g. permanent main supplier, permanent auxiliary supplier, occasional supplier). However, the evaluation criteria will differ. Due to the number of criteria and their interrelation, the procedure can be executed in one stage or in multi-stages. A special group of decision issues are decisions of strategic nature that bind the recipient with the supplier for a long time.

Supplier evaluation and problem selection have been presented in the literature as the goal of the development research or as the task related with creation of the software applications. In most cases we have to do with proposals of creation of the Decision Support Systems (DSS) which could be applied in business practice. Thus, an issue of essential significance is not only the way of solving the problem of multi-criteria evaluation but also assumptions and methods of execution of an application meeting user's requirements. These both decisions must be considered together, because the characteristics of the evaluation algorithms exerts an essential influence on the way of realization of the application and particularly on creation of conditions for current updating of the assumptions without the necessity of modification of the very structure of the programming modules. Therefore, the way of proper knowledge modeling and management is very important.

The object of our research is the problem of classification of potential suppliers of the primary charging materials. Taking the above into consideration and assuming that the adopted methodological solutions should permit one to achieve easy creation of flexible application serving, inter alia, in practical realization of the current purchase strategy, we have made the following assumptions specifying the research problem:

- any potential supplier will be evaluated individually and not with use of the method of pairwise comparisons, as it happens in the Analytic Hierarchy Process (AHP) and similar methods;
- criteria related with complexity of the deliveries are not taken into consideration;
- while evaluating the deliveries one takes into account the destination of the materials earmarked to produce different groups of product;
- the way of presentation of the criteria in the model and criteria evaluation must be readable for the managers and allow them to present the purchase strategy;
- the application based on the proposed solution must allow one to perform simple updating of the evaluation criteria with no interference in the code.

The fulfillment of every of the above presumptions places specific requirements to the method used in knowledge representation and consequently - also to the inference mechanisms. We have acknowledged that the most profitable solution would be use of the rule based knowledge representation. A symbolic (linguistic) mapping of the assessment criteria and declarative nature of the knowledge optimally corresponds with the requirements and needs of the managers responsible for execution of the business operations.

III. Literature review

In the face of acute global competition, supplier management is rapidly emerging as a crucial issue to any enterprises striving for business success and sustainable development. As it was mentioned above, supplier evaluation and order allocation are complex, multi-criteria decisions.

Incorporating multi-dimensional information into vendor evaluation is important and well established in both academic and practitioner's literature [3], [4]. Over the years, several multi-criteria techniques have been proposed for the effective evaluation and selection of vendors. According to the literature, some supplier selection criteria are found to vary in different situations, and experts agree that there is no one best way to evaluate, select suppliers and that organizations use a variety of different approaches in their evaluating processes.

As it has been previously mentioned, the evaluation of a supplier is realized in different phases of the process of supply management and may concern different special cases. Depending on the purpose of the evaluation and the adopted assumptions, different criteria are taken into consideration.

Ha and Krishnan [3] summarizes some of these criteria which have appeared in literature since 1966. Among them one can mention (ordered according to the frequency of quoting in the literature): price, quality, delivery warranties and claims, after sales service, technical support, training aids, attitude performance history, financial position, geographical location, management and organization, labor relations, communication system, response to customer request, e-commerce capability, JiT capability, technical capability, production facilities and capacity, packaging ability, operational controls, ease-of-use maintainability, amount of past business, reputation and position in industry, reciprocal arrangements, impression, environmentally friendly products, product appearance, catalog technology.

In an overall analysis of 181 articles referenced within the studies made by Erdem and Göçen [2], AHP related

methodologies seem to be the most popular techniques which are applied in over 36% of the studies. This is mostly due to the fact that AHP incorporates both qualitative and quantitative evaluation of the decision maker by use of tangible and intangible factors designed in a hierarchical manner. It is suitable, flexible and easy-to-use for multi-criteria decision making and can be applied in group decision making environments as well.

Along with usage of the AHP method one can find in the literature other solutions [1], [5], [6], [7]: multicriteria classification and sorting methods (among other sorting method based on the PROMETHEE methodology), Game Theory, Decision Trees, Factor Analysis, Structural Equations, Loss Functions, Process Capability Index, Expert Systems, Case Based Reasoning (CBR), data envelopment analysis (DEA), and neural network (NN).

Although several techniques and models have been utilized for the selecting and evaluating of vendors, efficient partner selection, combining multiple techniques (AHP, DEA, and NN), has not been suggested previously with regard to the purchasing evaluation process [3]. The hybrid method uses an AHP to assign weight to the qualitative selection criteria, and it uses a DEA, NN or other methods in order to choose efficient vendors in the final selection process. Exemplary, the study [1] aims to develop models and generate a decision support system (DSS) for the improvement of supplier evaluation and order allocation decisions in a supply chain. Initially, an analytic hierarchy process (AHP) model is developed for qualitative and quantitative evaluation of suppliers. Based on these evaluations, a goal programming (GP) model is developed for order allocation among suppliers.

As it has already been stated decisions concerning organization of the supply of strategic nature are of specific character. An example of such situation one can find in work [8] where the problem of warehouse selection for a company was presented. This is a valuable and realistic decision problem in logistic and supply chain management (LSCM). The authors provided a solution for solving the raised problem via knowledge discovery and utilization. The decision knowledge in the form of ”if. . . then . . .” rules are generated based on known information of owned warehouses and then utilized for predicting the preference order of alternatives according to their profitability. The process of solving of the problem is realized in four stages. In the two first stages the expert knowledge and the knowledge derived from previous experiences is gathered. The third stage relies on elaboration of rules purposed for evaluation of the decision variants, whereas the fourth stage relies on their implementation in the specific case. Because both certain and uncertain information are taken into account, the authors introduce interval-valued intuitionistic fuzzy set (IVIFS), which consists of a membership function and a non-membership function, whose values are intervals rather than exact numbers. The presented procedures are sophisticated, time- and cost-consuming (due to engagement of external experts) and serve to solving individual problems.

One can also find some works which discuss the use of the rule-based approach in less complex problems related with undertaking multi-criteria decisions. Vokurka, Choobineh, and Vadi [9] develop a prototype expert system to evaluate the potential suppliers. Interesting approach to the preference modeling in the form of a set of "if..., then..." decision rules discovered from the data by inductive learning is presented in [10]. To structure the data prior to induction of rules, the authors use the Dominance-based Rough Set Approach (DRSA).

Summarizing the review of the literature one may find that the dominant method applied in the multi-criteria evaluation of the suppliers is the AHP method. Its imperfection one tries to level with use of supplementing method that permit objectification of the inherently subjective evaluations of experts. Nevertheless, there is lack of reports on the problem of the suppliers selection which would permit to treat them as activities aiming at business processes standardization. There are also no reports regarding usage of the concept of management with business rules in the matter under discussion.

IV. Business and Technological Rules Management System

Rebit System belongs to the category of Business Rule Management System (BRMS). It consists of rule and workflow engines, knowledge base editor, generic client, testing, validation and simulation tools, knowledge base repositories and resource management module. All these components may be configured and integrated into a standalone application. However, the main advantage is that they are also a set of loosely coupled components working in Service Oriented Architecture (SOA).

Rebit System supports all stages of knowledge base development process. Knowledge base editor is generally the first tool used in this process. It allows knowledge base creating and editing in a graphical or textual way with the help of intelligent prompts. All classic elements used in knowledge representation are available in Rebit editor. The main building blocks are rules, variables and functions. Rebit rules belong to the category of productions rules. Rule premises are logical conditions based on variables and functions. Rule conclusions are simple assignments. Rules are organized in so called rule sets, i.e. a group of logically connected rules. Rebit language provides more sophisticated elements, such as grids and decision tables. They allow for more concise and user friendly knowledge representation. The knowledge contained in decision tables and grids may be converted into ordinary rules. Rebit System is equipped with algorithms of learning by examples which allow for rational translation of decision tables into an effective rule set. The translation process usually takes place just before deployment.

The next steps in the process of knowledge base development are validation and testing. Rebit System provides an efficient validation and testing tools. Testing and validation algorithms allows for finding most inconsistencies and incompleteness. In order to perform validation and testing the knowledge base must be translated to Prolog language. The translation process is automatic and transparent to the

user. The user sees only the results and some additional statistics. The standard verification procedure include testing for knowledge base integrity, consistency and completeness. This procedure may be optionally extended by looking for hidden cycles and other unsafe phenomena.

Rebit inference engine handles three modes of inference: forward, backward and mixed, i.e., inference with predefined target variable. In the first inference mode the engine tries to infer all possible facts (variable values) from input data. The inference session may be continued after entering new input data. Backward reasoning is the verification of the hypothesis (concerning the value of some variable) on the base of information entered by the user at the request of the engine. The third inference mode, co-called mixed, combines forward and backward reasoning. The user specifies the final variable, i.e. the variable which must have the value. The inference engine finds the most efficient path and the set of variables that must have values. After that the mixed process of forward and backward chaining is accomplished. The process is stopped when the goal is reached or there is an evidence that the goal cannot be reached. The unique feature of Rebit engine is the possibility of controlling the reasoning process. It allow to reduce the total number of questions asked to the user.

Generic client is integrated with simulation module which enables automatic or semi-automatic simulation. The simulation allows finding groups of "not optimal rules", unused variables or rules, repeating or overlapping rules and other harmful elements. The first step in simulation procedure is the setup of statistical properties of all input variables. In the next step constraints and typical simulation properties like exit conditions are defined. The result of simulation procedure is a detailed report containing many useful statistics relating to variables and rules.

The knowledge base development process has not been as extensively explored as the software development process. However, general guidelines on how to proceed are the same in both processes. The iterative and incremental approach known from software development seems to be the most appropriate also in knowledge base development. The general idea is as follow: each iteration consists of identification, conceptualization and formalization of a selected portion of the domain. Next iteration starts after successful testing. It usually extends the previous portion of the domain. The iteration process ends when the entire domain is covered in knowledge base.

A special role in Rebit model of knowledge representation play elements called resources. They are introduced to enable access to data stored in SQL databases and other data sources. The current version of Rebit System includes connection strings, queries, variables and bindings. Query combined with variable - which represents the result of this query - form a new element called binding.

V. RULE-BASED APPROACH IMPLEMENTATION

A. Exemplary problem description

Our proposal has been verified on an example of selection and grouping of suppliers in enterprises producing metal products. All the enterprises for which the supplier selection problems play key role and are operating currently on the market have their own, substantially formalized procedures of suppliers selection. Standards in this domain are formally specified by ISO 9001:2000 norms (Clouse 7.4 Purchasing). Below there are shortly discussed: a process of evaluation and qualification of a supplier applied by a chosen producer of structural closed cold formed steel profiles.

A new supplier is evaluated from the point of view of the foreseen quality of cooperation with the enterprise and from the point of view of possibility of purchasing from him the commodity from given assortment group.

The supplier, in view of the quality of cooperation, is classified to the one of three groups: permanent main supplier, permanent auxiliary supplier, and occasional supplier. The classification based on the possibility of delivery of different assortments of the charging materials distinguishes three group of product providers: low cost, standard and HQ (high quality).

The assessment of the supplier takes into consideration the following aspects of cooperation:

- the experience in the cooperation,
- the contractual cooperation,
- the effectiveness of the complaints,
- servicing procedures,
- the reputation of the supplier,
- the balance of liabilities and receivables.

The supplier evaluation model in exemplary case is as shown in Fig. 1.

B. Knowledge base formulation

The considerable number of the criteria brings about need of grouping them in accordance with partition presented in Fig. 1. Simultaneously it is necessary to provide independent updating, testing and verification of groups of rules evaluating individual criteria. The Rebit system allows one to group rules into relatively independent rule sets.

Individual partial criteria can be both of quantitative or qualitative. Taking into consideration the necessity of adaptation of the knowledge model to symbolic reasoning which is specific for human being, it was assumed that the "rough" model would operate exclusively on qualitative variables. An additional justification of such solution is the fact that quantities which could be recorded as a constant or numerical variables can undergo dynamic changes. The updating of the knowledge model would require then frequent verification of many rules, what in turn is time-consuming and can be a source of errors. On the other hand, one shall assume that to some extend the values necessary to rules evaluation will be collected from external sources of data and perforce will be of quantitative type. This apparent contradiction can be solved with use of the mechanisms and tools of the Rebit system.

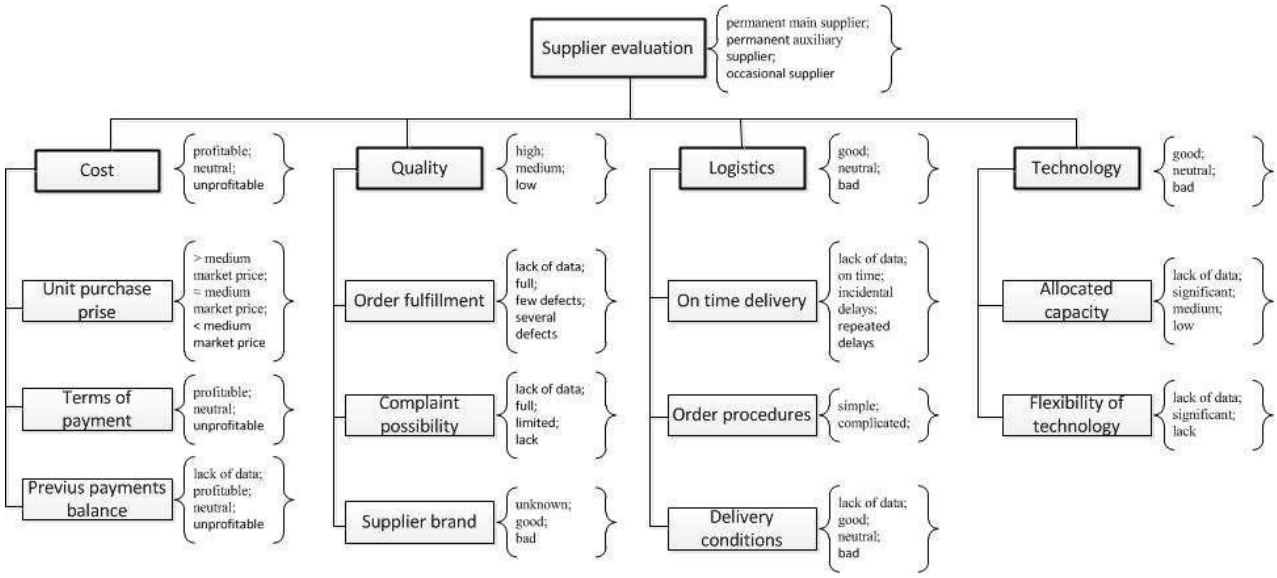


Fig. 1. The supplier evaluation

The adopted concepts can be illustrated on an example of knowledge acquisition for needs of evaluation of the criterion "Cost".

In case of the linguistic variables the most convenient form of knowledge representation is the decision table. In the Rebit system there is a possibility of generating such table after having previously declared appropriate enumerated types. Figure 2 presents a fragment of such table. The algorithm of learning by examples based on ID3, which is incorporated into the Rebit system, allows one to generate the "minimal" set of rules for the examples recorded in the decision table.

Fig. 2. Decision table for criterion Cost

Below one can find some exemplary rules:

```
RULE Cost_21
IF UnitePurchasePrice =
"greater then medium market price" AND
PreviousPaymentsBalance = "lack of data"
THEN Cost = "unprofitable"

RULE Cost_22
IF UnitePurchasePrice =
"greater then medium market price" AND
PreviousPaymentsBalance = "neutral"
THEN Cost = "unprofitable"

RULE Cost_23
IF UnitePurchasePrice =
"greater then medium market price" AND
PreviousPaymentsBalance = "unprofitable"
THEN Cost = "unprofitable"

RULE Cost_24
IF UnitePurchasePrice =
"greater then "medium market price" AND
PreviousPaymentsBalance = "profitable" AND
TermsOfPayment = "profitable"
THEN Cost = "neutral"
```

For the needs of symbolic representation of knowledge it was convenient to present the variable `UnitePurchasePrice` as a linguistic variable. In practice it is however compared with numerical quantities. This problem can be solved by introducing additional numerical variables `AskPrice`, `LowerPriceBound` and `UpperPriceBound`, as well as rules allowing for mapping a numerical to linguistic variable.

The pertinent rules have been presented below:

```
RULE CostPar_0
IF AskPrice <= LowerPriceBound
THEN UnitePurchasePrice =
"less then medium market price"

RULE CostPar_1
IF AskPrice >= LowerPriceBound AND
Ask_price <= UpperPriceBound
THEN UnitePurchasePrice =
"equal to medium market price"

RULE CostPar_2
IF AskPrice > UpperPriceBound
THEN UnitePurchasePrice =
```

```
"greater then medium market price"
```

As a result one obtains a model of knowledge which on the one hand is readable and easy to updating from the point of view of the manager but on the other hand, it permits one to data acquisition without the participation of the user (quantitative values can be downloaded directly from databases by means of the module of resources management). It is also worth pointing out that updating of the knowledge (e.g. in case of change of the purchase strategy) takes place on level of decision tables and does not require any interference into rules which are updated "automatically".

As it has been previously mentioned, one of the main criteria of evaluation of the supplier is the destination of the materials acquired from him. Depending on the requirements placed against the final goods which will be produced from the purchased materials, the way of the evaluation of the supplier's offer will also be different. The simplest solution would be to construct three separate models of the knowledge for: low cost, standard and high quality products. However, it is not justified whereas in case of a suitably expressive model of knowledge - which can be recorded in the Rebit system - not necessary. Some of the criteria are independent from the destination of the purchased materials (e.g. economic parameters). However, in case of some materials one can, thanks to parametrization, construct a model of knowledge that is common for different cases.

One can illustrate this on an example of quality evaluation of the delivered materials. Like in the former case the "rough" - linguistic knowledge base is created with the aid of a decision table transformed into the form of rules:

```
RULE Quality_12
IF OrderFulfillment = "full" AND
ComplaintPossibility = "full"
THEN Quality = "high"

RULE Quality_13
IF OrderFulfillment = "full" AND
ComplaintPossibility = "lack"
THEN Quality = "medium"

RULE Quality_14
IF OrderFulfillment = "full" AND
ComplaintPossibility = "limited" AND
SuplierBrand = "unknow"
THEN Quality = "medium"

RULE Quality_15
IF OrderFulfillment = "full" AND
ComplaintPossibility = "limited" AND
SuplierBrand = "good"
THEN Quality = "high"
```

In this case, parameterization of rules will consist of introduction of numerical variables and appropriate rules:

```
QualityPar_0
IF DefectsRatio <= FirstDefectsBound
THEN OrderFulfillment = "full"

RULE QualityPar_1
IF DefectsRatio >= FirstDefectsBound AND
Defects_ratio <= SecondDefectsBound
THEN OrderFulfillment = "few defects"

RULE QualityPar_2
IF DefectsRatio > SecondDefectsBound
THEN OrderFulfillment = "several defects"

RULE QualityPar_3
IF ProductRange = "low cost"
THEN FirstDefectsBound = 10

RULE QualityPar_4
IF ProductRange = "low cost"
THEN SecondDefectsBound = 15

RULE QualityPar_5
IF ProductRange = "standard"
THEN FirstDefectsBound = 8

RULE QualityPar_6
IF ProductRange = "standard"
THEN SecondDefectsBound = 10

RULE QualityPar_7
IF ProductRange = "high quality"
THEN FirstDefectsBound = 4

RULE QualityPar_8
IF ProductRange = "high quality"
THEN SecondDefectsBound = 7
```

Let us notice that quality of the materials from the supplier with a nine percent level of discard will be evaluated as meeting the requirements in case their destination is a low-cost product, as "few defects" in case of "standard" products and as "several defects" for "high quality" products.

In this case, the restrictions concerning the level of the expected discards have been entered as constants in the rules. This is justified by relative permanence of these values, as well as by the fact that each of them appeared in one rule only.

The problem of modification of the way of inference, depending on certain parameters, can be more complex. Let us assume that the criteria grouped to the class "Technology" are not evaluated in case of destination of the materials on "low cost" products. In order to not complicate the universal set of rules designed for final classification of the supplier one may assume that in case of "low cost" products the parameter `Technology` is set to "good" whereas parameters `AllocatedCapacity` and `FlexibilityOfTtechnology` are not verified. Some properties of the inference engine of the Rebit system allow

one to carry out this task in a very simple way. It is enough that to the set of rules defining the value of the variable Technology and generated on the basis of appropriate decision table:

```
RULE Technology_9
IF AllocatedCapacity = "lack of data" AND
FlexibilityOfTechnology = "lack of data"
THEN Technology = "bad"

RULE Technology_10
IF AllocatedCapacity = "lack of data" AND
FlexibilityOfTechnology = "significant"
THEN Technology = "neutral"

RULE Technology_11
IF AllocatedCapacity = "lack of data" AND
FlexibilityOfTechnology = "low"
THEN Technology = "bad"
```

one adds the following rule:

```
RULE TechnologyPar_01
IF Product_range = "low cost"
THEN Technology = "good"
```

The inference engine of the Rebit system, at each stage of evaluation of the rules, searches for the least expensive (i.e. requiring the least number of "questions" for variables) path of premises confirmation. In case of rules defining the value of the variable `Technology` "the cheapest" rule will be the rule `TechnologyPar_01`. This allows the engine will always have to do with "low cost" products, it will set up the value of the variable "Technology" on "good" and will not verify the consecutive rules defining this variable.

According to the principles identical with the above presented there are constructed all rule sets which describe four partial criteria. Then one constructs superior rule set which connects partial evaluations so as the final classification of the recipient is possible. In this case, one can also take advantage of the decision table (Fig. 3).

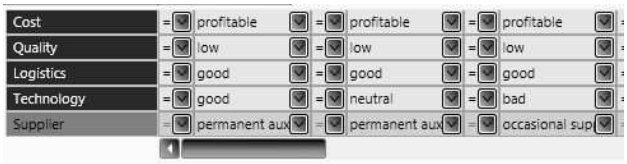


Fig. 3. Decision table for final evaluation criterion

As a result of all these actions one gets the model of knowledge composed of 146 rules.

Rebit inference engine, working in mixed mode and using this knowledgebase, allows for evaluation of each case of suppliers description in much effective way.

C. Business Application Recommendation

The interactive Rebit environment allows for creating, validation, testing and simulation of knowledge bases in a user friendly way. Although there is a possibility to use Rebit System as a standalone solution, it seems that the tighter integration of Rebit components with target environment would be more useful in cases when reasoning and knowledge base are a part of more complex business activity.

It is worth to consider two scenarios of such integration. They differ in scope and depth. The first one is based on SOA architecture of Rebit components. In this scenario rule engine acts as an external, independent component providing services for inference for a selected knowledge base. Knowledge bases may be stored in Rebit or in local repository. This scenario requires an implementation of SOA client and its integration with the target application. The main advantage of this form of integration is that the resulting system consists of loosely-coupled components which are easy to manage and update. Rebit package supports this form of integration by providing library for building a dedicated SOA client.

In the second integration scenario rule engine work as an integral part of the target application. There are two ways to communicate with the rule engine: directly or by means of inter-process communication based on pipes. Since this scenario is moving towards tight integration with the target application, it is recommended to store knowledge bases in local repository. The main limitation of this scenario is that it can only be realized on the .NET platform. As in the previous case, it is necessary to implement the client code. Rebit package provides library supporting integration based on direct access as well as the one based on pipe communication.

VI. Conclusion and Further Works

The problem of evaluation, selection and classification of suppliers is generally considered as one of the more essential issues in practice of enterprises management. Such statement is confirmed by numerous publications. On the other hand, in the organizational documentation (procedures) of all significant enterprises much attention is paid to procedures of supplier selection. Most often one can find in the literature examples of application of the AHP method and related ones, as the most effective ways of solving the problem of multi-criteria evaluation. Therefore, effective methodologies that have the capability of evaluating and continually monitoring suppliers' performance are still needed.

The above mentioned AHP method is the subject of many scientific research studies which confirm its usability and correctness. Nevertheless, there are also critical opinions. Among them the following issues are worth mentioning:

- the existence of large number of pairwise comparisons characteristic for this method brings some limitations on the number of criteria used [2],
- high degree of subjectivity of the evaluations and scale conventionality,
- lack of possibility of verification and reasoning of the evaluations resulting from numerical nature of the aggregation procedures,
- problems related to the phenomenon called rank reversal.

In majority of publications with critical approach to the AHP method a special attention is paid on the problem of objectivation of the evaluation criteria or their more flexible expression, particularly in case when the information on criteria may be deficient, uncertain and incomplete. There are proposed, inter alia, solutions basing on Fuzzy Sets Theory or Rough Sets Theory.

In our opinion an issue of much greater importance, from the point of view of the needs of the users responsible for the management processes, are these restrictions of the AHP method and similar ones which hinder expressing conscious and desirable preferences in the decision making model. While designing the management system of an enterprise as a set of rules, the manager realizes his own preferences and sees no need for their examination. Such situation is diametrically different when compared with research basing on the evaluation of external experts. Experts foresee that e.g. prices can exert influence on the efficiency of the supply greater than quality of the materials. The purchasing officer determines the weight of these criteria.

Therefore, in our opinion, the methods used in Business Rules Management Systems may be successfully applied in case of solving the problem of supplier evaluation and selection. They can be used provided that the tools are expressive enough and there is a possibility of an easy generation of useful business application. The Rebit system presented in the this paper has got these properties.

Rule based approach to the multi-criteria evaluation creates however some problems. They are related with exponentially growing number of examples which should be analyzed in case of formulating the knowledge. Solving this problem through segmentation of the evaluation on increasingly detailed partial criteria creates similar problems as in the AHP method. It is true that explicit presentation of principles of aggregation allows one to avoid most of the problems specific to the methods which, in this case, use computational procedures. Nevertheless, it is a source of flattering of the results of successive aggregation. Therefore, the target of the successive works will be examining, based on the data describing real examples of selection and their results, how big is the scale of this phenomenon and its influence on the correctness and repeatability of decision- making.

REFERENCES

[1] C. Araz and I. Ozkarahan, “Supplier evaluation and management system for strategic sourcing based on a new multicriteria sorting procedure”, *Int. J. Prod. Econ.*, vol. 106, 2007, pp. 585-606.

[2] A. S. Erdem and E. Göçen, “Development of a decision support system for supplier evaluation and order allocation”, *Expert Syst. Appl.*, vol. 39, 2012, pp. 4927-4937.

[3] S. H. Ha and R. Krishnan, “A hybrid approach to supplier selection for the maintenance of a competitive supply chain”, *Expert Syst. Appl.*, vol. 34, 2008, pp. 1303-1311.

[4] L. de Boer, E. Labro and P. Morlacchi, “A review of methods supporting supplier selection”, *Eur. J. Purch. Supply. Manag.*, vol. 7, 2001, pp. 75-89.

[5] G. Bruno, E. Esposito, A. Genovese, Re. Passaro, “AHP-based approaches for supplier evaluation: Problems and perspectives”, *J. Purch. Supply Manag.*, vol. 18, 2012, pp. 159-172.

[6] B. Srdjevic, “Combining different prioritization methods in the analytic hierarchy process synthesis”, *Comput. Oper. Res.*, vol. 32, 2005, pp. 1897-1919.

[7] R-H. Lin, “An integrated FANP-MOLP for supplier evaluation and order allocation”, *Appl. Math. Model.*, vol. 33, 2009, pp. 2730-2736.

[8] J. Chai, J.N.K. Liu and Z. Xu, “A rule-based group decision model for warehouse evaluation under interval-valued intuitionistic fuzzy environments”, *Expert Syst. Appl.*, vol. 40, 2013, pp. 1959-1970.

[9] R. J. Vokurka, J. Choobineh, L. and L. Vadi, “A prototype expert system for the evaluation and selection of potential suppliers”, *Int. J. Oper. Prod. Man.*, vol. 16, 1996, pp. 106-127.

[10] R. Slowinski, S. Greco and B. Matarazzo, “Rough Set and rule-based multicriteria decision aiding”, *Pesqui. Oper.*, vol. 32(2), 2012, pp. 213-269.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1203–1209

Applying Big Data and Linked Data Concepts in Supply Chains Management

Silva Robak
Uniwersytet Zielonogórski, ul. prof. Z. Szafrana 4a, 65-516 Zielona Góra, Poland
Email: s.robak@wmie.uz.zgora.pl

Bogdan Franczyk
Uniwersytet Ekonomiczny we Wrocławiu, ul. Komandorska 118/120, 53-345 Wroclaw, Universität Leipzig, Germany
Email: franczyk@wifa.uni-leipzig.de, bogdan.franczyk@ue.wroc.pl

Marcin Robak
XLogics Sp. z o.o., ul. Kostrzyńska 4, 65-127 Zielona Góra, Poland
Email: m.robak@xlogics.eu, Uniwersytet Zielonogórski, WEIiT, m.robak@weit.uz.zgora.pl

***Abstract*—One of the contemporary problems, and at the same time a big opportunity, in business networks of supply chains are the issues associated with the vast amounts of data arising there. The data may be utilized by the decision support systems in logistics; nevertheless, often there is an information integration problem. The problems with information interchange are related to issues with exchange between independently designed data systems. The networked supply chains will need appropriate IT architectures to support the cooperating business units utilizing structured and unstructured big data and the mechanisms to integrate data in heterogeneous supply chains. In this paper we analyze the capabilities of the big data technology architectures with cloud computing under usage of Linked Data in business process management in supply chains to cope with unstructured near-time data and data silos problems. We present our approach on a 4PL (Fourth-party Logistics) integrator business process example.**

I. Introduction

In the contemporary world the business companies have to face unprecedented challenges. As a result of globalization the amount of data arising in supply chains is raising, the competition is becoming fiercer and the customers often expect integrated services, what requires a close cooperation between several involved organizations. The companies have to adapt to new, such as networked, business models and rethink their role and position in their value chain regarding the potential possibilities given by the utilization of big data to add value for their customer and suppliers. This requires some changes from companies in their organizational view, but at the same time in their information technology view. The appropriate technology environment is needed to support their interoperable business cooperation.

The problem of the appropriate information technology environments for collaborative processes between business participants is twofold. Firstly, the appropriate IT infrastructure for utilization of big data is needed, and secondly, there are data 'siloes' from diverse applications. The last problem has been approached with several solutions like common IT platforms consisting of (possibly common) components based on established standards, standard enterprise information systems, and standard business protocols. Nevertheless, the IT environment platforms still contain proprietary applications like enterprise resource planning systems (ERP), customer relationship management systems (CRM), etc. The foreseen scale of collaboration between business partners may require undertaking further steps for IT environment integration, such as one of the known enterprise application integration solutions or usage of the Web services [1].

The inter-company networks are defined as complex arrays of relationships between companies, which establish these relationships by interacting with each other [2]. Whilst the markets are expanding toward inter-company networks (webs) of collaborating organizations, mentioned above IT integration solution approaches, do not seem to be sufficing and satisfactory. This level of an organizational form of the market participants requires mutual adjustment in information sharing and data management, and further a coordination of collaborative business processes of the supply chain's participants.

In the paper we will approach the problem of possibly advantageous utilization of vast amounts of data (in variety of formats) in supply chains and also the information integration issues in order to overcome the data silo problem. We will investigate the appropriate IT architectures for big data used in association of cloud computing facilities [3] and the utilization of common (open stated) data format as it is offered by Linked Data [4] for data silos integration purposes. We consider the network from a supply chain perspective with emphasis on the value-adding partnerships. The proposal of possible utilization of the Linked Data as an integration solution for business process management BPM in supply chains networks we have already presented in [5]. In this paper we investigate the usage of big data IT architectures, appropriate for supply chains in conjunction with data silos integration possibilities for supply chains on the basis of (open) Linked Data.

A supply chain is defined as a network that comprehends all the organizations and activities associated with the flow and transformation of goods, starting from raw material stage through the whole process, to the end user, as well as the associated information flow [6]. In the paper we will concentrate on the networked supply chain activities and information flow.

In the inter-organizational information systems, which link companies to their suppliers, distributors and customers, a movement of information through electronic links (e.g. XML/EDI - Extensible Markup Language/ Electronic Data Interchange) takes place across organizational boundaries

978-83-60810-53-8/$25.00

between separately owned organizations. It requires not only electronic linkage in form of basic electronic data interchange systems (as for purchase orders, delivery notes, cash flows, etc.), but also interactions between complex cash management systems or by accessing shared technical databases. So the problems with sharing and exchange of information are still viable in supply chains contexts.

The existing EDI standard [7], as message-centric solution, has limited possibilities in enabling interactions in the value chain. The representation of business processes and vocabularies in a domain to potentially automate the trading partners interactions is missing. Another important aspect regarding supply chains networks is integration of additional data from semantic Web applications into logistic systems.

A business process consists of one or more than one related activities that combined together respond to the need for a business action [6]. The processing steps in a workflow might go through numerous data transformations (geographic, technological, linguistic, syntactical and semantic transformations). Communication is an important part of the process and (e-) business processes exist within certain environments. In the dynamic business environment, such as networks of venture participants involved in logistic value chains, where coordination problems in the business process management plays the key role, the appropriate IT architecture, data amount and format, are essential.

Therefore, as stated previously in our paper, we will analyze big data architectures and Linked Data for business processes in supply chains in business networks. For this aim the rest of the paper is organized as follows.

In Section 2 we characterize main features of big data and architectural elements needed by IT infrastructures to support big data during business process management in supply chains. In Section 3 we summarize the Linked Data principles and concepts. In Section 4 we provide an example scenario for 4PL (Fourth-party Logistics) integrator managing the package delivery from Webshops in Asia to customers in Europe. We examine the possible added value resulting from usage of big data and open Linked Data elements that may possibly be useful for the purpose of decision support in supply chain networks. We will also try to show (in Section 5) how the information integration on the base of the Linked Data in conjunction with the big data IT architectures may be applied to achieve the improvement of supply chain environments. In the last Section we conclude our work.

II. Big Data

A. Big Data Features

The big scale usage of available and generated data is made possible for organizations owing to cloud computing paradigms, such as Infrastructure as a Service (IaaS), Storage as a Service (SaaS), which revolutionized the way the computing infrastructures are used [3]. Big data is referred to data that goes beyond the processing capacity of the conventional database systems. In addition to the aspect that it is big (e.g. a huge number of small transactions, or continuous data streams from sensors, mobile devices etc.) it may move too fast, or does not fit the structure of traditional (i.e. relational) database architectures. Big data also may have a low value for further usage before processing it [8].

According to [8] when we denote a big amount of data as "big data" it has to cover the three "Vs" (features) such as: volume, velocity and variety. Other authors (e.g. [9], [10]) add the fourth V-feature: value.

The first feature - volume of big data - denotes its massive character. The big volume of data is beneficial for the data analysts. It may improve the analytics models by having more cases available for forecasts and increase the number of factors to be considered in the models making them more accurate. Nevertheless, the volume bears potential challenge for IT infrastructures to deal with big amounts of data, especially when taking into account its second feature – velocity.

The second feature of big data is the velocity in which data flows into organization or the expected response time to the data. Big data may arrive quickly - in real-time, or near real-time (denoted in this paper as near-time). If data arrives too quickly the IT infrastructures of the organization may be not able to respond timely to it, or even to store all of it. Such situations may lead to data inconsistencies. We will regard further the issue of possible velocity consequences in the next section considering the suitable architectures for big data applications.

The third feature of big data is the variety of data. Big data may have diverse structures and forms, not falling into the rigid relational structures of SQL databases without loss of information. Some of data may be saved as blobs in inside traditional data bases. Therefore the IT infrastructures for big data are denoted as NoSQL, which means that data is "not only SQL" [10]. Several examples for diverse kinds of data are standard business documents, transactional records, and unstructured data in form of images, recordings, HTML documents (web pages), text and email messages, streams from meters and environment sensors, GPS tracks, click streams from Web queries, social media updates, data streams from machines' communication or wearable computing sensors, and many others.

The big data value feature denotes the need for processing it before using it in order to make it valuable for analysis purposes.

B. Big Data Architectures

In the previous section the four characteristic features of big data have been discussed. It is apparent that conventional IT structures may encounter problems with storing variety of data and immediately reacting to it. Firstly, it is because of big data amounts on unstructured data arriving in near-time. The fact that data is unstructured, or rather, it lacks a structure appropriate for storage in conventional SQL databases, implies that other solutions will be needed.

The first issue to consider is the common usage of SQL data bases. IT infrastructures in supply chains include structured data in form of OLTP (Online Transaction Processing) and OLAP (Online Analytical Processing) systems. While the traditional OLTP systems support the transactional systems with highly structured SQL databases, the OLAP sys-

tems contain aggregated historical data in form of cubes. The OLTP systems deliver simple reports, while OLAP systems (known as Data Warehouses) are suited for (traditional) business intelligence applications with reporting facilities on business statistics, performance, etc. on the basis of structured (analytical) historical data. These both databases forms are unsuitable for big data purposes. The data stored there has to have fixed structure, which is conflicting with variety of big data. The OLAP and OLTP contain only high quality data, what is not the situation in case of low value big data (on the opposite to low value big data).

Another problem arises due to the velocity of big data. For the reason that analytical OLAP systems contain only historical data, they are unsuitable for big data applications.

The rigid SQL data structures are insufficient for big data applications, but there are other OLT solutions like "NoSQL OLTP" – MongoDB, AmazonDynamo or Windows Azure Table Storage [10]. This type of database is known as the 'key-value stores' where the data is stored by key and its value is a blob and this solution is widely adopted by enterprises [3].

There is also a known solution for "NoSQL warehousing" for storing and analyzing massive data sets – Apache Hadoop [11]. The Hadoop is a framework for development of open-source software with its own highly distributed HDFS file system, MapReduce framework for writing and executing distributed algorithms and its own query languages – Hive and Pig. The Hadoop components are not only highly distributed but also high tolerant.

Another aspect of big data which is different from SQL databases, is that the results from big data analysis are immediately used and often discarded after that. If not, some bridging solutions are needed, e.g. SQQOP for connecting SQL and Hadoop [10], but they may turn inefficient. Moreover, for handling variety of big data, another solutions like dedicated XML store or graph databases are available [12].

The volume feature of big data can be handled with the usage of capacities and platforms offered by cloud computing [3]. With the pay-as-you-go and low time-to-market solutions, it became affordable even for small organizations.

The big data applications are possible as combinations of diverse technologies (products mash-ups) [12].

The volume, variety and value problems of big data can be tackled by solutions mentioned above. There still remains the dilemma with big data velocity. If data streams arrive quickly in real-time (or near-time) there may be a problem with storing them. One solution possibility would be temporary batch of data in the data pools. This problem is resolved by the Lambda architecture proposed in [13]. Its authors assume that a query is a function on the whole data pool:

```
query = function (all data)
```

Therefore Lambda is a three layers architecture with the batch layer, serving layer and speed layer (see Fig. 1).

The bottom layer (the batch layer) is dedicated to the precomputing on the whole data pool. Its two functions are storing master data set and computation of arbitrary (i.e. any) views. The batch processing principle is well known; for big data the Hadoop is a canonical example [13]. In the Lambda architecture the batch layer continuously recomputes the batch view from scratch.

The middle layer, the serving layer delivers the random accesss to batch views and also is updated by the batch layer.

The top layer, the speed layer deals only with the latest data received during running precomputations in batch layers. According to [13], it compensates for high latency of updates to the serving layer and uses fast incremental algorithms; the batch layer ultimately overrides the speed layer.

In [13], there are big data applications examples as combinations of diverse technologies (products mash-ups) and providers of big data solutions, such as SAP, Oracle, IBM, HP, etc.

With the usage of Hadoop a very fast historical data analysis will be possible based on the Hadoop file system and the MapReduce technology. Nevertheless, it would not be sufficient for some kinds of applications, because the data analysis of current incoming data would be missing. Therefore, in addition to Hadoop we may use the Complex event processing technology (CEP) [14] for dealing with the huge amounts of other real-time data incoming during the running processes. With data supplied to the system and with the appropriate rule sets a dedicated decision support system would be able to react in a suitable way in critical situations, each time when a near-real-time decision will be needed.

The main idea of the introduced concept includes two aspects: the historical context of the Hadoop data, while reacting to the current situation with the CEP technology

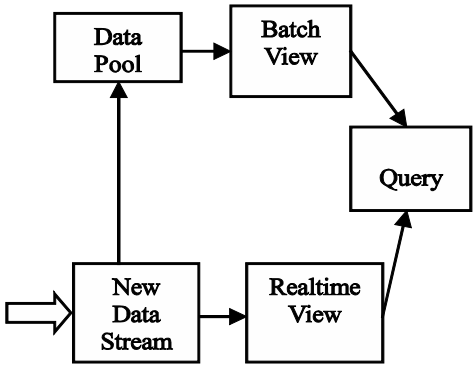


Fig. 1 Lambda architecture diagram [12]

usage. Merging of actual and historical data would add a new value in the decision making processes.

III. LINKED (PPEN) DATA

The Linked Data principles were introduced by Tim Berners-Lee at the TED 2009 presentation [15]. He outlined four rules for making human or machine-readable links for the exploration of web of data. The first rule refers to the usage of Uniform Resource Identifier URI [16] for identification of items (called "things"). The second given rule specifies that only HTTP URIs are meant, so that people can look at them and these can be found by the standard established Domain Name Space (DNS) system. The third rule was formu-

lated for the purpose of providing additional useful information for the items defined by URI. The information should be denoted in a standard format, such as Resource Description Framework RDF* [17], in form of RDF/XML or an alternative serialization (N3, Turtle). The last, fourth rule concerns providing the linkage of such described items with other related items (data), so that the related information on the Web can be discovered more easily.

The further development was the Linked *Open* Data LOD, the concept recommended by the World Wide Web Consortium W3C [4]. It is a star rating system of linked data that allows for proving to which extent the linked data can be regarded as open. The rating is formulated as a five principles scheme, where each next scheme principle extends (for the next star added) the former one by integrating an additional feature. The first principle states that the data should be available on the Web, no matter in which format, but it should be one with an open license. The second principle adds that it should be machine-readable structured data. The third principle adds that the format of the data should be a non-proprietary format. The fourth principle assumes achievement of the former three principles and additionally presumes the usage of an open W3C standard for identification of items, like RDF (RDF/XML, N3 or Turtle) or SPARQL [4] formats (SPARQL Protocol and RDF Query Language) for larger data amounts of data sets. The last principle, required for getting the five star grading assumes contextual linkage of rated data to other resources described in the same way.

IV. 4PL Integrator Example

The development in contemporary logistic networks leans toward possible outsourcing of various logistic functions or services. Further trends include possible integration of outsourced functions/services or even the outsourcing of the whole business processes. At present the dominant are so-called 3PL (Third-party Logistics) solutions [18]. At the next developmental stage the concept of the 4PL (Fourth-party Logistics) emerged. It encompasses the functions offered by a 4PL logistic provider, which is acting as an integrator, assembling the resources, capabilities and technology needed for design, building and running of the comprehensive supply chain solutions [19]. The international 4PL do not need to have their own transport [20]. They may work directly with companies offering transport, or with the 3PL providers, what includes different kinds of carriers, consolidators and forwarders such as ocean carriers, airfreight forwarders and local carriers. The 4PL govern the settlements of the agreements with all involved partners.

As an example of the ideas presented in previous Sections, we consider an example of a 4PL logistic provider which is managing the shipping of commodities bought from the Webshops located in Asia by the customers residing in Europe. The Webshops in our example are located in different Asian cities and offer toys and consumer electronics goods. Our 4PL logistic integrator outsources a warehouse (a hub) in Asia, so that he can consolidate the shipments, which are sent from different Webshops for their further transport to Europe. He also outsources space and commissioning capacity at few hubs in Europe (i.e., near London and Lyon), where the goods first arrive from Asia. The 4PL integrator operates a software platform that integrates the orders from diverse shops and processes the communication with the integrated shipping software dedicated for labeling the shipments for the European carriers. It includes up-to-datedness of the solutions (i.e., carrier-dedicated label layouts) for the European market. Thus, the carrier integration purpose of the 4PL-software fulfills one of most important roles of the 4PL.

Below we will show a standard solution, which is traditionally offered by the 4PL and in a next Section we describe an improvement bringing added value, which could be potentially achieved by the supplementary usage of big data and open Linked Data.

In the basic scenario the 4PL relies on the Webshop`s order data and on the agreements with the freight forwarders and the carriers. At first the goods are ordered by the customers in Europe, who choose a particular European carrier company while ordering products. We consider the situation of delivering of valuable, bulky goods equipped with RFID. The ordered goods are then labeled with the European carrier shipping labels by the Webshops, which download these outputs from the 4PL's IT platform. This results in every single parcel having a shipping label, which fully complies with particular European carrier labeling specification and is augmented with the corresponding Webshop logo.

In the next step the labeled shipments from a particular Webshop are consolidated on palette(s) and brought per freight forwarders to the Asian hub, where the goods are re-consolidated (individual shipments from different Webshops are packed on palettes for a particular hub in Europe) and sent overseas to appropriate European hubs. In Europe the palettes are unpacked and the individual shipments are scanned as ready for a European carrier pickup. The carrier, which was selected by the Webshop customer, will transport the goods within Europe, after they have reached European hub. The carrier's driver receives the printed list with information about the parcels he takes from the hub. At the same time electronic information with the data of the shipment is sent via EDI to the carrier system. From this moment on tracking in Europe is possible, but on the carrier website only.

A. Enhanced 4PL Scenario

The enhanced 4PL's platform scenario performs the same tasks as in the above basic scenario, but the 4PL also gathers a lot of additional data, which will be used for the improvement of the transport decisions through the route.

Every individual shipment prepared by the Webshop is tracked in the 4PL's platform, starting from the point it leaves the Webshop and is picked by the Asian freight forwarder. The freight forwarder provides GPS tracking for the road carried pallet. Further scans are made in the Asian and in the European hubs. The tracking after this point is processed through the carrier tracking system; the corresponding data is imported into the 4PL's IT platform through web service requests, sftp status file transfers or by automated

read outs of the carrier tracking website - depending on the solutions provided by the particular carrier company.

Import and analysis of tracking data from different carriers is a demanding task, which may be supported by Linked Data. Over 17 thousand status event descriptions, which are used by European carriers, can be synthesized to less than one hundred events. Thanks to the full tracking transparency the 4PL is able to collect and analyze the data, and deduct how long it takes to transport goods from point A to point B. In case of delayed shipments, which are reported in his platform in real-time, the 4PL has the possibility of picking out an additional feature such as an express route for further parcel transportation. The gathered tracking data also enables finding out the reasons for the delays, which could help avoiding these in the future.

Another important big data source gathered and analyzed by the 4PL's platform comprise the Asian weather reports and also road and airport traffic reports (available as open Linked Data). Such information is inevitable to support real-time decisions of choosing appropriate transport way (air, road) within Asia. For example, the road transport may turn faster, if the nearest airport is expected to be covered in fog for the next two days.

The 4PL may also use social media and blog data, so that the trends in popularity of e.g. the toys and electronics offered by the Webshops can be regularly observed and evaluated for different regions of Europe. Based on this information, order and transport volumes can be better forecasted, enabling preparation of appropriate transport routes (changing the agreements with the freight forwarders, carriers and warehouses/hubs) in case of forecasted booming or collapsing demand.

Among others tasks, the 4PL has to manage big amounts of data coming from the carriers and other participants in the SCM with different formats and further with diverse semantic interpretation for each identifier. For instance, each carrier has its unique scope of the services (in addition, not always available for all cases) with its own sets of identifiers and furthermore, of the possible package statuses.

Thus, there is a lot data mainly associated with the delivery status of the parcel, denoted in proprietary format. For instance the package, which has been delivered to the client, can get the status "delivered", "closed", "ready", etc. dependent on the carrier firm, etc. So the Integrator has to perform the task of mapping all the unique package status names to one standard format in order to be able to further process the associated data. This way also all the data with the same semantics gets reduced to one uniform internal name and format of the status. Broadly speaking, a company dealing with about hundred carriers in different countries has to understand about 20 thousand possible package statuses, which can be reduced by the 4PL integrator to about one hundred mapped status descriptions. These will be further needed in the supply chain EDI data exchange. The usage of the Linked Data could facilitate the mapping of the equivalent data, not only on the 4PL side, but also among other supply chain participants.

In the next Section we show the possibility of integration of logistic data by using open Linked Data facilities.

V. Data Integration With Linked Data In Value Chains

As stated at the previous Sections, the process steps in a workflow could undertake numerous transformations of data. A common format could improve the communication between the participants collaborating in the process environments and serve as a broker between SQL and NoSQL data, especially in the big data environments.

In our example, the source data acquired from a Webshop is, until delivery of the commodities to the customer, administered in the subsequent stages, changing format and being adjusted and enriched through numerous additional transformations, which are needed for accomplishment of the activities of the participants in a joint business venture.

The Semantic Web concept supports the basic idea of the Web considered as an open community sharing information around the world. As pointed out in our hypothetic Webshop parcels example, one part of the data integrated into the data exchange flow of business networks could be the data supplied into the 4PL integrator software platform from Web applications like the Geo, metrological or traffic data, partially enriched with semantic information described with RDF triples.

Since there still are no known established business solutions successfully working on the base of ontologies we consider application of Linked Data concepts as a more of a lightweight solution than the semantic description of data which is exchanged in networks connected through Internet and enriched with data from web applications like OpenStreetMap [22] or DBpedia [23].

The information assumed is to be presented as Linked Open Data, presumes that the data not in a proprietary format. Therefore it is important that the communication software (e.g. of 4PL) will support open data formats, such as CSV (Comma-separated values) [24] or transformation to such an open format.

The data exchanged between supply chain participants may be enriched with semantic information by means of the RDF graphs. At present time information is stored mostly in relational databases. There are some solutions for data transformations, i.e. Triplify for transforming of the data stored in transactional SQL databases into RDF representations [25]. Other possibilities, like object serialization or hierarchical representation, should be mapped into the graph data models. Meanwhile there are multiple semantic database implementations known, such as Triplestores, a purpose-built database type dedicated for the storage and retrieval of triples, e.g. Virtuoso [26]. In addition to queries the triples can be imported and exported using RDF or other formats.

The mapping between customized IT solutions and different data formats into the triple representation could be undertaken by means of dedicated software.

The usage of open formats with RDF-defined semantic could support easier data entries into the digital value chains. The enrichment of data with the semantic informa-

tion can help with communication and mediation between multiple points. The semantic enriched (big) data stored in an open format can be made widely available for the participants of the value chains if it could be further managed by using of cloud computing – the web-based, dynamical IT services. Cloud computing solutions moreover warrant the security on the infrastructure and data level, and also eliminate the need of initial investments in IT infrastructures and shorten the time-to-market.

The usage of open formats may considerably contribute to rising of flexibility and content transfer within supply chains, organized as webs, and simplifying the data transformation into diverse e-business standards.

The drawback of the given approach is the need of its integration into various IT solutions. To be useful it should be supported by the numerous diverse E-business standards as shown in [5].

VI. Conclusion

In the past times the vendors had to exploit earlier period's structured data (stored in SQL OLTP or OLAP systems) in order to analyze customer's attitudes and increase sales. Nowadays, a raising all-embracing connectivity with potentially all stakeholders in supply chains networks results in the possibility of accessing to all needed current data in real-time and also in getting a near-time feedback. This bears the genuine chances for almost immediate improvement of the relationships with the supply chain's stakeholders and therefore increases the agility and ability for just-in-time in reactions to the changing requirements [27]. Accordingly, the high quality decision support becomes possible, which enables achieving optimal performance.

It became feasible to take advantage of this situation facilitated owing to usage of cloud computing and big data by the organizations taking part in the supply chain networks. Nevertheless, the amassed data encountered in the supply chains demand solutions for suited processing of coexistent structured and unstructured data (NoSQL) on the base of appropriate software architectures, and also require a common base as the exchange format of the data shared and exchanged in the supply chains networks.

In the paper we have analyzed the nowadays common solutions for structured and unstructured data storage options for the decision making support. With the opportunity of big data and cloud computing technologies application, the amount of partially unstructured data increases and it needs to be taken into account while making logistic decisions the previous solutions are not enough to cope with the problem dealing with them in the real time or in the near-time.

We see big chances in using such architectures as the layered Lambda architecture (for big data processing), which was designed to cope with the near-time exploitation of big amounts of data arriving. The data exchanged in supply chains has diverse formats, therefore we further propose using an open common data format in supply chains. For an additional advantage a standard solution with Linked Data may be further enriched with semantic information for further support of supply chain collaboration.

It is expected that the application of open Linked Data may substantially support the automated extraction of the information published on the Web by using open standards and additionally describing with semantic meaning and contextual relationships of the data.

We have shown on the 4PL integrator example the need for the integration of social media data for forecasting aims. Also diverse Linked Data from Virtuoso databases can be applied in the decision making support. As a common standard for data exchange between the various IT applications interacting in the logistic value chains we have considered incorporating the semantic concepts associated with the Linked Data into the supply chain management, especially for the aims of the common format integration between the SQL data silos in the value chains using big data.

The usage of the Linked Data by a broker may contribute to the data integration and transfer speed up. The saving of the costs previously needed for the transformations between diverse formats will create the added value for the network participants.

The suggested improvements raise new possibilities for adding value in supply chains. The network effect causes that with increased number of participations the added value for the participants of the network grows.

The Lambda architecture allows merging huge amounts of historical data with near real-time data creating context-oriented data needed for reacting and appropriate responding to different situations like transport events in the logistics.

In the paper we presented the ongoing research work. In the future work the further aspects like economic evaluation of Applying Big Data and Linked Data Concepts in Supply Chain Management should be considered. Also the integration of the further open Linked Data instances from the Virtuoso databases into the supply chains may be investigated.

References

[1] Web Services Activity, W3C Working Group, http://www.w3.org/ws

[2] J. C. Jarillo, "On strategic networks", in *Strategic Management Journal*, 9, 1988.

[3] D. Agrawal, S. Das and A. E. Abbadi, „Big data and cloud computing: current state and future opportunities". *EDBT 2011*, March 22-24, 2011, Uppsala, Sweden. ACM 978-1-4503-0528-0/11/0003.

[4] W3C LinkedData, 2011, www.w3.org/wiki/LinkedData

[5] S. Robak, B. Franczyk, and M. Robak, Applying Linked Data concepts in BPM, *Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS*. Wrocław 2012. IEEE Conference Publications, ISBN: 978-1-4673-0708-6, pp. 1105-1110.

[6] M. P. Papazoglou, and P. M. A Ribbes, *E-business: organizational and technical foundations, John Wiley and sons*. London 2006, pp.88-90.

[7] M. Kantor, and J. Burrows, *Electronic Data Interchange (EDI)*, National Institute of Standards and Technology, 1996.

[8] E. Dumbill, "*What is big data? An introduction to the big data landscape*", Strata O'Reilly, 11 January 2012, http://strata.oreilly.com/2012/01/what-is-big-data.html

[9] S. Wrobel, „*Big Data – Vorsprung durch Wissen*", Fraunhofer-Institut für Intelligente Analyse- und Informationsverarbeitungsysteme IAIS. Presentation, www.iais.fraunhofer.de, 2012.

[10] I. Mitchell and M. Wilson, "Linked Data. Connecting and exploiting big data", Fujitsu Services Limited, March 2012, www.fujitsu.com.uk.

[11] The Apache Hadoop Project. http://hadoop.apache.org/core/ , 2009.

[12] M. May, "*Living Big Data. Konzeption einer Experimentierplattform*". Fraunhofer-Institut für Intelligente Analyse- und Informationsverar-

beitungsysteme IAIS. Presentation, www.iais.fraunhofer.de, Berlin, 2012.
[13] N. Marz and J. Warren, "*Big data. Principles and best practices of scalable realtime data systems"*. Manning Publications, MEAP Edition, Manning Early Access Program Big Data version 7, 2012.
[14] D. C. Luckham, *Event processing for business: organizing the real-time enterprise*. Hoboken, New Jersey: John Wiley & Sons, Inc., p.3. 2012.
[15] T. Berners-Lee, "On the next Web", Talk on the TED Conference 2009, www.ted.com/talks/tim_berners_lee_on_the_next_web.html
[16] W3C Architecture domain, Naming and Addressing: URIs, URLs, …, http://www.w3.org/Addressing/
[17] W3C Semantic Web, RDF Working Group, *Resource Description Framework (RDF)*, 2004, www.w3.org/RDF
[18] H. Baumgarten, *Das beste der Logistik*. Springer Verlag, Berlin 2008.
[19] S. Chopra, and P. Meindl, "*Supply chain management: strategy planning, and operation*", 3rd ed., Prentice Hall, 2007, pp.427.
[20] A. Matopoulos, and E. -M. Papadopoulou, "The evolution of logistics service providers and the role of Internet-based applications in facilitating global operations" in *Enterprise Networks and Logistics for Agile Manufacturing*, L. Wang, and S.C.L. Koh, Eds., Springer, 2007, pp. 298-304.
[21] D. Allemang, and J. Hendler, *Semantic Web for the working ontologist modeling in RDF, RDFS and OWL*. Morgan Kaufman, 2008.
[22] OpenStreetMap Project, available at www.openStreetMap.org
[23] DBpedia Project, Free University of Berlin, and University of Leipzig, OpenLink Software, http://wiki.dbpedia.org/About
[24] MastPoint, *CSV-1203, CSV File Format Specification*, Best practice for business-to-business operations, available at http://mastpoint.curzonnassau.com/csv-1203/index.html
[25] Triplify expose semantics!, http://www.triplify.org/Overview
[26] Virtuoso Universal Server, Universal Server Platform for Enterprise Data Integration, Web Services, & Process Orchestration, openLink software, available at http://virtuoso.openlinksw.com/
[27] R. Sethuraman and S. K. Kundharaju, "Top 7 Tips for Big data to optimize Supply Chains. 5 Februar 2013, http://risnews.edgl.com/retail-trends/Top-7-Tips-for-Utilizing-Big-Data-to-Optimize-Supply-Chains86163.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1211–1218

A hybrid approach to supply chain modeling and optimization

Paweł Sitek
Kielce University of Technology
Al. 1000-lecia PP 7, 25-314
Kielce, Poland Institute of
Management Control Systems
e-mail:sitek@tu.kielce.pl

Jarosław Wikarek
Kielce University of Technology
Al. 1000-lecia PP 7, 25-314
Kielce, Poland Institute of
Management Control Systems
e-mail:j.wikarek@tu.kielce.pl

***Abstract*—The paper presents the concept and an outline of the implementation of a hybrid approach to supply chain modeling and optimization. Two environments mathematical programming (MP) and logic programming (LP) were integrated. The strengths of integer programming (IP) and constraint logic programming (CLP), in which constraints are treated in a different way and different methods are implemented, were combined to use the strengths of both. The proposed approach is particularly important for the decision models with an objective function and many discrete decision variables added up in multiple constraints. In order to verify the proposed approach, the optimization models were presented and implemented in both traditional mathematical programming and the hybrid environment.**

I. Introduction

A SUPPLY chain is referred to as an integrated systems which synchronizes a series of related business processes in order to acquire raw materials and parts, transform them into finished products and distribute to customers and retailers. The supply chain plays an important role in the automotive, electronics and food industries.

Huang et al. [1] studied information sharing in supply chain management. They considered and proposed four classification criteria: supply chain structure, decision level, modeling approach and shared information.

Supply chain structure: It defines the way various organizations within the supply chain are arranged and related to each other.

Decision level: Three decision levels may be distinguished in terms of the decision to be made: strategic, tactical and operational, with their corresponding period, i.e., long-term, mid-term and short-term.

Supply chain analytical modeling approach: This approach focuses on type of representation, in this case, mathematical relationships, and the aspects to be considered in the supply chain. The literature mostly describes and discusses mathematical programming-based modeling: linear programming, integer programming or mixed integer linear programming models [2]–[6]. Minimization of integrated costs is the main purpose of the models presented in the literature [6]–[10]. Maximization of revenues or sales is considered to a lesser extent [4], [11].

Shared information: Information is shared between network nodes determined by the model. This enables production, distribution, inventory and transport planning, depending on the purpose. The information sharing process is a vital aspect in an effective supply chain. The following groups of parameters are taken into account: resources, inventory, production, transport, demand, time, etc.

This paper focuses on the modeling approach to optimization problems in supply chain. A type of representation together with aspects to consider in the supply chain makes up a modeling approach. The vast majority of the works reviewed have formulated their models as linear programming (LP), integer programming (IP) and mixed integer linear programming (MILP) problems and solved them using the Operations Research methods. Nonlinear programming, multi-objective programming, fuzzy programming with stochastic programming are used much less frequently [12].

Problems related to the design, integration and management of the supply chain affect many aspects of production, distribution, warehouse management, supply chain structure, transport modes etc. Those problems are usually closely related to each other, some may influence one another to a greater or lesser extent. Because of the interconnectedness and a very large number of different constraints: resource, time, technological, and financial, the constraint–based environments are suitable for producing "natural" solutions for highly combinatorial problems. In the literature, references to modeling and optimizing supply chain problems using constraint–based environments are relatively few in number [11], [12].

A. Constraint-based environments

Constraint satisfaction problems (CSPs), constraint programming (CP) and constraint logic programming (CLP) [13]–[15] offer a very good framework for representing the knowledge and information needed to deal with supply chain problems.

Constraint satisfaction problems (CSPs) are mathematical problems defined as a set of elements whose state must satisfy a number of constraints. CSPs represent the entities in a problem as a homogeneous collection of finite constraints over variables, which are solved by constraint satisfaction methods. CSPs are the subject of intense study in both artificial intelligence and operations research, since the regularity in their formulation provides a common basis

978-83-60810-53-8/$25.00

to analyze and solve problems of many unrelated families [13]. Formally, a constraint satisfaction problem is defined as a triple (X,D,C), where X is a set of variables, D is a domain of values, and C is a set of constraints. Every constraint is in turn a pair (t,R) (usually represented as a matrix), where t is an n-tuple of variables and R is an n-ary relation on D. An evaluation of the variables is a function from the set of variables to the domain of values, $v:X \rightarrow D$. An evaluation v satisfies constraint $((x_1,\ldots,x_n),R)$ if $(v(x_1),..v(x_n)) \in R$. A solution is an evaluation that satisfies all constraints.

Constraint satisfaction problems on finite domains are typically solved using a form of search. The most used techniques are variants of backtracking, constraint propagation, and local search. Our experience as well as that of other researchers, confirms that constraint propagation is central to the process of solving a constraint problem [13], [14], [16]. Constraint propagation embeds any reasoning that consists in explicitly forbidding values or combinations of values for some decision variables of a problem because a given subset of its constraints cannot be satisfied otherwise.

CSPs are often used in constraint programming. Constraint programming is the use of constraints as a programming language to encode and solve problems. Constraint logic programming is a form of constraint programming, in which logic programming is extended to include concepts from constraint satisfaction. A constraint logic program is a logic program that contains constraints in the body of clauses. Constraints can also be present in the goal. These environments are declarative.

B. Organization and structure of the paper

In this paper, we focus on the problem of hybrid modeling and optimization of the supply chain problems in the hybrid environment. We propose a novel approach to supply chain modeling and optimization by developing integrated models and methods using the complementary strengths of MILP and CP/CLP (II, III). In this approach, both the hybrid model (V) and the hybrid framework (IV) to its efficient solution were developed.

In order to verify the proposed approach, the optimization model in mixed linear integer programming (MILP) was created and implemented in traditional (IP) and hybrid approaches. Finally, the hybrid model was optimized in the hybrid framework (V).

II. MOTIVATION

Based on [1], [2], [13], [15], [16] and our previous work [14], [17], [18] we observed some advantages and disadvantages of these environments.

An integrated approach of constraint programming (CP) and mixed integer programming (MIP) can help to solve optimization problems that are intractable with either of the two methods alone [20]–[22]. Although Operations Research (OR) and Constraint Programming (CP) have different roots, the links between the two environments have grown stronger in recent years.

Both MIP/MILP/IP and finite domain CP/CLP involve variables and constraints. However, the types of the variables and constraints that are used, and the way the constraints are solved, are different in the two approaches [23], [24].

III. STATE OF THE ART

As mentioned earlier, the vast majority of decision-making models for the problems of production, logistics, supply chain are formulated in the form of mathematical programming (MIP, MILP, IP).

Due to the structure of these models (adding together discrete decision variables in the constraints and the objective function) and a large number of discrete decision variables (integer and binary), they can only be applied to small problems. Another weakness is that only linear constraints can be used. In practice, the issues related to the production, distribution and supply chain constraints are often logical, nonlinear, etc. For these reasons the problem was formulated in a new way

In our hybrid approach to modeling and optimization supply chain problems, we proposed the environment, where:

- knowledge related to supply chain can be presented in a linear and logical constraints (implement all types of constraints of previous MILP/MIP models [18], [19] and introduce new types of constraints (logical, nonlinear, symbolic etc.));
- the optimization model solved by using the framework can be formulated as a pure MILP/MIP model, a CP/CLP model or as a hybrid model;
- the novel method of constraint propagation is introduced (obtained by transformation of the optimization model to explore its structure (feasible routes, capacities, etc.));
- constrained domains of decision variables, new constraints and values for some variables are transferred from CP/CLP to MILP/MIP;
- the efficiency of finding solutions to the problems of larger sizes is increased.

As a result, we obtained the hybrid optimization environment that ensures a better and easier way of modeling and optimization, and more effective search solution for a certain class of optimization problems. This class includes quantitative models related to costs, customer service and inventories. Models of this class are characterized by adding up many discrete decision variables in both constraints and the objective function.

IV. HYBRID OPTIMIZATION ENVIRONMENT

In order to implement all the assumptions and requirements outlined in the previous chapter, both constraint logic programming (CLP) and integer programming (MILP/MIP) had to be combined and linked.

The hybrid environment consists of MILP/CLP/Hybrid models and hybrid optimization framework to solve them (Fig. 1). The concept of this framework and its phases (P1 .. P5, G1..G3) are presented in Fig. 2.

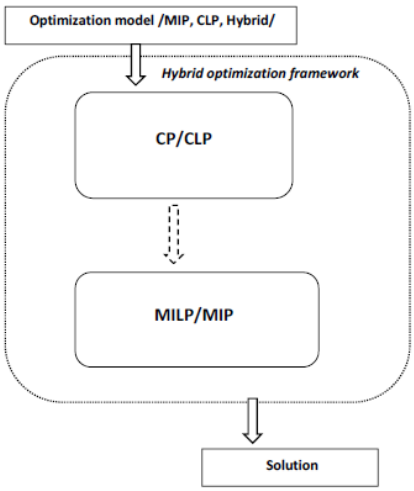


Fig. 1 Scheme of the hybrid optimization environment

The details of the hybrid environment have been discussed in [24]. The motivation was to offer the most effective tools for model–specific constraints and solution efficiency.

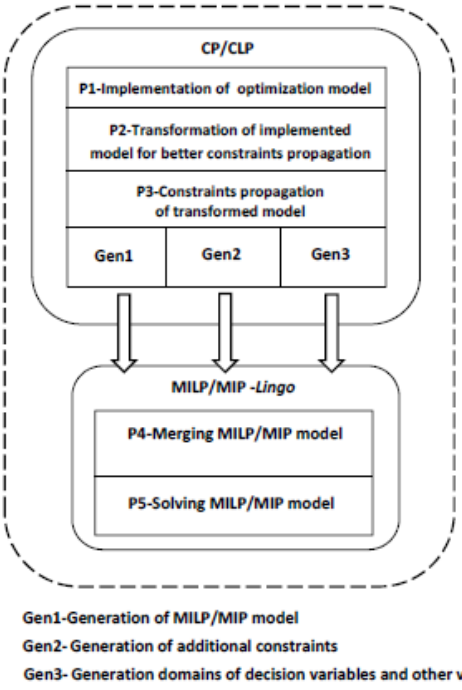


Fig. 2 Scheme of the hybrid optimization framework

The constraints propagation of the transformed model (phase–P3) largely affected the efficiency of the solution. Therefore phase P2 was introduced. During this phase, the transformation was performed using the structure and properties of the model. The details of this transformation are described in the following chapter. From a variety of tools for the implementation of the CP/CLP environment, ECLiPSe software [25] was selected. ECLiPSe is an open-source software system for the cost–effective development and deployment of constraint programming applications. Environment for the implementation of MILP/MIP was LINGO by LINDO Systems. LINGO Optimization Modeling Software is a powerful tool for building and solving mathematical optimization models [26].

V. EXAMPLES OF SUPPLY CHAIN OPTIMIZATION

The proposed HSE environment was verified and tested on two models.

First model was formulated as a mixed linear integer programming (MILP) problem [18], [19] under constraints (2) .. (23) in order to test the proposed environment (Fig. 1) against the classical integer programming environment [26]. Then the hybrid model (1) .. (25) was implemented and solved. Indices, parameters and decision variables used in the models together with their descriptions are summarized in Tab. 1. The simplified structure of the supply chain network for this model, composed of producers, distributors and customers is presented in Fig. 3.

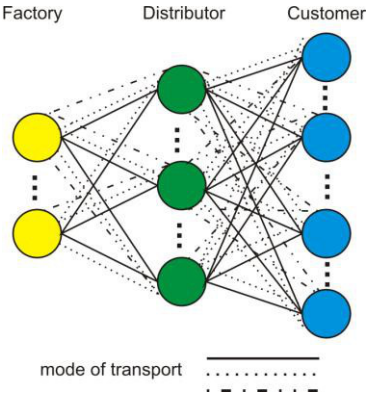


Fig. 3 The simplified structure of the supply chain network

Both models are the cost models that take into account three other types of parameters, i.e., the spatial parameters (area/volume occupied by the product, distributor capacity and capacity of transport unit), time (duration of delivery and service by distributor, etc.) and the transport mode.

The main assumptions made in the construction of these models were as follows:

- the shared information process in the supply chain consists of resources (capacity, versatility, costs), inventory (capacity, versatility, costs, time), production (capacity, versatility, costs), product (volume), transport (cost, mode, time), demand, etc;
- part of the supply chain has a structure as in Fig. 3.;
- transport is multimodal (several modes of transport, a limited number of means of transport for each mode);
- the environmental aspects of use of transport modes are taken into account;
- different products are combined in one batch of transport;
- the cost of supplies is presented in the form of a function (in this approach, linear function of fixed and variable costs);

- the models have linear or linear and logical constraints.

TABLE I
INDICES, PARAMETERS AND DECISION VARIABLES

Symbol	Description
Indices	
k	product type (k=1..O)
j	delivery point/customer/city (j=1..M)
i	manufacturer/factory (i=1..N)
s	distributor /distribution center (s=1..E)
d	mode of transport (d=1..L)
N	number of manufacturers/factories
M	number of delivery points/customers
E	number of distributors
O	number of product types
L	number of mode of transport
Input parameters	
F_s	the fixed cost of distributor/distribution center s
P_k	the area/volume occupied by product k
V_s	distributor s maximum capacity/volume
$W_{i,k}$	production capacity at factory i for product k
$C_{i,k}$	the cost of product k at factory i
$R_{s,k}$	if distributor s can deliver product k then $R_{s,k}$=1, otherwise $R_{s,k}$=0
$Tp_{s,k}$	the time needed for distributor s to prepare the shipment of product k
$Tc_{j,k}$	the cut-off time of delivery to the delivery point/customer j of product k
$Z_{j,k}$	customer demand/order j for product k
Zt_d	the number of transport units using mode of transport d
Pt_d	the capacity of transport unit using mode of transport d
$Tf_{i,s,d}$	the time of delivery from manufacturer i to distributor s using mode of transport d
$K1_{i,s,k,d}$	the variable cost of delivery of product k from manufacturer i to distributor s using mode of transport d
$R1_{i,s,d}$	if manufacturer i can deliver to distributor s using mode of transport d then $R1_{i,s,d}$=1, otherwise $R1_{i,s,d}$=0
$A_{i,s,d}$	the fixed cost of delivery from manufacturer i to distributor s using mode of transport d
$Koa_{s,j,d}$	the total cost of delivery from distributor s to customer j using mode of transport d
$Tm_{s,j,d}$	the time of delivery from distributor s to customer j using mode of transport d
$K2_{s,j,k,d}$	the variable cost of delivery of product k from distributor s to customer j using mode of transport d
$R2_{s,j,d}$	if distributor s can deliver to customer j using mode of transport d then $R2_{s,j,d}$=1, otherwise $R2_{s,j,d}$=0
$G_{s,j,d}$	the fixed cost of delivery from distributor s to customer j using mode of transport d
$Kog_{s,j,d}$	the total cost of delivery from distributor s to customer j using mode of transport d
Od_d	the environmental cost of using mode of transport d
Decision variables	
$X_{i,s,k,d}$	delivery quantity of product k from manufacturer i to distributor s using mode of transport d
$Xa_{i,s,d}$	if delivery is from manufacturer i to distributor s using mode of transport d then $Xa_{i,s,d}$=1, otherwise $Xa_{i,s,d}$=0
$Xb_{i,s,d}$	the number of courses from manufacturer i to distributor s using mode of transport d
$Y_{s,j,k,d}$	delivery quantity of product k from distributor s to customer j using mode of transport d
$Ya_{s,j,d}$	if delivery is from distributor s to customer j using mode of transport d then $Ya_{s,j,d}$ =1, otherwise $Ya_{s,j,d}$ =0
$Yb_{s,j,d}$	the number of courses from distributor s to customer j using mode of transport d
Tc_s	if distributor s participates in deliveries, then Tc_s=1, otherwise Tc_s=0
CW	Arbitrarily large constant

A. Objective function

The objective function (1) defines the aggregate costs of the entire chain and consists of five elements. The first element comprises the fixed costs associated with the operation of the distributor involved in the delivery (e.g. distribution centre, warehouse, etc.). The second element corresponds to environmental costs of using various means of transport. Those costs are dependent on the number of courses of the given means of transport, and on the other hand, on the environmental levy, which in turn may depend on the use of fossil fuels and carbon-dioxide emissions.

The third component determines the cost of the delivery from the manufacturer to the distributor. Another component is responsible for the costs of the delivery from the distributor to the end user (the store, the individual client, etc.). The last component of the objective function determines the cost of manufacturing the product by the given manufacturer.

Formulating the objective function in this manner allows comprehensive cost optimization of various aspects of supply chain management. Each subset of the objective function with the same constrains provides a subset of the optimization area and makes it much easier to search for a solution.

$$\sum_{s=1}^{E} F_s \cdot Tc_s + \sum_{d=1}^{L} Od_d \left(\sum_{i=1}^{N}\sum_{s=1}^{E} Xb_{i,s,d} + \sum_{s=1}^{E}\sum_{j=1}^{M} Yb_{j,s,d}\right) + \sum_{i=1}^{N}\sum_{s=1}^{E}\sum_{d=1}^{L} Koa_{i,s,d} + \sum_{s=1}^{E}\sum_{j=1}^{M}\sum_{d=1}^{L} Kog_{s,j,d} + \sum_{i=1}^{N}\sum_{k=1}^{O}\left(C_{ik} \cdot \sum_{s=1}^{E}\sum_{d=1}^{L} X_{i,s,k,d}\right) \quad (1)$$

B. Constraints

The model was based on constraints (2) .. (24) Constraint (2) specifies that all deliveries of product k produced by the manufacturer i and delivered to all distributors s using mode of transport d do not exceed the manufacturer's production capacity.

Constraint (3) covers all customer j demands for product k ($Z_{j,k}$) through the implementation of delivery by distributors s (the values of decision variables $Y_{i,s,k,d}$). The flow balance of each distributor s corresponds to constraint (4). The possibility of delivery is dependent on the distributor's technical capabilities - constraint (5). Time constraint (6) ensures the terms of delivery are met. Constraints (7a), (7b), (8) guarantee deliveries with available transport taken into account. Constraints (9), (10), (11) set values of decision variables based on binary variables Tc_s, $Xa_{i,s,d}$, $Ya_{s,j,d}$. Dependencies (12) and (13) represent the relationship based on which total costs are calculated. In general, these may be any linear functions. The remaining constraints (14)..(23) arise from the nature of the model (MILP).

Constraint (24) allows the distribution of exclusively one of the two selected products in the distribution center s. Similarly, constraint (25) allows the production of exclusively one of the two selected products in the factory i.

Those constraints result from technological, marketing, sales or safety reasons. Therefore, some products cannot be distributed and/or produced together. The constraint can be re-used for different pairs of product k and for some of or all

distribution centers s and factories i. A logical constraint like this cannot be easily implemented in a linear model. Only declarative application environments based on constraint satisfaction problem (CSP) make it possible to implement constraints such as (24), (25).

The addition of constraints of that type changes the model class. It is a hybrid model.

$$\sum_{s=1}^{E}\sum_{d=1}^{L} X_{i,s,k,d} \cdot R_{s,k} \le W_{i,k} \text{ for } i=1..N, k=1..O \tag{2}$$

$$\sum_{s=1}^{E}\sum_{d=1}^{L} (Y_{s,j,k,d} \cdot R_{s,k}) \ge Z_{j,k} \text{ for } j=1..M, k=1..O \tag{3}$$

$$\sum_{i=1}^{N}\sum_{d=1}^{L} X_{i,s,k,d} = \sum_{j=1}^{M}\sum_{d=1}^{L} Y_{s,j,k,d} \text{ for } s=1..E, k=1..O \tag{4}$$

$$\sum_{k=1}^{O} (P_k \cdot \sum_{i=1}^{N}\sum_{d=1}^{L} X_{i,s,k,d}) \le Tc_s \cdot V_s \text{ for } s=1..E \tag{5}$$

$$Xa_{i,s,d} \cdot Tf_{i,s,d} + Xa_{i,s,d} \cdot Tp_{s,k} + Ya_{s,j,d} \cdot Tm_{s,j,d} \le Tc_{j,k} \text{ for } i=1..N,\ s=1..E, j=1..M,\ k=1..O, d=1..L \tag{6}$$

$$R1_{i,s,d} \cdot Xb_{i,s,d} \cdot Pt_d \ge X_{i,s,k,d} \cdot P_k \text{ for } i=1..N, s=1..E, k=1..O, d=1..L \tag{7a}$$

$$R2_{s,j,d} \cdot Yb_{s,j,d} \cdot Pt_d \ge Y_{s,j,k,d} \cdot P_k \text{ for } s=1..E,\ j=1..M,\ k=1..O, d=1..L \tag{7b}$$

$$\sum_{i=1}^{N}\sum_{s=1}^{E} Xb_{i,s,d} + \sum_{j=1}^{M}\sum_{s=1}^{E} Yb_{j,s,d} \le Zt_d \text{ for } d=1..L \tag{8}$$

$$\sum_{i=1}^{N}\sum_{d=1}^{L} Xb_{i,s,d} \le CW \cdot Tc_s \text{ for } s=1..E \tag{9}$$

$$Xb_{i,s,d} \le CW \cdot Xa_{i,s,d} \text{ for } i=1..N, s=1..E, d=1..L \tag{10}$$

$$Yb_{s,j,d} \le CW \cdot Ya_{s,j,d} \text{ for } s=1..E, j=1..M, d=1..L \tag{11}$$

$$Koa_{i,s,d} = A_{i,s,d} \cdot Xb_{i,s,d} + \sum_{k=1}^{O} K1_{i,s,k,d} \cdot X_{i,s,k,d} \text{ for } i=1..N, s=1..E, d=1..L \tag{12}$$

$$Kog_{s,j,d} = G_{s,j,d} \cdot Yb_{js,d} + \sum_{k=1}^{O} K2_{s,j,k,d} \cdot Y_{s,j,k,d} \text{ for } s=1..E, j=1..M, d=1..L \tag{13}$$

$$X_{i,s,k,d} \ge 0 \text{ for } i=1..N, s=1..E, k=1..0,\ d=1..L \tag{14}$$

$$Xb_{i,s,d} \ge 0 \text{ for } i=1..N, s=1..E, d=1..L, \tag{15}$$

$$Yb_{s,j,d} \ge 0 \text{ for } s=1..E, j=2..M,\ d=1..L, \tag{16}$$

$$X_{i,s,k,d} \in C \text{ for } i=1..N, s=1..E,\ k=1..0, d=1..L, \tag{17}$$

$$Xb_{i,s,d} \in C \text{ for } i=1..N, s=1..E, d=1..L \tag{18}$$

$$Y_{s,j,k,d} \in C \text{ for } s=1..E, j=1..M,\ k=1..0, d=1..L \tag{19}$$

$$Yb_{s,j,d} \in C \text{ for } s=1..E, j=1..M,\ d=1..L, \tag{20}$$

$$Xa_{i,s,d} \in \{0,1\} \text{ for } i=1..N, s=1..E, d=1..L, \tag{21}$$

$$Ya_{s,j,d} \in \{0,1\} \text{ for } s=1..E, j=2..M,\ d=1..L, \tag{22}$$

$$Tc_s \in \{0,1\} \text{ for } s=1..E \tag{23}$$

$$\text{ExclusionD}(X_{i,s,k,d}, X_{i,s,l,d}, s) \text{ for } k \neq l,\ s=1..S \tag{24}$$

$$\text{ExclusionP}(X_{i,s,k,d}, X_{i,s,l,d}, i) \text{ for } k \neq l,\ i=1..N \tag{25}$$

C. Model transformation

Due to the nature of the decision problem (adding up decision variables and constraints involving a lot of variables), the constraint propagation efficiency decreases dramatically. Constraint propagation is one of the most important methods in CLP affecting the efficiency and effectiveness of the CLP and hybrid optimization environment (Fig. 1). For that reason, research into more efficient and more effective methods of constraint propagation was conducted. The results included different representation of the problem and the manner of its implementation. The classical problem modeling in the CLP environment consists in building a set of predicates with parameters. Each CLP predicate has a corresponding multi-dimensional vector representation. While modeling both problems, (1) .. (23) and (1) .. (25), quantities i, s, k, d and decision variable $X_{i,s,k,d}$ were vector parameters. The process of finding the solution may consist in using the constraints propagation methods, labeling of variables and the backtracking mechanism. The quality of constraints propagation and the number of backtrackings are affected to a high extent by the number of parameters that must be specified/labeled in the given predicate/vector. In both models presented above, the classical problem representation included five parameters: i, s, k, d and $X_{i,s,k,d}$. Considering the domain size of each parameter, the process is complex and time-consuming. Our idea was to transform the problem by changing its representation without changing the very problem. All permissible routes were first generated based on the fixed data and a set of orders, then the specific values of parameters i, s, k, d were assigned to each of the routes. In this way, only decision variables $X_{i,s,k,d}$ (deliveries) had to be specified. This transformation fundamentally improved the efficiency of the constraint propagation and reduced the number of backtracks. A route model is a name adopted for the models that underwent the transformation.

D. Decision-making support

The proposed models can support decision-making in the following areas:

- the optimization of total cost of the supply chain (objective function, decision variables-Appendix A2);
- the selection of the transport fleet number, capacity and modes for specific total costs;
- the sizing of distributor warehouses and the study of their impact on the overall costs (Appendix A3, Fig. 4, Fig. 5, Tab. V);
- the selection of transport routes for optimal total cost.

Detailed studies of these topics are being conducted and will be described in our future articles. We use the hybrid approach to both modeling and solving.

VI. NUMERICAL EXPERIMENTS

In order to verify and evaluate the proposed approach, many numerical experiments were performed. All the examples relate to the supply chain with two manufacturers (i=1..2), three distributors (s=1..3), five customers (j=1..5), three modes of transport (d=1..3), and ten types of products (k=1..10). Experiments began with three examples of P1, P2, P3 for the optimization MILP model (1) .. (23). The examples differ in terms of capacity available to the distributors s (V_s), the number of transport units using the mode of transport d (Zt_d) and the number of orders (No). The first series of experiments was designed to show the benefits and advantages of the hybrid approach. For this purpose the

model (1) .. (23) was implemented in both the hybrid and integer programming environments. In addition, hybrid implementation of the transformed model was performed with and without constraint propagation, (MILPT2) and (MILPT1), respectively. The experiments that follow were conducted to optimize examples P4, P5, which are implementations of the model (1) .. (25) for the hybrid approach. Examples P4, P5 were obtained from P1, P3 by the addition of logical constraints (24), (25).

Numeric data of input parameters for examples P1, P2, P3, P4, P5 are shown in Appendix A1. The results in the form of the objective function and the computation time are shown in Table II. Other results including the decision variables for the optimal value of the objective function are given in Appendix A2.

TABLE II
THE RESULTS OF NUMERICAL EXAMPLES FOR BOTH APPROACHES

P(No)	Hybrid				Integer Programming	
	MILPT1		MILPT2		MILP	
	Fc	T	Fc	T	Fc	T
P1(10)	22401*	600**	22394^{O}	18	22404*	600**
P2(10)	21167*	600**	21142^{O}	150	21343*	600**
P3(20)	45654^{O}	95	45654^{O}	8	45710*	600**

P(No)	MH	
	Fc	T
P4(10)	22397^{O}	255
P5(20)	46419^{O}	43

Fc	the value of the objective function
T	time of finding solution (in seconds)
^{O}	the optimal value of the objective function after the time T
*****	the feasible value of the objective function after the time T
****	calculation was stopped after 600 s
MILP	MILP model implementation in the IP environment.
MILPT1	MILP model after transformation - implementation in the hybrid optimization framework without phase P3
MILPT2	MILP model after transformation-implementation in the hybrid optimization framework.
MH	Hybrid model after transformation-implementation in the hybrid optimization framework.

For each example the solution for the MILPT2 implementation was found faster than that for the MILP implementation. Moreover, for examples P1 .. P3, the traditional approach based on integer programming gives only feasible solution (calculation was stopped after 600 s) despite using highly efficient LINGO solvers. It is obvious that the solution of the hybrid model (MH) was, due to its nature, only possible using the hybrid environment. Also, the proposed environment brought the expected results. The results were obtained in only a slightly longer period of time than that necessary for examples P1 and P3.

VII. CONCLUSION

The efficiency of the proposed approach is based on the reduction of the combinatorial problem. This means that using the hybrid approach practically for all models of this class, the same or better solutions are found faster (the optimal instead of the feasible solutions). Another element contributing to the high efficiency of the method is a possibility to determine the values or ranges of values for some of the decision variables (phase P3). All effective LINGO solvers can be used in the hybrid method.

Therefore, the proposed solution is highly recommended for all types of decision problems in supply chain or for other problems with similar structure. This structure is characterized by the constraints of many discrete decision variables and their summation. Furthermore, this method can model and solve problems with logical constraints.

Further work will focus on running the optimization models with non-linear and other logical constraints, multi-objective, uncertainty etc. in the hybrid optimization framework.

APPENDIX A1

TABLE III
DATA FOR COMPUTATIONAL EXAMPLES P1, P2, P3,P4,P5

k	V_k
P1	1
P2	1
P3	3
P4	2
P5	3
P6	1
P7	1
P8	3
P9	2
P10	3

j
1
2
3
4
5

i
F1
F2

s	V_s			F_s
	P_1,P_4	P_2,	P_3,P_5	
C1	200	300	800	600
C2	200	300	1000	700
C3	200	400	1000	900

d	Pt_s	Zt_s			Od_s
		P_1,P_4	P_2	P_3,P_5	
S1	10	30	50	60	125
S2	20	20	35	35	180
S3	40	10	20	20	240

i	s	d	$K_{i,s,d}$	$T_{i,s,d}$
F1	C1	S2	2	3
F1	C1	S3	4	4
F1	C2	S2	4	2
F1	C2	S3	8	3
F1	C3	S2	6	2
F1	C3	S3	8	3
F2	C1	S2	5	4
F2	C1	S3	7	4
F2	C2	S2	2	6
F2	C2	S3	4	7
F2	C3	S2	2	6
F2	C3	S3	3	6

i	k	$W_{i,k}$	$C_{i,k}$
F1	P1	350	10
F1	P2	300	40
F1	P3	500	30
F1	P4	600	40
F1	P5	400	50
F1	P6	300	60
F2	P5	400	50
F2	P6	300	60
F2	P7	400	70
F2	P8	500	80
F2	P9	600	90
F2	P10	650	90

s	j	d	$K_{s,j,d}$	$T_{s,j,d}$	s	j	d	$K_{s,j,d}$	$T_{s,j,d}$
C1	M1	S1	2	1	C2	M3	S2	6	2
C1	M1	S2	4	2	C2	M4	S1	3	1
C1	M2	S1	2	1	C2	M4	S2	6	2
C1	M2	S2	4	2	C2	M5	S1	3	1
C1	M3	S1	2	1	C2	M5	S2	6	2
C1	M3	S2	4	2	C3	M1	S1	4	1
C1	M4	S1	2	1	C3	M1	S2	8	2
C1	M4	S2	4	2	C3	M2	S1	4	1
C1	M5	S1	2	1	C3	M2	S2	8	2
C1	M5	S2	4	2	C3	M3	S1	4	1
C2	M1	S1	3	1	C3	M3	S2	8	2
C2	M1	S2	6	2	C3	M4	S1	4	1
C2	M2	S1	3	1	C3	M4	S2	8	2
C2	M2	S2	6	2	C3	M5	S1	4	1
C2	M3	S1	3	1	C3	M5	S2	8	2

k	i	k
P5	F1	P6
P5	F2	P6
P2	F1	P8
P2	F2	P8

k	s	k
P5	C1	P6
P5	C2	P6
P5	C3	P6
P2	C1	P8
P2	C2	P8
P2	C3	P8

s	k	$T_{s,k}$	s	k	$T_{s,k}$	s	k	$T_{s,k}$
C1	P1	2	C2	P1	1	C3	P1	3
C1	P2	2	C2	P2	1	C3	P2	3
C1	P3	2	C2	P3	1	C3	P3	3
C1	P4	2	C2	P4	1	C3	P4	3
C1	P5	2	C2	P5	1	C3	P5	3
C1	P6	2	C2	P6	1	C3	P6	3
C1	P7	2	C2	P7	1	C3	P7	3
C1	P8	2	C2	P8	1	C3	P8	3
C1	P9	2	C2	P9	1	C3	P9	3
C1	P10	2	C2	P10	1	C3	P10	3

Name	k	j	$T_{j,k}$	$Z_{j,k}$	Name	k	j	$T_{j,k}$	$Z_{j,k}$
Z_01	p1	m1	8	10	Z_11	p1	m3	8	15
Z_02	p2	m2	12	10	Z_12	p2	m4	12	20
Z_03	p3	m3	10	25	Z_13	p3	m5	10	25
Z_04	p4	m4	8	30	Z_14	p4	m1	8	30
Z_05	p5	m5	12	10	Z_15	p5	m2	12	30
Z_06	p6	m1	8	15	Z_16	p6	m3	8	15
Z_07	p7	m2	12	20	Z_17	p7	m4	12	20
Z_08	p8	m3	10	25	Z_18	p8	m5	10	25
Z_09	p9	m4	8	30	Z_19	p9	m1	8	30
Z_10	p10	m5	12	30	Z_20	p10	m2	12	35

APPENDIX A2

TABLE IV
RESULTS OF OPTIMIZATION FOR COMPUTATIONAL EXAMPLES P1, P2 (FULL) AND P3,P4,P5 (PART)

Example P1 Fc^{opt} = 22394

Name	i	k	s	j	d1	d2	X_{iskd1}	Y_{sjkd2}
Z_01	F1	P1	C1	M1	S3	S1	10.00	5.00
Z_01	F1	P1	C1	M1	S3	S2		5.00
Z_02	F1	P2	C1	M2	S3	S2	10.00	10.00
Z_03	F1	P3	C1	M3	S3	S1	25.00	5.00
Z_03	F1	P3	C1	M3	S3	S2		20.00
Z_04	F1	P4	C1	M4	S3	S2	30.00	30.00
Z_05	F2	P5	C2	M5	S3	S2	10.00	10.00
Z_06	F2	P6	C1	M1	S3	S2	15.00	15.00
Z_07	F2	P7	C1	M2	S3	S2	10.00	10.00
Z_07	F2	P7	C3	M2	S3	S1	10.00	10.00
Z_08	F2	P8	C1	M3	S3	S1	5.00	5.00
Z_08	F2	P8	C2	M3	S3	S2	20.00	20.00
Z_09	F2	P9	C2	M4	S2	S1	30.00	30.00
Z_10	F2	P10	C2	M5	S3	S2	10.00	10.00
Z_10	F2	P10	C3	M5	S3	S2	20.00	20.00

i	s	d	$Xb_{i,s,d}$	i	s	d	$Xb_{i,s,d}$
F1	C1	S3	4.00	F2	C2	S3	3.00
F2	C1	S3	1.00	F2	C3	S3	2.00
F2	C2	S2	3.00				

s	j	d	$Yb_{s,j,d}$	s	j	d	$Yb_{s,j,d}$
C1	M1	S1	1.00	C2	M3	S2	3.00
C1	M1	S2	1.00	C2	M4	S1	6.00
C1	M2	S2	1.00	C2	M5	S2	3.00
C1	M3	S1	3.00	C3	M2	S1	1.00
C1	M3	S2	3.00	C3	M5	S2	3.00
C1	M4	S2	3.00				

Example P2 Fc^{opt} = 2142

Name	i	k	s	j	d1	d2	X_{iskd1}	Y_{sjkd2}
Z_01	F1	P1	C1	M1	S2	S1	8.00	8.00
Z_01	F1	P1	C1	M1	S3	S1	2.00	1.00
Z_01	F1	P1	C1	M1	S3	S2		1.00
Z_02	F1	P2	C1	M2	S3	S2	10.00	10.00
Z_03	F1	P3	C1	M3	S3	S2	25.00	25.00
Z_04	F1	P4	C1	M4	S3	S2	30.00	30.00
Z_05	F1	P5	C1	M5	S2	S2	4.00	8.00
Z_05	F1	P5	C1	M5	S3	S2	4.00	
Z_06	F1	P6	C1	M1	S3	S1	1.00	1.00
Z_05	F2	P5	C1	M5	S3	S2	2.00	2.00
Z_06	F2	P6	C1	M1	S3	S2	14.00	14.00
Z_07	F2	P7	C1	M2	S3	S2	10.00	10.00
Z_07	F2	P7	C2	M2	S3	S1	10.00	10.00
Z_08	F2	P8	C2	M3	S3	S2	25.00	25.00
Z_09	F2	P9	C1	M4	S3	S2	30.00	30.00
Z_10	F2	P10	C1	M5	S3	S2	10.00	10.00
Z_10	F2	P10	C2	M5	S3	S2	20.00	20.00

i	s	d	$Xb_{i,s,d}$	i	s	d	$Xb_{i,s,d}$
F1	C1	S2	1.00	F2	C1	S3	3.00
F1	C1	S3	4.00	F2	C2	S3	4.00

s	j	d	$Yb_{s,j,d}$	s	j	d	$Yb_{s,j,d}$
C1	M1	S1	1.00	C1	M5	S2	3.00
C1	M1	S2	1.00	C2	M2	S1	1.00
C1	M2	S2	1.00	C2	M3	S2	4.00
C1	M3	S2	4.00	C2	M5	S2	3.00
C1	M4	S2	6.00				

Example P3 Fc^{opt} = 45654

Name	i	k	s	j	d1	d2	X_{iskd1}	Y_{sjkd2}
Z_01	F1	P1	C2	M1	S3	S1	25.00	10.00
Z_11	F1	P1	C2	M3	S3	S2		15.00
.....								
Z_20	F2	P10	C2	M2	S2	S2	1.00	35.00
Z_20	F2	P10	C2	M2	S3	S2	64.00	
Z_10	F2	P10	C2	M5	S3	S2		30.00

i	s	d	$Xb_{i,s,d}$	i	s	d	$Xb_{i,s,d}$
F1	C2	S2	1.00	F2	C2	S2	8.00
F1	C2	S3	8.00	F2	C2	S3	12.00

s	j	d	$Yb_{s,j,d}$	s	j	d	$Yb_{s,j,d}$
C2	M1	S1	15.00	C2	M4	S1	14.00
C2	M2	S1	11.00	C2	M4	S2	1.00
C2	M2	S2	6.00	C2	M5	S1	9.00
C2	M3	S1	8.00	C2	M5	S2	9.00
C2	M3	S2	5.00				

Example P4 Fc^{opt} = 22397

Name	i	k	s	j	d1	d2	X_{iskd1}	Y_{sjkd2}
Z_01	F1	P1	C1	M1	S3	S1	10.00	10.00
Z_02	F1	P2	C1	M2	S3	S1	10.00	10.00
Z_03	F1	P3	C1	M3	S3	S2	20.00	20.00
Z_03	F1	P3	C2	M3	S2	S1	5.00	4.00
Z_03	F1	P3	C2	M3	S2	S2		1.00
....								
Z_10	F2	P10	C2	M5	S3	S2	10.00	10.00
Z_10	F2	P10	C3	M5	S3	S2	20.00	20.00

i	s	d	$Xb_{i,s,d}$	i	s	d	$Xb_{i,s,d}$
F1	C1	S3	4.00	F2	C2	S2	2.00
F1	C2	S2	1.00	F2	C2	S3	3.00
F2	C1	S3	1.00	F2	C3	S3	2.00

s	j	d	$Yb_{s,j,d}$	s	j	d	$Yb_{s,j,d}$
C1	M1	S1	3.00	C2	M3	S2	3.00
C1	M2	S1	1.00	C2	M4	S1	2.00
C1	M3	S2	3.00	C2	M5	S2	3.00
C1	M4	S1	2.00	C3	M2	S2	1.00
C1	M4	S2	4.00	C3	M5	S2	3.00
C2	M3	S1	3.00				

Example P5 Fc^{opt} = 46419

Name	i	k	s	j	d1	d2	X_{iskd1}	Y_{sjkd2}
Z_01	F1	P1	C1	M1	S2	S1	10.00	10.00
Z_11	F1	P1	C1	M3	S3	S1	15.00	15.00
Z_01	F1	P2	C1	M2	S3	S2	30.00	10.00
Z_12	F1	P2	C1	M4	S3	S2		20.00
....								
Z_20	F2	P10	C1	M2	S3	S2	56.00	31.00

Z_10	F2	P10	C1	M5	S3	S1		25.00
Z_20	F2	P10	C2	M2	S2	S1	4.00	4.00
Z_10	F2	P10	C2	M5	S3	S1	5.00	5.00

i	s	d	$Xb_{i,s,d}$	i	s	d	$Xb_{i,s,d}$
F1	C1	S2	7.00	F2	C2	S2	3.00
F1	C1	S3	8.00	F2	C2	S3	4.00
F2	C1	S3	8.00				

s	j	d	$Yb_{s,j,d}$	s	j	d	$Yb_{s,j,d}$
C1	M1	S1	13.00	C1	M5	S1	12.00
C1	M2	S1	1.00	C1	M5	S2	3.00
C1	M2	S2	10.00	C2	M1	S1	2.00
C1	M3	S1	9.00	C2	M2	S1	2.00
C1	M4	S1	2.00	C2	M3	S1	9.00
C1	M4	S2	7.00	C2	M5	S1	9.00

APPENDIX A3

TABLE V
ANALYSIS OF THE IMPACT PARAMETER V_S FOR FC (EXAMPLE P2)

$V=V_1=V_2=V_3$	Fc^{opt}	Distributor capacity (Vs)utilization		
		V_1	V_2	V_3
200	22 058	199	176	70
300	21 142	300	145	0
400	21 137	435	10	0
450	20 439	445	0	0
500	20 439	445	0	0
550	20 439	445	0	0

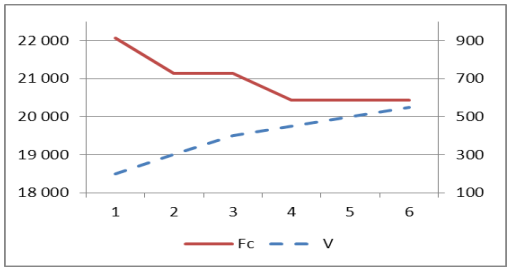


Fig. 4 The value of the objective function depending on the parameter V (Example P2)

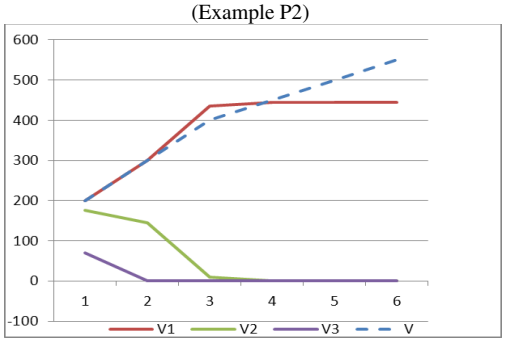


Fig. 5 Capacity utilization (Vs) for distributor s=1, s=2,s=3 (Example P2)

References

[1] Huang, G.Q., Lau, J.S.K., Mak, K.L., The impacts of sharing production information on supply chain dynamics: a review of the literature. International Journal of Production Research 41, 2003, pp.1483–1517.

[2] Kanyalkar, A.P., Adil, G.K., An integrated aggregate and detailed planning in a multi-site production environment using linear programming. International Journal of Production Research 43, 2005, pp. 4431–4454.

[3] Perea-lopez, E., Ydstie, B.E., Grossmann, I.E., A model predictive control strategy for supply chain optimization. Computers and Chemical Engineering 27, 2003, pp. 1201–1218.

[4] Park, Y.B.,. An integrated approach for production and distribution planning in supply chain management. International Journal of Production Research 43, 2005, pp. 1205–1224.

[5] Jung, H., Jeong, B., Lee, C.G., An order quantity negotiation model for distributor-driven supply chains. International Journal of Production Economics 111, 2008, pp. 147–158.

[6] Rizk, N., Martel, A., D'amours, S., Multi-item dynamic production–distribution planning in process industries with divergent finishing stages. Computers and Operations Research 33, 2006, pp. 3600–3623.

[7] Selim, H., Am, C., Ozkarahan, I., Collaborative production–distribution planning in supply chain: a fuzzy goal programming approach. Transportation Research Part E-Logistics and Transportation Review 44, 2008, pp. 396–419.

[8] Lee, Y.H., Kim, S.H., Optimal production–distribution planning in supply chain management using a hybrid simulation-analytic approach. Proceedings of the 2000 Winter Simulation Conference 1 and 2, 2000, pp. 1252–1259.

[9] Chern, C.C., Hsieh, J.S., A heuristic algorithm for master planning that satisfies multiple objectives. Computers and Operations Research 34, 2007, pp. 3491–3513.

[10] Jang, Y.J., Jang, S.Y., Chang, B.M., Park, J., A combined model of network design and production/distribution planning for a supply network. Computers and Industrial Engineering 43, 2002, pp. 263–281.

[11] Timpe, C.H., Kallrath, J., Optimal planning in large multi-site production networks. European Journal of Operational Research 126, 2000, pp. 422–435.

[12] Mula J., Peidro D., Diaz-Madronero M., Vicens E., Mathematical programming models for supply chain production and transport planning, European Journal of Operational Research 204, 2010, pp. 377–390.

[13] Apt K., Wallace M., Constraint Logic Programming using Eclipse, Cambridge University Press, 2006.

[14] Sitek P., Wikarek J., A Declarative Framework for Constrained Search Problems. New Frontiers in Applied Artificial Intelligence, Lecture Notes in Artificial Intelligence, Nguyen, NT., et al. (Eds.), Vol. 5027, Springer-Verlag, Berlin-Heidelberg, 2008, pp. 728-737.

[15] Bocewicz G., Wójcik R., Banaszak Z., AGVs distributed control subject to imprecise operation time". In: Agent and Multi-Agent Systems: Technologies and Applications, Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, Vol. 4953, 2008, pp. 421-430.

[16] Rossi F., Van Beek P., Walsh T., Handbook of Constraint Programming (Foundations of Artificial Intelligence), Elsevier Science Inc. New York, NY, USA, 2006.

[17] Sitek P., Wikarek J., Cost optimization of supply chain with multimodal transport, Federated Conference on Computer Science and Information Systems (FedCSIS), 2012, pp. 1111-1118.

[18] Sitek P., Wikarek J., Supply chain optimization based on a MILP model from the perspective of a logistics provider, Management and Production Engineering Review, 2012 pp. 49-61.

[19] Sitek P., Wikarek J., The Declarative Framework Approach to Decision Support for Constrained Search Problems, INTECH, 2011, pp. 163-182.

[20] Jain V., Grossmann I.E., Algorithms for hybrid MILP/CP models for a class of optimization problems, INFORMS Journal on Computing 13(4), 2001, pp. 258–276.

[21] Milano M., Wallace M., Integrating Operations Research in Constraint Programming, Annals of Operations Research vol. 175 issue 1, 2010, pp. 37 – 76.

[22] Achterberg T., Berthold T., Koch T., Wolter K., Constraint Integer Programming. A New Approach to Integrate CP and MIP, Lecture Notes in Computer Science Volume 5015, 2008, pp. 6-20.

[23] Bockmayr A., Kasper T., Branch-and-Infer, A Framework for Combining CP and IP, Constraint and Integer Programming Operations Research/Computer Science Interfaces Series, Volume 27, 2004, pp. 59-87.

[24] Sitek P., Wikarek J., A hybrid method for modeling and solving constrained search problems, Federated Conference on Computer Science and Information Systems (FedCSIS), 2013.pp.385-392.

[25] www.eclipse.org

[26] www.lindo.com

19th Conference on Knowledge Acquisition and Management

We have the pleasure to invite you to contribute to and participate in the conference "Knowledge Acquisition and Management" (KAM'2012). The predecessor of the KAM conference has been organized for the first time in 1992, as a venue for scientists and practitioners to address different aspects of usage of advanced information technologies in management, with focus on intelligent techniques and knowledge management. In 2003 the conference changed somewhat its focus and was organized for the first under its current name. Furthermore, the KAM conference became an international event, with participants from around the world. The aim of the conference is to create possibility of presenting and discussing approaches, methods, techniques and tools in the knowledge acquisition and the knowledge management areas. We expect that the conference will enable exchange of information and experiences, and delve into current trends of methodological, technological and implementation aspects of knowledge acquisition and management processes.

Topics

The following group topics, concerning both theory and applications, will be included (unavoidably incomplete):

- Data mining and knowledge discovery from databases and data warehouses
- Methods and tools for knowledge acquisition
- New emerging technologies for management
- Organizing the knowledge centers
- Knowledge creation and validation
- Knowledge dynamics and machine learning
- Distance learning and knowledge sharing
- Knowledge representation models
- Management of enterprise knowledge versus personal knowledge
- Knowledge managers and workers
- Knowledge coaching
- Knowledge diffusion
- Knowledge engineering and software engineering
- Managerial knowledge evolution
- Knowledge grid and social networks
- Knowledge modeling and visualization
- Knowledge management and e-government
- Business Intelligence environment for supporting knowledge management
- Knowledge management in virtual agents

Event Chairs

Hauke, Krzysztof, Wroclaw University of Economics, Poland

Nycz, Małgorzata, Wrocław University of Economics, Poland

Owoc, Mieczyslaw, Wroclaw University of Economics, Poland

Pondel, Maciej, Wroclaw University of Economics, Poland

Program Committee

Abramowicz, Witold, Poznan University of Economics, Poland

Andres, Frederic, National Institute of Informatics, Tokyo, Japan

Bodyanskiy, Yevgeniy, Kharkiv National University of Radio Electronics, Ukraine

Chmielarz, Witold, Warsaw University, Poland

Christozov, Dimitar, American University in Bulgaria, Bulgaria

Goluchowski, Jerzy, University of Economics in Katowice, Poland

Helfert, Markus, Dublin City University, Ireland

Korczak, Jerzy, Wroclaw University of Economics, Poland

Mach-Król, Maria, Katowice University of Economics, Poland

Perechuda, Kazimierz, Wroclaw University of Economics, Poland

Vanhoof, Koen, Hasselt University, Belgium

Vanthienen, Jan, KU Leuven, Belgium

Zhelezko, Boris, Belarussian Economic State University, Minsk, Belarus

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1221–1226

Inconsistency Handling in Collaborative Knowledge Management

Weronika T. Adrian
AGH University of
Science and Technology
wta@agh.edu.pl

Antoni Ligęza
AGH University of
Science and Technology
ligeza@agh.edu.pl

Grzegorz J. Nalepa
AGH University of
Science and Technology
gjn@agh.edu.pl

***Abstract*—One of the challenges of knowledge management is handling inconsistency. Traditionally, it was often perceived as indication of invalid data or behavior and as such should be avoided or eliminated. However, there are also numerous situations where inconsistency is a natural phenomena or carry useful information. In order to decide how to manage inconsistent knowledge, it is thus important to recognize its origin, aspect and influence on the behavior of the system. In this paper, we analyze a case of collaborative knowledge management with hybrid knowledge representation. This serves as a starting point for a discussion about various types of inconsistencies and approaches to handle it. We analyze sources, interpretation and possible approaches to identified types of inconsistencies. We discuss practical use cases to illustrate selected approaches.**

I. Introduction

INCONSISTENCY, defined in various sources as inability for all conceived statements or beliefs to be simultaneously true, is a major issue in knowledge management (KM). Contradiction, often used as a synonym, is an especially strong kind of inconsistency between sentences such that one sentence must be true and the other must be false [1].

Inconsistency management has found application in various areas including: knowlege-based systems analysis [2], [3] and verification [4], multiagent systems, information retrieval, recommender systems, and intelligent tutoring systems [5].

While inconsistency of data is usually undesirable, inconsistency of knowledge constitutes a more complex challenge [6]. Firstly, it is not always easy to discover, because it may appear on different levels of knowledge representation and reasoning. Secondly, in distributed and dynamic environments, it may be a natural phenomena that carry useful information.

There exist various approaches to handle inconsistency [7], [8], from elimination, through consensus methods [5], [9] and argumentation frameworks [10], [11], up to paraconsistent reasoning tolerating inconsistent information [12]. Sometimes it is useful to first measure the inconsistency [13], [14] and based on the results decide what to do with it [15], [16].

In our research, we analyze inconsistency in a collaborative knowledge management environment (see Fig. 1). With the rapid development of new technologies, collaborative environments for knowledge management become increasingly complex. Knowledge in such environments is represented with use of diversified formal, semi-formal and informal methods [18].

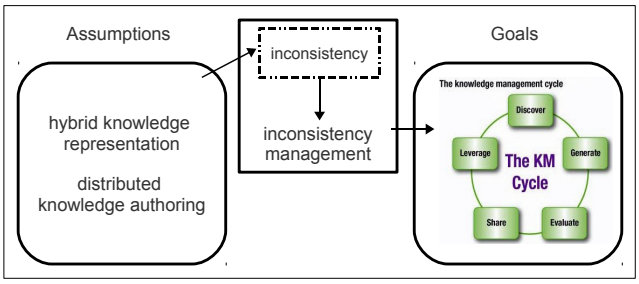


Figure 1. Inconsistency in Collaborative Knowledge Management (the knowledge management cycle as in [17]).

In this work, we concentrate on knowledge representation with Semantic Web technologies. Within this area, we investigate:

- Challenges and problems related to inconsistency,
- Sources of inconsistency, and
- Approaches to handle inconsistency.

This paper is organized as follows: Motivation for our research is given in Section II. Theoretical background of handling inconsistency is outlined in Section III. Selected problems and approaches to handle inconsistency on the Semantic Web are then discussed in Section IV. In Section V, an analysis of the approaches with respect to knowledge management is given. Use case scenarios are presented in Section VI. The paper is concluded in Section VII.

II. Motivation

Motivation for research presented in this work stems from the experiences of BIMLOQ[1] and INDECT [19], [20] research projects. The former was focused on quality of knowledge represented with business processes, rules and semantics. Within the latter a collaborative knowledge management environment was developed that used semantic technologies and social features to foster collaboration. The projects revealed the importance of practical challenges of inconsistency in knowledge management, related specifically to:

1) Dynamics of the system (integration, revision, merge),
2) Distributed knowledge authoring, and
3) Different methods of knowledge representation.

[1]See http://bimloq.ia.agh.edu.pl.

978-83-60810-53-8/$25.00

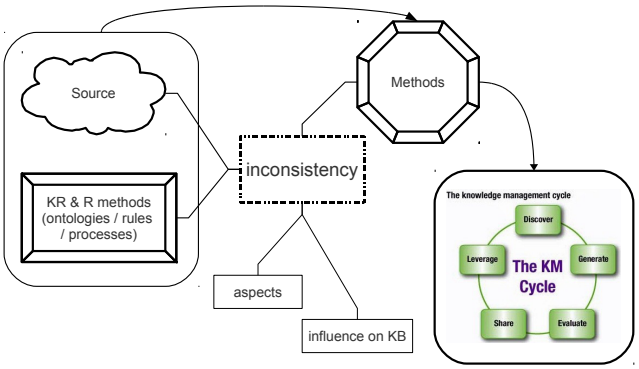


Figure 2. Map of areas for inconsistency analysis.

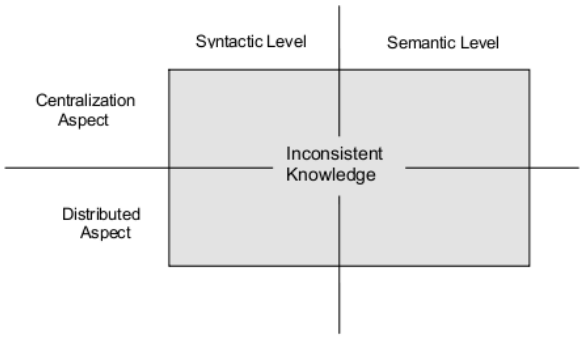


Figure 3. Aspects and levels of inconsistency [29].

Therefore, when considering a collaborative knowledge management environment with hybrid knowledge representation we aim to analyze (see Fig. 2):

1) How knowledge representation influence inconsistency?
2) What are the sources of inconsistency in selected areas?
3) What are possible approaches to handle it?

Our basic testing platform is a semantic knowledge-based wiki that supports semantic technologies and rule-based reasoning [21]. Semantic wikis proved to be useful in collaborative knowledge management and engineering [22]. They constitute a flexible tool for knowledge representation and reasoning [23], as well as distributed knowledge acquisition [24]. Our implementation supports semantic technologies and rule-based reasoning. Combining different method of knowledge representation within a semantic wiki have been proposed by authors in several works: business processes with rules [25], and rules with semantics [26], [27]. There are ongoing works on integrating semantics with business processes [28].

Currently, the system (available at http://loki.ia.agh.edu.pl) is not equipped with any mechanisms for handling inconsistencies. Ultimately, we aim to develop a system that will allow for knowledge representation with semantic technologies, rules and business processes, able to deal with inconsistency.

We claim that in such an environment handling inconsistency is a complex challenge and one should consider methods that *accept* it rather than eliminate. In the following section, we briefly review selected concepts, approaches and theoretical bases for the problem of inconsistency. This is the starting point for more detailed and focused review given in Section IV. In this paper, we analyze the case of the Semantic Web technologies and present selected approaches to deal with inconsistent knowledge.

III. Theoretical Foundations

A. Vocabulary related to Inconsistency

When talking about *inconsistency*, one can find several definitions and interpretations of the core terms used in the area. Here we assume the following definitions:

- Inconsistency – when set of sentences or beliefs cannot be true at the same time or under the same interpretation
- Contradiction – when having two sentences or beliefs if one is true, then the other cannot be true,
- Paraconsistency – way of amending classical logic to be able to reason (conclude meaningful statements) in presence of inconsistency (ECQ does not hold).

B. Aspects and Levels of Inconsistency

One can consider different aspects and levels of inconsistency (see Fig. 3). In *distribution aspect*, basic cause of inconsistency is the independence of knowledge agents or knowledge processing mechanisms, while in *centralization aspect* inconsistency is related to dynamic changes of a world.

Inconsistency can be identified and processed on a *syntactic* and *semantic* level [29]. Analogously, inconsistency can be checked for in a purely *logical* way (e.g. p and $\neg p$ are present in the knowledge under discourse), or as *material* inconsistency, when two pieces of knowledge are invalid together due to the assumed interpretation [30].

Taking into consideration possible actions in the presence of inconsistency, one may take a *actual-contradictions* view or *potential-contradictions* view [7]. The former assumes that contradictions can appear in a knowledge base and no ”degenerate” reasoning should occur when contradictory statements are jointly asserted. Every statement should be treated equally. However there are two main approaches: First is that contradictions are ”bad”: if they appear, then reasoning collapses and results are trivial. The other is that the contradictions arise naturally and can be more informative than any consistent revision of the theory. Potential-contradictions view claims that contradictions do not actually exist, so there are either some statements ”responsible” for the inconsistency or the contradictions can be resolved by using argumentation. Concrete realization of potential condradictions view is in defeasible reasoning.

C. Formal Representation of Inconsistency

In order to formalize inconsistency handling in logic, there must be a formal representation of inconsistency itself. In [7], where logics are defined as ”formal systems consisting of a language L (in the form of a set of formulas) on which

an inference operation C is defined." three approaches to represent inconsistency are distinguished:

- *C-scheme*: to relate contradictions to inference, stating that inconsistency arises when all formulas are inferred,
- *A-scheme*: to pick a subset of the language, and use each element of the subset as a representation of absurdity, and
- *N-scheme*: contradictions are captured through an auxiliary notion of negation ($A, \neg A$ or $A \wedge \neg A$ if conjunction is available, is a syntactical account for inconsistency).

Inconsistency may be also represented in a form of *conflict profiles* as explained in [5]. Finally, contradictions can be *incorporated* into the formal logic by augmenting the classical logic. One of the most successful is Belnap's 4-valued logic [31] in which we can represent a statement that can be inferred to be true and false at the same time. This logic has been successfully used in the context of DL ontologies as will be shown in Section IV-B.

IV. Inconsistency on the Semantic Web

Semantic technologies are used to represent and process data, information and knowledge on several abstraction levels. Relations between objects are described with RDF [32] and simple classification can be done in RDFS [33]. Ontologies constitute the main method of knowledge representation. Integration with rules as well as incorporation of them within ontologies is an active research area. Inconsistency can be therefore considered on various levels as presented in the following subsections.

A. Inconsistency in RDF/S

RDF allows to describe objects by means of statements about their attributes and relations to other objects. The statements build a graph representing positive knowledge about the conceived world. In order to ascribe a category and build simple taxonomies, RDF Schema was introduced. It supports basic relation that define the *type* of an object, *domain* and *range* of certain relations etc. Reasoning in RDF/S is based on their defined semantics and set of entailment rules.

In order to avoid inconsistency, it was not allowed to use explicit negation in RDF/S. Open World Assumption that traditionally holds on the Semantic Web means that if something is not stated, it does not mean that it is not true. However, despite this restriction, there still exist inconsistent RDFS statements, e.g., ones that violate domain/range restrictions.

To deal with inconsistency, Extended RDF (ERDF) [34] has been proposed. Its semantics is based on *partial logic* that allows to express both *strong* and *weak* negation and can support reasoning with Closed-World Assumption as well as Open-World Assumption (this can be set by author of a semantic knowledge base). Stable model semantics of Extended RDF (ERDF) ontologies has also been proposed.

The approach is realized in MWeb framework[2]. The tool uses *restricted propagation of local inconsistencies*, making it possible to reasoning even in the presence of an inconsistency, local to a Web rule base and reasoning mode.

[2]See http://centria.di.fct.unl.pt/~cd/mweb/.

B. Inconsistent Ontologies

Formal ontologies, understood as logical theories, should typically be consistent. However, there are different situations where inconsistency may appear.

1) Problems and Sources of Inconsistency: Several problems related to inconsistent ontologies are distinguished [35]: ontology mismatch and conflict, ontology merging, and integration. Two former are related to specific relations between two or more ontologies. Two latter refer to some activities performed on ontologies. Main scenarios for a formation of inconsistency are given in [36]:

- *Multiple sources*, for example if the ontology is built by several authors, or during such processes as merging, integrating and aligning ontologies.
- *Mis-representation of defaults*, for instance if a more general concept is inconsistent with more specific facts.
- *Moving from other formalism*, for example if in the target formalism there are restrictions that make the translated information contradictory.
- *Polysemy*, if the same name refers to different concepts with inconsistent definitions.

2) Inconsistency Levels: Differences between ontologies may appear on various levels. While instances can be identified on the physical level (referring to their being in the real world), concepts are identified only on the logical one, that is referring to their names and structures. Nguyen [35] distinguishes the following levels of inconsistency between ontologies (called *ontology conflicts*):

- *Inconsistency on the instance level*: the same instance belonging to different ontologies does not satisfy the instance integration condition which states that if the instance is described differently in different concepts then in referring to the same attributes they should have the same value.
- *Inconsistency on the concept level*: There are several concepts with the same name having different structures in different ontologies.
- *Inconsistency on the relation level*: Between the same two concepts there are inconsistent relations in different ontologies.

3) Selected Approaches to Handle Inconsistency: Several approaches to inconsistency have been adapted for ontologies. On the one hand, inconsistency in ontologies can be diagnosed and repaired before reasoning [37] (including forgetting-based approach [38]). On the other, one can use conflicts to generate new knowledge or perform meaningful reasoning over inconsistent ontologies. Sometimes, it is not practical or even possible to resolve inconsistency (due to the access restrictions or possible information loss). The following examples illustrate selected approaches that seems especially suitable for collaborative knowledge management:

a) Consensus-based methods: Consensus-based methods have been discussed in [35] as a way to resolve inconsistency during ontology integration. Algorithms to determine a consensus on the instance, concept and relation level have

been proposed. For a set of different versions of data (so called *conflict profile*) they determine such a version that best represents the given versions.

b) Argumentative frameworks: Argumentative framework for reasoning with inconsistent DL ontologies has been proposed in [10], [11]. The proposal involves expressing DL ontologies as Defeasible Logic Programs (DeLP). Once a query is posed to an inconsistent ontology, a dialectical analysis on a DeLP program (obtained from such ontology) is performed and all arguments in favor and against the final answer of the query are taken into account [10].

c) Selecting consistent subsets: This approach, introduced in [36] and extended in [39] is based on selecting consistent sub-theories from inconsistent ontologies using selection functions based on syntactic [36] or semantic [39] relevance. Reasoning is then executed on the consistent subset and if a satisfying answer cannot be found, the sub-theory is appropriately extended. In [40], minimal inconsistent sets (MIS) and a resolution method are proposed to improve the run-time performance of the inconsistency reasoner. The approaches have been implemented in PION (Processing Inconsistency ONtologies) tool within the LarKC project.[3]

d) Paraconsistent reasoning: One can also extend the classical logic for OWL (Description Logics are subsets of First Order Logic) to many-valued paraconsistent logic. A proposal of representing inconsistent ontologies with 4-valued logic has been proposed in [41], [42]. In this approach, two additional truth values, namely *underdefined* and *overdefined* (i.e. contradictory) are used. Thanks to the mapping between the logics, it is possible to use classical OWL reasoners to operate on inconsistent knowledge. This idea has been implemented in RaDON NeoN Toolkit Plugin.

V. Analysis and Discussion

Traditionally, in analysis and verification of knowledge-based systems, inconsistency was undesirable and negatively influenced the quality of a knowledge. With the advent of modern Web-based environments, collaborative knowledge management becomes vital. Inconsistency in such environments seems unavoidable. It can be considered in distribution and centralization aspect, i.e., related to the multiple sources or dynamic changes over time. Inconsistency may be discovered during a process of knowledge integration, e.g. ontology aligning or merging, or observed in a static knowledge base, e.g., a fact base may contain inconsistent statements.

Even within the narrowed field of knowledge representation with Semantic Web technologies, there exist various approaches to handle inconsistency that partially depend on the representation level. In case of RDF/S, the knowledge base consists of positive statements, assertions with limited semantics. Inconsistency here is closely linked to the issue of negation. ERDF aims to solve this problem by introducing strong and weak negation into the language and allowing the knowledge engineer to state whether Closed- or Open-World Assumption should be adopted. OWL Ontologies are logical theories based on formal logic. If classical semantics is adopted, then inconsistency make them unusable and thus should be suppressed, e.g., by consensus finding or repairing methods. However, if the semantics is re-defined, one can tolerate and represent inconsistent information e.g., by using one of paraconsistent logics.

Particular phases of knowledge management cycle [17], poses various challenges related to inconsistency, including:

- *Discover*: information fusion from multiple sources, independent experts, independent knowledge processing,
- *Generate*: various views or opinions expressed in mutually inconsistent concept descriptions, self-contradictions,
- *Evaluate*: measuring inconsistency, meaningful query answering and reasoning in inconsistent knowledge bases,
- *Share*: merging, aligning, integrating knowledge,
- *Leverage*: searching for a consensus or argumentation.

Circumstances in which inconsistency appears and the level of knowledge representation it affects may have a deciding influence on the method one will choose to manage the inconsistency. Selected approaches, presented in this paper and summarized in Table I, illustrate various aspects and scenarios that may be adapted in collaborative knowledge management.

VI. Use Case Examples

In this section, we present selected use cases of handling inconsistency in collaborative knowledge management. Let us consider a semantic wiki with underlying logical knowledge representation that supports semantic technologies, rule-based reasoning and business process modeling [21], [23], [28].

A. Collaborative Ontology Development

In this case, autonomous experts jointly model a single knowledge base (e.g. an ontology). During the process, the ontology becomes inconsistent.

1) Case: Inconsistency on relation/concept/instance level.

2) Source of the Inconsistency: Distributed authoring.

3) Interpretation of the Inconsistency: Authors may have different opinions or diversified knowledge about the subject.

4) Suggested Approach: If classical semantics is adopted, consistency should be regained. Consensus methods (taking into account different opinions) or argumentation framework (identifying the best option) could be used to resolve conflicts.

B. Collaborative Recommendation System

In this case, autonomous knowledge agents independently assert their opinions and ratings about movies in the system.

1) Case: Pieces of knowledge (facts) are inconsistent.

2) Source of the Inconsistency: Distributed authoring / Dynamic updates.

3) Interpretation of the Inconsistency: Authors may have different opinions and there is no way to arbitrarily say which one is good and which one should be eliminated. / Authors add, remove or change their opinions over time.

4) Suggested Approach: Inconsistency should be accepted (all opinions should be represented, regardless of contradictions). Paraconsistent reasoning, e.g. using four-valued logic could be used.

[3] See http://larkc.eu.

Table I
SELECTED APPROACHES TO INCONSISTENCY ON THE SEMANTIC WEB

Reference	Problem	Knowledge Representation	Approach	Tool
[10], [11]	ontology integration	DL	argumentative framework	–
[34]	reasoning with inconsistent information	RDF/S	stable model semantics, partial logic (strong and weak negation)	MWeb
[35]	ontology integration	not specified (solution on a general level)	determining consensus on instance/concept/relation level	–
[36], [39] [40]	querying inconsistent KB	DL	selecting consistent subsets	PION (plugin for LarKC tool)
[41], [42]	reasoning with inconsistent KB	$SROIQ$ DL	mapping to 4-valued logic	NeoN Toolkit Plugin RaDON

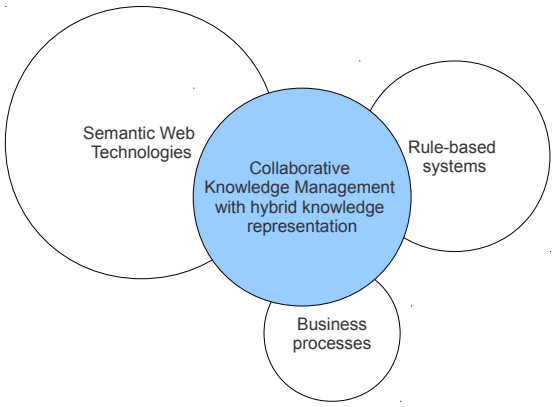


Figure 4. Inconsistency in Collaborative Knowledge Management.

VII. SUMMARY AND FUTURE WORK

Collaborative Knowledge Management poses numerous challenges related to inconsistency. It may result from the distributed authoring as well as dynamic changes of the world and the system. Inconsistency may be considered on a purely syntactic level or on a level where the semantics play a significant role. While in some situations inconsistency signals erratic data or behavior and should be resolved, sometimes it is natural or even useful. In order to apply appropriate technique to handle inconsistency, it is necessary to recognize and understand its origin and influence on the system. In this paper, we analyzed inconsistency in collaborative knowledge management. We presented selected problems of inconsistency and approaches to handle it.

In order to build a comprehensive collaborative environment with hybrid knowledge representation that is capable of handling various sorts of inconsistency, we consistently investigate each area of considered knowledge representation (see Fig. 4). In this paper, we have analyzed the area of Semantic Web taking into account various levels of knowledge representation and different situations in which inconsistency may arise.

In future, we plan to analyze inconsistency handling in rule-based systems [43], [44], business processes, and Multi-Context Systems [45]. Quality of knowledge will also be addressed, taking into consideration Information Quality criteria.

REFERENCES

[1] B. Dowden, "What is inconsistency?" http://www.csus.edu/indiv/d/dowdenb/misc/inconsistency.htm, 2013.

[2] F. Coenen, T. Bench-Capon, R. Boswell, J. Dibie-Barthélemy, B. Eaglestone, R. Gerrits, E. Grégoire, A. Ligęza, L. Laita, M. Owoc, F. Sellini, S. Spreeuwenberg, J. Vanthienen, A. Vermesan, and N. Wiratunga, "Validation and verification of knowledge-based systems: report on eurovav99," *Knowl. Eng. Rev.*, vol. 15, no. 2, pp. 187–196, Jun. 2000. [Online]. Available: http://dx.doi.org/10.1017/S0269888900002010

[3] G. Nalepa, S. Bobek, A. Ligęza, and K. Kaczor, "HalVA – rule analysis framework for XTT2 rules," in *Rule-Based Reasoning, Programming, and Applications*, ser. Lecture Notes in Computer Science, N. Bassiliades, G. Governatori, and A. Paschke, Eds., vol. 6826. Springer Berlin / Heidelberg, 2011, pp. 337–344. [Online]. Available: http://www.springerlink.com/content/c276374nh9682jm6/

[4] M. Szpyrka, G. J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 249–255. [Online]. Available: http://www.springerlink.com/content/m181144037q67271/

[5] N. T. Nguyen, *Advanced Methods for Inconsistent Knowledge Management (Advanced Information and Knowledge Processing)*. Springer London, 2008.

[6] A. Ligęza, "Intelligent data and knowledge analysis and verification; towards a taxonomy of specific problems," in *Validation and Verification of Knowledge Based Systems*, A. Vermesan and F. Coenen, Eds. Springer US, 1999, pp. 313–325. [Online]. Available: http://dx.doi.org/10.1007/978-1-4757-6916-6_21

[7] P. Besnard and A. Hunter, "Introduction to actual and potential contradictions," in *Reasoning with Actual and Potential Contradictions*, ser. Handbook of Defeasible Reasoning and Uncertainty Management Systems, P. Besnard and A. Hunter, Eds. Springer Netherlands, 1998, vol. 2, pp. 1–9. [Online]. Available: http://dx.doi.org/10.1007/978-94-017-1739-7_1

[8] L. Bertossi, A. Hunter, and T. Schaub, "Introduction to inconsistency tolerance," in *Inconsistency Tolerance*, ser. Lecture Notes in Computer Science, L. Bertossi, A. Hunter, and T. Schaub, Eds. Springer Berlin Heidelberg, 2005, vol. 3300, pp. 1–14. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30597-2_1

[9] A. Czyszczoń and A. Zgrzywa, "Consensus as a tool supporting customer behaviour prediction in social crm systems," *Computer Science*, vol. 13, no. 4, 2012. [Online]. Available: http://journals.agh.edu.pl/csci/article/view/49

[10] S. A. Gomez, C. I. Chesnevar, and G. R. Simari, "An argumentative approach to reasoning with inconsistent ontologies," in *Knowledge Representation Ontology Workshop (KROW 2008)*, ser. CRPIT, T. Meyer and M. A. Orgun, Eds., vol. 90. Sydney, Australia: ACS, 2008, pp. 11–20.

[11] S. Alejandro Gomez, C. Ivan Chesnevar, and G. R. Simari, "Reasoning with inconsistent ontologies through argumentation," *Appl. Artif. Intell.*, vol. 24, no. 1-2, pp. 102–148, Jan. 2010.

[12] G. Priest, K. Tanaka, and Z. Weber, "Paraconsistent logic," in *The Stanford Encyclopedia of Philosophy*, summer 2013 ed., E. N. Zalta, Ed., 2013.

[13] J. Grant and A. Hunter, "Measuring inconsistency in knowledgebases," *J. Intell. Inf. Syst.*, vol. 27, no. 2, pp. 159–184, Sep. 2006.

[14] Y. Ma, G. Qi, and P. Hitzler, "Computing inconsistency measure based on paraconsistent semantics," *Journal of Logic and Computation*, vol. 21, no. 6, pp. 1257–1281, 2011.

[15] A. Hunter, "How to act on inconsistent news: ignore, resolve, or reject," *Data Knowl. Eng.*, vol. 57, no. 3, pp. 221–239, Jun. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.datak.2005.04.005

[16] J. Grant and A. Hunter, "Measuring consistency gain and information loss in stepwise inconsistency resolution," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, ser. Lecture Notes in Computer Science, W. Liu, Ed. Springer Berlin Heidelberg, 2011, vol. 6717, pp. 362–373. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22152-1_31

[17] F. Bouthillier and K. Shearer, "Understanding knowledge management and information management: the need for an empirical perspective," *Information research*, vol. 8, no. 1, pp. 8–1, 2002.

[18] J. Baumeister, J. Reutelshoefer, and F. Puppe, "Engineering intelligent systems on the knowledge formalization continuum," *International Journal of Applied Mathematics and Computer Science (AMCS)*, vol. 21, no. 1, 2011. [Online]. Available: http://ki.informatik.uni-wuerzburg.de/papers/baumeister/2011/2011-Baumeister-KFC-AMCS.pdf

[19] W. T. Adrian, P. Ciężkowski, K. Kaczor, A. Ligęza, and G. J. Nalepa, "Web-based knowledge acquisition and management system supporting collaboration for improving safety in urban environment," in *Multimedia Communications, Services and Security: 5th International Conference, MCSS 2012: Kraków, Poland. May 31-June 1, 2012. Proceedings*, ser. Communications in Computer and Information Science, A. Dziech and A. Czyżewski, Eds., vol. 287, 2012, pp. 1–12. [Online]. Available: http://link.springer.com/book/10.1007/978-3-642-30721-8/page/1

[20] S. Bobek, G. J. Nalepa, and W. T. Adrian, "Mobile context-based framework for monitoring threats in urban environment," in *Multimedia Communications, Services and Security: 6th International Conference, MCSS 2013: Kraków, Poland. June 6-7, 2013. Proceedings*, 2013.

[21] G. J. Nalepa, "PlWiki – a generic semantic wiki architecture," in *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems, First International Conference, ICCCI 2009, Wroclaw, Poland, October 5-7, 2009. Proceedings*, ser. Lecture Notes in Computer Science, N. T. Nguyen, R. Kowalczyk, and S.-M. Chen, Eds., vol. 5796. Springer, 2009, pp. 345–356.

[22] G. J. Nalepa, "Collective knowledge engineering with semantic wikis," *Journal of Universal Computer Science*, vol. 16, no. 7, pp. 1006–1023, 2010. [Online]. Available: http://www.jucs.org/jucs_16_7/collective_knowledge_engineering_with

[23] W. T. Adrian, S. Bobek, G. J. Nalepa, K. Kaczor, and K. Kluza, "How to reason by HeaRT in a semantic knowledge-based wiki," in *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2011*, Boca Raton, Florida, USA, November 2011, pp. 438–441. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6103361&tag=1

[24] G. J. Nalepa, W. T. Adrian, S. Bobek, and P. Maślanka, "Combining AceWiki with a CAPTCHA system for collaborative knowledge acquisition," in *ICTAI 2012: 24th IEEE International Conference on Tools with Artificial Intelligence : November 7-9, 2012, Athens, Greece*, 2012, pp. 405–410. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6495074

[25] A. Ligęza, K. Kluza, G. J. Nalepa, W. T. Adrian, and T. Potempa, "Artificial intelligence for knowledge management with bpmn and rules," in *AI4KM 2012 : 1st international workshop on Artificial Intelligence for Knowledge Management at the biennial European Conference on Artificial Intelligence (ECAI 2012): August 28, 2012, Montpellier, France*, E. M.-L. [et al.], Ed., 2012, pp. 27–32.

[26] G. J. Nalepa and W. T. Furmańska, "Pellet-HeaRT – proposal of an architecture for ontology systems with rules," in *KI 2010: Advances in Artificial Intelligence: 33rd annual German conference on AI: Karlsruhe, Germany, September 21-24, 2010*, ser. Lecture Notes in Artificial Intelligence, R. Dillmann and et al., Eds., vol. 6359. Berlin; Heidelberg: Springer-Verlag, 2010, pp. 143–150. [Online]. Available: http://www.springerlink.com/content/r46p8m40432n7342/

[27] G. J. Nalepa and W. T. Furmańska, "Integration proposal for description logic and attributive logic – towards Semantic Web rules," in *Transactions on Computational Collective Intelligence II*, ser. Lecture Notes in Computer Science, N. T. Nguyen and R. Kowalczyk, Eds. Springer Berlin / Heidelberg, 2010, vol. 6450, pp. 1–23. [Online]. Available: http://www.springerlink.com/content/m388651626832551/

[28] G. J. Nalepa, K. Kluza, and U. Ciaputa, "Proposal of automation of the collaborative modeling and evaluation of business processes using a semantic wiki," in *Proceedings of the 17th IEEE International Conference on Emerging Technologies and Factory Automation ETFA 2012, Kraków, Poland, 28 September 2012*, 2012.

[29] N. T. Nguyen, "Inconsistency of knowledge," in *Advanced Methods for Inconsistent Knowledge Management*, ser. Advanced Information and Knowledge Processing. Springer London, 2008, pp. 1–12.

[30] A. Ligęza, "A 3-valued logic for diagnostic applications," in *Diagnostic REAsoning: Model Analysis and Performance, August, 27th, Montpellier, France*, A. G. Yannick Pencolé, Alexander Feldman, Ed., 2012. [Online]. Available: http://dreamap.sciencesconf.org/conference/dreamap/eda_en.pdf

[31] N. D. Belnap Jr, "A useful four-valued logic," in *Modern uses of multiple-valued logic*. Springer, 1977, pp. 5–37.

[32] E. Miller and F. Manola, "RDF primer," W3C, W3C Recommendation, Feb. 2004, http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

[33] P. Hayes, "RDF semantics," W3C, W3C Recommendation, Feb. 2004, http://www.w3.org/TR/2004/REC-rdf-mt-20040210/.

[34] A. Analyti, G. Antoniou, C. V. Damásio, and G. Wagner, "Extended rdf as a semantic foundation of rule markup languages," *Journal of Artificial Intelligence Research*, vol. 32, no. 1, pp. 37–94, 2008.

[35] N. T. Nguyen, "Ontology integration," in *Advanced Methods for Inconsistent Knowledge Management*, ser. Advanced Information and Knowledge Processing. Springer London, 2008, pp. 241–262.

[36] Z. Huang, F. van Harmelen, and A. ten Teije, "Reasoning with inconsistent ontologies," in *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*, Edinburgh, Scotland, August 2005, pp. 454–459.

[37] S. Schlobach and R. Cornet, "Non-standard reasoning services for the debugging of description logic terminologies," in *International Joint Conference on Artificial Intelligence*, vol. 18. LAWRENCE ERLBAUM ASSOCIATES LTD, 2003, pp. 355–362.

[38] G. Qi, Y. Wang, P. Haase, and P. Hitzler, "A forgetting-based approach for reasoning with inconsistent distributed ontologies," *Institute AIFB, Institute AIFB, University of Karlsruhe, Germany*, 2008.

[39] Z. Huang and F. Harmelen, "Using semantic distances for reasoning with inconsistent ontologies," in *Proceedings of the 7th International Conference on The Semantic Web*, ser. ISWC '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 178–194.

[40] J. Fang and H. Zhisheng, "A new approach of reasoning with inconsistent ontologies," in *CSWS'10*, 2010.

[41] Y. Ma, Z. Lin, and Z. Lin, "Inferring with inconsistent OWL DL ontology: A multi-valued logic approach," in *Current Trends in Database Technology – EDBT 2006*, ser. Lecture Notes in Computer Science, T. Grust, H. Höpfner, A. Illarramendi, S. Jablonski, M. Mesiti, S. Müller, P.-L. Patranjan, K.-U. Sattler, M. Spiliopoulou, and J. Wijsen, Eds. Springer Berlin Heidelberg, 2006, vol. 4254, pp. 535–553.

[42] Y. Ma and P. Hitzler, "Paraconsistent reasoning for OWL 2," in *Web Reasoning and Rule Systems*, ser. Lecture Notes in Computer Science, A. Polleres and T. Swift, Eds. Springer Berlin Heidelberg, 2009, vol. 5837, pp. 197–211.

[43] G. J. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[44] A. Ligęza and G. J. Nalepa, "A study of methodological issues in design and development of rule-based systems: proposal of a new approach," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 117–137, 2011.

[45] T. Eiter, M. Fink, and A. Weinzierl, "Preference-based inconsistency assessment in multi-context systems," in *Logics in Artificial Intelligence*, ser. Lecture Notes in Computer Science, T. Janhunen and I. Niemelä, Eds. Springer Berlin Heidelberg, 2010, vol. 6341, pp. 143–155. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15675-5_14

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1227–1234

Internet as the Source for Acquiring the Medical Information

Magdalena Czerwińska
Lublin University of Technology
ul. Nadbystrzycka 38, 20-618
Lublin, Poland
Email: m.czerwinska@pollub.pl

***Abstract*—The purpose of the present paper is to discuss the results of research conducted in 2012 in order to determine the role of Internet as the source for acquiring the medical information in a group of students. Obtained results confirm the assumed working hypothesis about big role of Internet as the source of medical information that the searching for medical information in Internet is the principal sign of the use of ICT solutions in health protection (as one of e-health elements). It has been examined what kind of information is searched by the patients, what Internet sources are used, what is the reason of the searching for medical information in Internet, what are the barriers encountered by the patients. Details are included in this paper.**

I. Introduction

IN contemporary world, information and communication technologies (ICT) become more and more popular and are used in almost all areas of human activity. E – trade, bank transactions (e-banking), communication (e-mail and mobile telephony), e-learning etc. are the examples of specific applications. The information and communication technologies can be also successfully used in the health care systems and in the whole sector associated with medical services. Thanks to the development of computer technologies and Internet it was possible to create new possibilities for doctor, health service managers and the patients. In case of patients, it is possible to apply e-health tools in services addressed directly to this group (electronic health accounts, medical information and education accounts, Internet pharmacies). E-health is one of the most significant and complex types of e-business [1-4].

As a result of the creation of Internet network we have been introduced into the period of "information society". Continuously increasing amount of media messages (information) is perceived by an average recipient like "information noise" generally occurring in mass-media, particularly in Internet [5]. Therefore the evaluation, selection and conscious choice of valuable, credible and reliable information is an extremely important ability today [6].

Internet became one of essential sources of medical and health information [7]. Said information is quickly and easily accessible in network resources. However their quality, reliability and credibility may give rise to many concerns [8]. Therefore the use of media for pro-health education should be associated with the formation of a positive but simultaneously critical attitude toward media messages.

Because the Internet more and more frequently is used as a serious and often the first source of information about diseases and health, the reliability of information available on the Internet is problematic [9]. The growing role of the Internet as a source of medical information and the quality of the information is highlighted in the literature [10–17]. Another alarming phenomenon is the use of information obtained on-line instead of real doctor's visit.

Increasing group of Internet users search on line for the medical information concerning healthy style of life, nutrition, diseases and their treatment. Internet health services constitute a dynamically developing sector – five years ago they have been visited by every fifth Internet user and currently by the every other [18].

The spectrum of proposals offered in that scope by the Internet is very wide: encompasses comprehensive information about the symptoms of diseases and their treatment, possibility of consultation with the specialist or exchange of experience with other patients, searching of location of medical entities, checking of patients opinions about specified doctor or entity, possibility of drugs purchase etc. without the necessity to leave home.

The fact that the Internet is often used by the Internet users as the source of information about health has been showed by studies carried out in a narrow local scope as well as studies on international level for example Global Health Survey 2011 – the studies carried out in 28 countries worldwide by research institutes associated in Iris network (International Research Institutions). The survey was carried out on population consisting of 22493 adults between August and October 2011 by means of diversified research methods: online questionnaires, CATI and direct interviews. The following countries participated in the survey: Ireland, Hungary, Slovenia, United Kingdom, Finland, Poland, Holland, Lithuania, Greece, France, Romania, Germany, Turkey, Russia, Italy, Ukraine, USA, Chile, Canada, Columbia, Thailand, Indonesia, Pakistan, Malaysia, China, India, Egypt, Australia. In Poland, the survey was completed by ARC Rynek and Opinia on the group consisting of 812 respondents.

The information on health is increasingly popular among Internet users – and searched more often than information concerning the culture and entertainment. The searching spectrum is very wide (information about diseases, their symptoms and treatment, opinions about doctors and medical entities, services offered by the entities, consultations

978-83-60810-53-8/$25.00

with specialists and other Internet users, healthy style of life, diet etc.). Acquired knowledge is very often used by the Internet users in practice. Particular attention deserves huge impact of opinions expressed by other persons not associated with medical sector on behaviours associated with the health and treatment – as many as 63% Internet users declare that they search for the opinions of other Internet users who had similar problem [19]. Obtained result have been also confirmed by the studies carried out for the needs of the present article where analogical indicator was equal to 59.6%.

The searching for information about health by the patient out of the doctor's office is nothing new. Only the alternative information sources are changed. Before the family, friends, neighbours were the information sources. Currently, in the period of ICT technology and society development, such role is performed by Internet. Its character is more anonymous. The knowledge obtained from Interned is verified by a part of patients in course of their visit in doctor's office. Unfortunately, some patients use the information found in the network without contacting the doctor and try to cure themselves. It is not problematic in case of minor diseases but is very alarming and negative phenomenon in case of major health problems. The quality of obtained information is also important. The functioning of specialized portals and forums dedicated for specific diseases should be evaluated positively, because they are visited not only by the patients but also by high class medical specialists and can be used as the source of reliable, useful and practical information.

II. Results of Empirical Research

The present article is an attempt to examine the role of Internet as the source of medical information.

The purpose of conducted research was to determine the types of information searched for by the patients, the kinds of Internet sources used, the reasons of medical information searching in the Internet and the barriers encountered by them. The intention of the author was to determine the attitudes and opinions concerning the acceptance of Internet as the source of medical information, the inclination to use this type of solution and awareness of importance of and potential benefits from the information obtained this way.

A. Own research methodology

The results presented herein constitute a fragment of wider studies concerning the social awareness, level of knowledge and perception of e-health.

The respondents group consisted of students subdivided into two groups. The members of the first group were the full-time students of public health field of study in the Medical University in Lublin (MU) and the members of the second group were the students of extramural management studies in Lublin University of Technology (PL).

The first group of respondents has been selected due to the fact that they will employed in the environment associated with the rendering of medical services in future and that they will among others responsible for efficient implementation of e-health applications and for use of Internet for the needs associated with medical services. Because e-health is the future of medical services, future doctors, health service managers as well as the patients will be to some extent "sen-

Table I.
Social and demographic features of the respondents

Variable	MU		PL		Total	
	Number	%	Number	%	Number	%
Age						
18-25	35	74.5	42	89.4	77	81.9
26-40	12	25.5	5	10.6	17	18.1
Gender						
Female	28	59.6	33	70.2	61	64.9
Male	19	40.4	14	29.8	33	35.1
Size of household						
1 person	3	6.4	5	10.6	8	8.5
2 persons	11	23.4	0	0	11	11.7
3 persons	12	25.5	19	40.4	31	33
4 persons	15	31.9	18	38.4	33	35.1
5 persons	4	8.5	0	0	4	4.3
6 persons and more	2	4.3	5	10.6	7	7.4
Employment status						
Student	47	100	47	100	94	100
Working student	19	40.4	33	70.2	52	55.3
Residence						
Village	1	2.1	23	49	24	25.5
Town up to 20 000 inhabitants.	1	2.1	9	19.2	10	10.6
Town up to 20-50 000 inhabitants.	3	6.4	5	10.6	8	8.5
Town up to 50-100 000 inhabitants.	6	12.8	5	10.6	11	11.7
Town up to 100-200 000 inhabitants.	7	14.9	0	0	7	7.5
Town up to 200-500 000 inhabitants.	29	61.7	5	10.6	34	36.2
Town above 500 000 inhabitants.	0	0	0	0	0	0

Source: own study on the basis of conducted research

tenced" to functioning in virtual environment. Therefore it seems to be interesting to examine among others how is e-health perceived by the students of medical disciplines, what is their knowledge and awareness in that field and what are their attitudes to Internet as the source of medical information.

Obtained results have been compared with those obtained from the group of students in Lublin University of Technology in order to check whether the professional profile of the respondent diversifies the respondents attitudes to Internet as the source of medical information.

The survey was conducted in the both groups in June 2012. The size of each group of respondents was identical and equal to 47 persons. A structured questionnaire form containing the open and closed questions has been used as the research tool (41 questions in the main questionnaire form and 6 questions in demographics).

B. Profile of respondents

The social and demographic features of the respondents are presented in Table I. The women were prevailing group in the survey (64.9%) i.e. in the group of students in the Medical University in Lublin (59.6%) as well as in the group of students in Lublin University of Technology (70.2%). More than one third of respondents were the members of four persons household (35.1%). 55.3% respondents have a regular job. It is obvious that the percentage of working respondents was higher in the group of students of extramural management studies (PL) and was equal to 70.2%. The inhabitants of Lublin prevailed in the group of full-time students (MU) (61.7%) and the students of extramural management studies came from the rural areas (49%).

C. Use of the Internet for health purposes

The place and intensity of the use of Internet network as well as aims of its use in the context of medical information searching were analyzed in examined group of students. The intensity has been determined through the frequency and the period of experience in the use of Internet network. There is no significant difference in the scope of a/m indicators between the students of Medical University and Lublin University of Technology.

Table II contains the information about the use of computers and Internet by the respondents. All respondents have their own computers and are active users of Internet. Majority of respondents use the Internet everyday (77.7%). Almost one fourth of respondents gained more than 10 years' experience in the use of Internet. Mainly the following places of Internet use are specified: home (90.4%), work place (40.4%), university (24.5%) and the use in any place by means of wireless access (24.5%). 93.6% of respondents evaluated their skills associated with the use of computers and Internet as good or very good. The students of Medical University were more critical in their assessments. Part of them evaluated their skills as sufficient (10.6%) or poor (2.1%).

Internet is used by the respondents mainly for searching of information about health, diseases and treatment methods. Therefore it is used by 89.4% of respondents from MU and 80.9% of respondents from Lublin University of Technol-

TABLE II.
INFORMATION ABOUT COMPUTERS AND INTERNET

Variable	MU		PL		Total	
	Qty	%	Qty	%	Qty	%
Possession of computer						
Yes	47	100	47	100	94	100
No	0	0	0	0	0	0
Frequency of Internet use						
Every day	40	85,1	33	70,2	73	77,7
Several times per week	7	14,9	14	29,8	21	22,3
Experience						
2-5 years	6	12,8	10	21,3	16	17
5-10 years	28	59,6	28	59,6	56	59,6
Above 10 years	13	27,6	9	19,1	22	23,4
Place of use						
Home	47	100	38	80,9	85	90,4
Work	19	40,4	19	40,4	38	40,4
Family / place	3	6,4	13	27,7	16	17
School / university	18	38,3	5	10,6	23	24,5
Everywhere (wireless access)	13	27,6	10	21,3	23	24,5
Access points; hot-spot type	10	21,3	5	10,6	15	16
Café internet	0	0	0	0	0	0
Assessment of skills in the scope of computer and Internet						
Very good	27	57,5	24	51,1	51	54,2
Good	14	29,8	23	48,9	37	39,4
Sufficient	5	10,6	0	0	5	5,3
Poor	1	2,1	0	0	1	1,1

Source: own study on the basis of conducted research

ogy. The aims of the use of Internet are presented in Table III. Obtained results confirmed the studies [20], carried out previously and demonstrating that e-health mainly consists in searching the information concerning the health and healthy style of life and, in further alternative, shopping in virtual pharmacies or performing other activities. Therefore the practical use of Internet in the area of medical services mainly consists in searching for information. High percentage of responses informing about the making use of advices given by other patients (almost 60% of respondents) is an alarming phenomenon without any impact of professional profile of the respondent. The scope of use of Internet is narrower in case of students from Lublin University of Technology – the following activities did not occur in this group at all: checking of dosages and undesired effects of drugs prescribed by doctor, asking questions concerning determined medical problem at discussion forum, advices given to other patients, ordering a prescription, ordering a doctor home visit.

The main Internet sources used by the respondents searching for information associated with the health and medicine are: search engines (70.2%), health portals for everyone (63.8%), general (48.9%) and specialized (31.9%) discussion forums, thematic websites prepared by large portals (39.4%), Internet pharmacies and medical shops (36.2%), websites of doctors and medical entities (34%), Internet portals for patients (17%), community services (8.5%), websites of scientific associations (5.3%), websites of drugs manufacturers (5.3%), websites of National Health Fund (8.5%) and Ministry of Health (4.3%). The search engines, health portals for everyone, the general (48.9%) and specialized discussion forums as well as Internet portals for patients are more frequently used by the students of Medical University. However the community services, websites of scientific associations, websites of drugs manufacturers and website of Ministry of Health did not occur at all in the group of sources indicated by the students from Lublin University of Technology.

The respondents asked to arrange various elements associated with health protection from the most important to completely insignificant ones gave the answers contained in the Table IV. In the group of students of Medical University, the largest group of respondents (51.1%) finds that the access to information / advices concerning health, diseases prevention or correct nutrition is very important.

The other issues found important by the respondents were: access via Internet to the list of medical entities and catalogue of their services, to personal health records and the possibility to purchase drugs and medical equipment via Internet. The possibility of on-line consultation with the doctor has been found less important and the possession of health card in electronic version completely unimportant.

But none of expressed opinions prevailed as very important in the group of the students from Lublin University of Technology. The access via Internet to the list of medical entities and their services, to information/advices concerning health, diseases prevention and correct nutrition etc. as well as the access to personal health/diseases records was found important by the respondents. The possibility of on-line consultation with the doctor, possibility to purchase drugs and medical equipment via Internet, possibility of discussion on forums about medical issues and appointment of visits at the doctor has been found less important and the possession of health card in electronic version completely unimportant.

The respondents identify the positive effects of applications of information technologies in the health service. The students from Medical University emphasize that said effects improve the health care system (70.2%), make it possi-

TABLE III.
AIM OF INTERNET USE

Aim of Internet use	MU		PL		Total	
	Freq.	%	Freq.	%	Freq.	%
Verification of opinions about specified doctor	28	59,6	13	27,7	41	43,6
Making use of advices given by other patients	28	59,6	28	59,6	56	59,6
Searching for addresses of medical entities	22	46,8	27	57,4	49	52,1
Verification of opinions about specified medical entities and doctors	21	44,7	23	48,9	44	46,8
Searching for information about the effect of specified drug	20	42,6	14	29,8	34	36,2
Checking undesired effects of drugs prescribed by doctor	18	38,3	0	0	18	19,1
Receipt of examinations results	15	31,9	18	38,3	33	35,1
Searching for information about the treatment or diagnostic method ordered by the doctor	15	31,9	9	19,1	24	25,5
Checking dosage of drug prescribed by doctor	15	31,9	0	0	15	16
Asking questions concerning determined medical problem at discussion forum	14	29,8	0	0	14	14,9
Purchase of drugs and medical equipment	11	23,4	6	12,8	17	18,1
Appointment of the visit in medical entity	7	14,9	13	27,7	20	21,3
Giving advices to other patients	7	14,9	0	0	7	7,4
Asking questions concerning health to experts accessible in Internet	4	8,5	5	10,6	9	9,6
Ordering a prescription	2	4,3	0	0	2	2,1
Checking a place in queue waiting for sanatorium or surgery	1	2,1	4	8,5	5	5,3
Online access to his/ her medical documentation	1	2,1	9	19,1	10	10,6
Ordering a doctor home visit	1	2,1	0	0	1	1,1

Source: own study on the basis of conducted research

ble to save the doctor's and patient's time (66%), improve the quality of medical services (57.4%). They also make the contacts with health service and medical services providers more efficient (55.3%), reducing the functioning costs of medical entities (51.1%) and increasing the access to medical services (46.8%). Almost every third respondent (29.8%) reported a positive impact of Internet on the society education in the scope of health protection. In the opinion of respondents the competition on medical services market is also positively affected by ICT technologies (23.4%). According to the natural order of things, the students from Medical University emphasized the positive aspects of computerization from the point of view of the suppliers or organizers of medical services not from the patients' point of view.

In the opinion of the students from Lublin University of Technology, the most important effect of health service computerization is better efficiency of the contacts with health service and medical services providers (70.2%) and saving the doctor's and patient's time (61.7%). Then they report the improvement of the health care system (48.9%) and increased access to medical services (48.9%). The next effects are: reduced costs of medical entity operation (40.4%), assistance in maintenance of health (21.3%) and improved quality of medical services (19.1%). The positive effects of applications of information technologies in the health service are identified with the society education in the scope of health protection in the scope of health protection and with the improvement in competition on medical services market by every tenth respondent only.

Due to the awareness of existence of the effects of applications of information technologies in the health service, 44.7% of respondents in the group of students from Medical University strongly believes that Internet is the future of medicine. Similar group (48.9%) is more sceptic but also can see the chances for health care development in Poland by means of Internet.

However the attitude presented by the students from Lublin University of Technology is a little bit different. The answer: Definitely YES to the question whether Internet is the future of medicine was given in this group by every tenth respondent only. The answer: YES was declared by 38.3% of respondents but as many as 40.4% did not expressed any opinion about this subject.

As appears from the conducted research, Internet is a medium most frequently and preferably used by the respondents gain the knowledge in the scope of health, diseases and treatment. The network as the preferred source of information about health, diseases and treatment is indicated by 77.7% respondents. The next answers were as follows: doctors/ representatives of health service – 74.5%, family – 56.4%, friends – 44.7%, books/handbooks – 29.8%.

The hierarchy of knowledge sources was different in the both groups of respondents. The following hierarchy occurred in case of MU students: doctors/ representatives of health service (80.9%), Internet (72.3%), family (40.4%), books/handbooks (40.4%). The following opinions prevailed in case of the students from Lublin University of Technology: Internet (83%), family (72.3%), doctors/ representatives of health service (68.1%), friends (63.8%).

As many as 85.1% of respondents (89.4% of MU students and 80.9% of the students from Lublin University of Technology) searched for health information in Internet. Such information is searched by 40.4% of respondents (MU students) frequently and as many as 61.7% of the students from Lublin University of Technology do it always if it is necessary.

Most frequently information about health encompass the information about diseases and their symptoms (78.7% MU and 91.5% PL), location and the offer of medical entities (59.6% MU and 59.6% PL) as well as effect and composi-

TABLE IV.
OPINIONS ABOUT VARIOUS ELEMENTS ASSOCIATED WITH HEALTH PROTECTION

Topics	Very important		Important		Less important		Completely unimportant	
	MU	PL	MU	PL	MU	PL	MU	PL
Access via Internet to the list of medical entities and their services in your voivodship	18	15	24	22	5	10	0	0
Access via Internet to the information / advices concerning health, diseases prevention and correct nutrition etc..	24	15	18	24	4	4	1	4
Appointment of visits at the doctor via Internet/email	12	0	23	18	10	25	2	4
Access via Internet to the to personal health / diseases records	13	4	24	24	8	10	2	9
possession of health card in electronic version	10	0	14	13	16	20	7	14
Possibility to purchase drugs and medical equipment via Internet	8	0	24	9	13	29	2	9
On-line consultation with the doctor	8	0	19	13	18	30	2	4
Possibility of discussion on forums about medical issues	7	0	18	15	17	28	5	4

Source: own study on the basis of conducted research

tion of drugs (55.3% MU and 61.7% PL). Other areas of medical information searched in Internet are presented in Table V.

The great majority of respondents (87.2%) reported Internet as the most popular information source due to the fact that the information are accessible quickly and easily. For almost the half of respondents (47.9%) it is also important that this medium makes it possible to enhance the knowledge about health. Such information is searched by 35.1% of respondents driven just by curiosity.

The respondents placed also the value on: lack of necessity of waiting in queue for doctor's visit (27.7%), time saving (25.5%), possibility to maintain anonymous status (9.6%) and quick diagnosis (8.5%). The quick and easy access to information prevails in the answers given by MU students (83%) and PL (91.5%). The next answers are characterized by differences. The next item specified by MU students is the lack of necessity of waiting in queue for doctor's visit (55.3%), time saving (40.4%), possibility to enhance the knowledge about health (36.2%) and curiosity (31.9%). The next important reasons for the students from Lublin University of Technology is the possibility to enhance the knowledge about health (59.6%) and curiosity (38.3%) only.

The respondents search for medical information in Internet but almost ¾ never arranged a doctor's visit by means of Internet or E-mail. The reason of this state of things in case of MU students is the fact that medical entity being their service provider does not accept the registration in such form (76.5%). However almost every third respondent informs that he/she never checked if it is possible to use this form of registration. However the students from Lublin University of Technology who do not arrange the doctor's visit in this manner prefer conventional methods (55.9%), do not use the entities offering such possibilities (44.1%), or did not check if it is possible to use this form of registration (44.1%). Every fifth respondent in this group informs that it is more quickly and convenient to arrange the doctor's visit by phone. It should be noted that the lack of computer skills or impeded access to Internet is not reported as the reason of failure to arrange the doctor's by means of Internet or E-mail in any group.

The barriers in use of the state of art technologies (Internet, telephony, computers) are also identified by the respondents. What's interesting is that no barriers were reported in that scope by as many as 42.6% respondents in MU students group but such answer was not given at all by the students from Lublin University of Technology. How to explain such optimistic opinions of MU students? Probably it is caused by poor knowledge and awareness of potential hazards. In the present research phase it is possible to suspect the reasons of this state of things i.e. wider knowledge and consequently awareness of potential hazards in the group of students of technical university. Certainly, the issues presented herein can constitute an interesting subject of further research. In the opinion of MU students reporting the barriers in use of the ICT technologies, the principal barriers are the personal data safety concerns (46.8%), hazards associated with viruses, hackers, undesired mail (34%) and doubts regarding Internet transactions safety (19.1%).

The same three barriers are also specified by the students from Lublin University of Technology but in different order i.e.: personal data safety concerns (78.7%), doubts regarding Internet transactions safety (59.6%) and hazards associated with viruses, hackers, undesired mail (21.3%).

III. Conclusion

Conducted research demonstrated that Internet is a valid source of medical information for potential patients. It also appears that searching for medical information in Internet is the principal sign of the use of ICT solutions in health protection and practical use of Internet in the area of medical services is limited to the searching for information mainly. It

Table V.
Thematic areas of medical information searched in Internet

Topics	MU		PL		Total	
	Freq.	%	Freq.	%	Freq.	%
Diseases and their symptoms	37	78,7	43	91,5	80	85,1
Interpretation of examinations results	17	36,2	24	51,1	41	43,6
Effect and composition of drugs	26	55,3	29	61,7	55	58,5
Alternative treatment methods	13	27,7	5	10,6	18	19,1
Physical activity – fitness, sport	20	42,6	23	48,9	43	45,7
Healthy style of life, diet, weight loss, vitamins, diet supplements	23	48,9	19	40,4	42	44,7
Stimulants, addictions	8	17	0	0	8	8,5
Stress, mental health, depression	9	19,1	10	21,3	19	20,2
Sexual life	8	17	19	40,4	27	28,7
Location and the offer of medical entities	28	59,6	28	59,6	56	59,6
Opinions about doctors	24	51,1	13	27,7	37	39,4
Health insurance	5	10,6	5	10,6	10	10,6
Alternative medicine, herbs, homeopathy, acupuncture	6	12,8	0	0	6	6,4
Medical procedures	9	19,1	9	19,1	18	19,1
Plastic surgery	2	4,3	5	10,6	7	7,4
Child health	6	12,8	0	0	6	6,4
Allergic reactions	8	17	5	10,6	13	13,8

Source: own study on the basis of conducted research

can be questioned whether it is caused by poor spectrum of e-health services or lack of awareness of the respondents in the scope of such opportunity or the respondents are unable to change their previous habits. These issues can constitute an interesting subject of further research.

There are no significant differences between the students from Medical University and the students from Lublin University of Technology in the scope of indices characterizing the place and intensity of Internet use.

It seems that the attitude of respondents toward Internet as the source of medical information is diversified by the Professional profile of respondent, as demonstrated in course of the research.

Significant differences in examined groups have been observed in the following:

1) subjective evaluation of their own computer and Internet skills:
   - the students from Lublin University of Technology better assess their ability to use computers and the Internet,
2) searching spectrum width:
   - the scope of use of Internet is narrower in case of students from Lublin University of Technology – the following activities did not occur in this group at all: checking of dosages and undesired effects of drugs prescribed by doctor, asking questions concerning determined medical problem at discussion forum, advices given to other patients, ordering a prescription, ordering a doctor home visit,
3) size of catalogue of Internet sources used by those who search for medical information:
   - students from Lublin University of Technology also exibit limited range of online sources used in search of medical information,
4) identification of positive effects of information technologies application in health service:
   - the students from Medical University emphasized the positive aspects of computerization from the point of view of the suppliers or organizers of medical services not from the patients' point of view,
5) forecasting of the future of medical services rendered by means of Internet:
   - the students from Medical University are more confident that the Internet is the future of medicine,
6) structure of medical knowledge sources:
   - for the students from Medical University the main sources of medical knowledge are doctors/ representatives of health service, for the students from Lublin University of Technology the main source of medical knowledge is Internet,
7) structure of reasons causing that the Internet is the most popular source of information:
   - for all respondents the main reason why the Internet is the most popular source of information is the fact that it provides quick and easy access to information; however, for other reasons, both groups of students give their different structure,
8) identification of barriers in use of the state of art technologies:
   - no barriers were reported in that scope by as many as 42.6% respondents in Medical University students group but such answer was not given at all by the students from Lublin University of Technology.

References

[1] S. Hinske, P. Ray, "Management of E-Health Networks for Disease Control: A Global Perspective", *IEEE 9th International Conference on e-Health Networking, Application and Services (Healthcom)*, 2007, pp. 52-57

[2] S. Chattopadhyay, Li Junhua, L. Land, P. Ray, "A framework for assessing ICT preparedness for e-health implementations", *IEEE 10th International Conference on e-Health Networking, Application and Services (Healthcom)*, 2008, pp. 124-129

[3] H.J. Wen, J. Tan, "The evolving face of telemedicine e-health: opening doors and closing gaps in e-health services opportunities challenges", *IEEE 36th Annual Hawaii International Conference on System Sciences*, 2003, 12 pp.

[4] S.N Khalifehsoltani, M.R. Gerami, "E-health Challenges, Opportunities and Experiences of Developing Countries", *International Conference on e-Education, e-Business, e-Management, and e-Learning (IC4E '10)*, 2010, pp.264–268

[5] T. Goban-Klas, *Media i komunikowanie masowe. Teorie i analizy prasy, radia i Internetu*, Wyd. Naukowe PWN, Warszawa, 2004, pp.158–289

[6] I. Laudańska-Krzemińska, A. Kaiser, "The internet as a source of information on health in the opinion of university students", *Annales Universitatis Mariae Curie-Skłodowska. Sectio D*. 60 (3 Suppl.16), Medicina 2005, pp.242–247

[7] M. van den Haak,C. van Hooijdonk, "Evaluating consumer health information websites: The importance of collecting observational, user-driven data", *IEEE International Professional Communication Conference (IPCC)*, 2010, pp. 333–338

[8] S.M. Akerkar, L.S. Bichile, "Health information on the Internet: patient epowerment or patient deceit?", *Indian J Med. Sci,* Vol. 58 No. 8, 2004, pp.321–326

[9] [x4] L. Weitzel, P. Quaresma, J.P.M. de Oliveira, "Evaluating Quality of Health Information Sources", *IEEE 26th International Conference on Advanced Information Networking and Applications (AINA)*, 2012 , pp. 655–662

[10] R. Bahati, S. Guy, M. Bauer, F. Gwadry-Sridhar, "Where's the Evidence for Evidence-based Knowledge in Ehealth Systems?", *Developments in E-systems Engineering (DESE)*, 2010, pp. 29–34

[11] D. Banciu, A. Alexandru, "Innovative research concerning eHealth products and services in Romania", *Wireless VITAE Conference*, 2009, pp. 68–72

[12] L. Daraz, J.C. MacDermid, S. Wilkins, J. Gibson, L. Shaw, "Health information from the web – assessing its quality: a KET intervention", *IEEE Toronto International Conference on Science and Technology for Humanity (TIC-STH)*, 2009, pp. 244–251

[13] A. Kralisch, A.W. Yeo, N. Jali, "Linguistic and Cultural Differences in Information Categorization and Their Impact on Website Use", *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS)*, 2006, pp. 93b

[14] J. Lutes, M. Park, Luo Bo Chen Xue-wen, "Healthcare Information Networks: Discovery and Evaluation", *First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, 2011, pp. 190–197

[15] L. Weitzel, P. Quaresma, J.P.M. de Oliveira, "Evaluating Quality of Health Information Sources", *IEEE 26th International Conference on Advanced Information Networking and Applications (AINA)*, 2012, pp. 655–662

[16] Lin Wen-Cheng, Lin Yu-Syuan, Chen Ming-Wei, Hong Wei-Cing, "Meta-Searching Chinese Health Information on the Internet", *9th International Conference on e-Health Networking, Application and Services*, 2007, pp. 100–104

[17] M. van den Haak, C. van Hooijdonk, "Evaluating consumer health information websites: The importance of collecting observational,

user-driven data", *IEEE International Professional Communication Conference (IPCC)*, 2010, pp. 333–338

[18] Report, "Internetowe serwisy o zdrowiu: zawartość, popularność, profil użytkowników, poszukiwane informacje, polskie badania Internetu", available at: pbi.org, March 2011

[19] ARC Rynek i Opinia Research Institute, Report on tendency for the Internet users to consult online, 2012, available at http://www.arc.com.pl

[20] A. Dąbrowska, M. Janoś-Kresło, A. Wódkowski, *E-usługi a społeczeństwo informacyjne*, Difin Warsaw, 2009, pp. 159

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1235–1239

Corporate Amnesia in the Micro Business Environment

Stephen J. Hall and Clifford De Raffaele *IEEE Member*
Middlesex University Malta,
Pembroke, Malta,
Email: stephenhall@bcs.org.uk, cderaffaele@ieee.org

***Abstract*—Corporate amnesia is a phenomenon that has persistently threatened the livelihood of business organizations and their success in commercial activity. Several substantial studies on this observable fact have been undertaken with focus primarily aimed at the large corporations and the small to medium sized organizations. This vulnerability is however evermore present and significant within the smaller of businesses. In the micro enterprise, the impact of corporate amnesia is realized when even a single member of staff is absent for any lengthy period of time or vacates their post altogether. With more than 80% of the workforce in the US and separately in the UK directly engaged within a micro enterprise, the competitive benefits that can potentially be realized by addressing corporate amnesia is significant. This paper will identify the main causes of corporate amnesia within the micro business environment and propose a suitable framework for the enterprise to effectively facilitate the adoption of Knowledge Management and realize the associated competitive benefits.**

I. Introduction

THE information economy has brought about a new wave of opportunities and challenges that have the potential to give organizations a competitive edge over their market rivals. Knowledge Management (KM) is one such potential opportunity, and although the grouping of such terminology is relatively recent, its concepts and methods have been in existence since time immemorial [8]. At a conceptual level, the management of knowledge is represented differently by academics and industrialists. However, it is generally well-agreed that fundamentally KM can deliver operational efficiency resulting in financial benefits to commercial activities. *Blair & Wallman* [3] found that properly implemented KM projects do result in substantial returns on investment (ROI), and *Stankosky* [20] determined that KM has the ability to enhance the performance of an organization by positively influencing intellectual resources.

Today, KM is generally accepted to be represented by a cycle, with iterations commencing with the identification of existing knowledge, and subsequently followed with planning the knowledge to collect, processing of the actual selected knowledge collected, distribution of new knowledge to where it is required, fostering the usage within the organization, controlling and maintaining its use and finally disposing of it when it is no longer required.

A. Corporate Memory

The storage of all knowledge pertaining to an organization is commonly referred to as Corporate Memory (CM). It is the result of collecting, storing and organizing knowledge in a way that it becomes of use (and consequently of value) to the organization. *Dalkir* calls this the repository of organizational knowledge [6]. Inversely, Corporate Amnesia (CA) is viewed as the loss of this organizational knowledge as a result of factors such as staff mobility, absenteeism, shift work and various others. *Kransdorff* describes CA as the failure of an organizations ability to efficiently and effectively use its experience and historical activity to its advantage [12]. This inevitably results in repeated mistakes and at times embarrassing and easily avoidable blunders. The CA phenomenon is further highlighted by *Tiwana* who states that organizations are not aware that they know what they already know [21].

Field describes a case where a large company was forced to withhold the launch of a product due to technical problems, only to find after their competitors beat them to the market, that they had developed a solution to this technical problem fifteen years earlier [10]. In a 2006 report, *Noria Corporation* forecast that by the year 2010, 60 percent of experienced managers will retire from the oil and gas industry resulting in the loss of an incalculable wealth of knowledge [17]. Similarly, NASA has publicly admitted that the knowledge of how to put a man on the moon has been lost and had it to attempt putting a man on the moon at any point in the future, all the research toward that objective would have to be redone [6]. *Andrade, et al.,* hence concludes that the benefits of KM can only be realized if a form of corporate memory is in place [2].

B. The Micro Enterprise

In contrast to the resources of these larger organizations, smaller firms – particularly the micro enterprise, face a rather different reality. Although the staff compliment is much smaller and therefore also the collective knowledge present within the organization, the share of knowledge per capita is often overlooked. In addition, the major economies of the world consist of a very high percentage of micro enterprises [23,24].

An organization that has a research department composed of a team of staff can share knowledge between them; hence if one member is absent other members of the team can utilize their combined knowledge to continue the work.

978-83-60810-53-8/$25.00

Conversely, a smaller organization employing only a single person for research would consequently be crippled if that person left the firm or is absent for any significant amount of time. Moreover, the limited resources found in the micro enterprise results in the excessive reliance on tacit knowledge. Thus implies that a micro-sized organization is consistently on the threshold of corporate amnesia with even the slightest of influence. Since staff members gather and harvest critical knowledge on the way processes are executed and how practices are applied without redundancy, the impact generated by the departure of a staff member inevitably yields severe knowledge gaps within the organization [4]. Moreover, current KM systems administratively overwhelm the micro enterprise and are as such a major contributor towards the reluctance factor this size of organizations face in employing KM systems. Hence, the micro enterprise typically reverts to over reliance on tacit knowledge and unconditional exploitation of generic knowledge found through internet resources.

Following an analysis and review of related articles in literature in Section II, this paper will address the issue of CA by initially identifying the main causes within a micro enterprise in Section III. Section IV will subsequently focus on the Knowledge Capture aspect of the KM cycle - Identification, Planning and Acquirement, since this is the most significant obstacle encountered by micro enterprises when attempting to employ KM to protect their organizations' memory. Finally, Section V will conclude this article by deriving the conclusions of the proposed framework.

II. Literature Review

The essential value of Corporate Memory to an organization was highlighted by *Dalkir* [6], with further research by *Kransdorff* adding detail to the concept by expanding its boundary of benefit [12]. This results in a definition and understanding of its common occurring converse Corporate Amnesia [12]. *Tiwana* [21] accents the reality that most organizations are unaware of the knowledge they possess [21], and confirms this by actual cases of corporate amnesia [6, 10, 17]. *Brossler* [4] undergoes a study to explain knowledge gaps derived as a result of staff mobility, a concern which *Moteleb* identifies and deems relevant for both large as well as small to medium sized organizations [15].

The shortage of material in the context of knowledge management specifically addressing the micro enterprise required a definition of this size of business to first be established. This was done by European legislation which considers organizations that have an annual turnover of below two million Euro and employ's less than ten people to be classified as a 'micro-enterprise' [9]. Seen within the KM context, this small size definition of a micro business brings inherent challenges to light such as; the selection of staff incentives for contribution toward knowledge collection [19], the hidden time factor and cost involved in maintaining a knowledge management system [11] and the learning disability found in this size of enterprise [6]. Although these challenges are shared in common with Small and Medium Enterprises, the issues of poor communication of knowledge, fear of knowledge loss and staff reluctant to sharing their knowledge are ever more pronounced in the micro environment [15].

The level of dependency that such organizations have on tacit knowledge and the recognition to its mobility is also worth noting [21]. An interesting study in [1] investigates the reason why Small and Medium Businesses (SMBs) are reluctant to transferring tacit knowledge into explicit knowledge and relevant tacit capture methods are evaluated against suitability qualities that impact the micro enterprise [6, 18].

III. The source of CA

The limited financial and human resources present within a micro enterprise leads it to perceive the benefits of KM and ultimately corporate memory as 'nice to have' but often hard to justify. This is primarily due to the excessive administrative overhead required to implement it. Following an analysis of the components leading to the implementation of KM practices within larger businesses [7], a number of factors hindering implementation in micro organizations are hereunder presented;

A. Incentive

The largest apparent hurdle in implementing a KM system is that of providing and maintaining sufficient incentive for staff to continually contribute knowledge to the system. In law firm environments, it was found that several attorneys see the product of their work as 'their own' rather than that of the firm [19]. This often results in an organizational culture in which employees refuse to share their knowledge in fear of losing the hold on their position. Furthermore, unless properly incentivized employees will seldom find the necessary time to transfer (document) their tacit knowledge into a KM system, a problem ever more pronounced within a micro enterprise due to the various roles each employee is assigned.

B. Cost

Another indirect hurdle which leads an organization to ponder on the applicability of KM is the cost of implementing and maintaining the KM system itself. Management may perceive the investment on the infrastructure and changes to procedures required to support a KM system as prohibitive and unjustifiable. Despite the tangible measures of predicted profitability and competitive advantage, it is nevertheless very difficult to justify the volume of time and money that is invested into managing disparate knowledge resources in a small budget firm [11].

C. Causes

Throughout its lifecycle, an enterprise experiences varied learning capacity difficulties. If an organization forgets its past endeavors and the reason why such tasks were even attempted, it would be equally unable to record and retrieve significant aspects of what it actually knew [6]. Despite the clear advantages of KM, the micro enterprise will typically

not employ a formal system but instead rely exclusively on the tacit knowledge of its staff for the purpose of CM. As a result of this over-reliance on tacit knowledge, CA in a micro enterprise can be summarized as being caused by the downsizing of staff levels, shift-work rotations, high staff turnover, outsourcing of processes and the fact that tacit knowledge is forgotten or not used because it was not associated to the location of its use. Since these identified causes represent the inherent nature of the human workforce in a corporate environment they are not directly preventable. However, by optimizing the conversion and storage of knowledge purposefully, a KM system can help minimize the impact that each of these causes can have on the micro enterprises' corporate memory.

IV. Optimizing K-Capture

Like any other form of business organisation, a micro enterprise harvests two fundamental types of knowledge - tacit and explicit. Tacit KM is the process of capturing, managing and sharing one's experience and expertise when and where it is required [6]. Tacit knowledge is itself split into two areas – Individual and organizational. Individual tacit knowledge, usually present within the minds of 'knowers' and often contributed voluntarily by individual members of staff. Organizational tacit knowledge is carried by the collective grouping of individual tacit knowledge. The other fundamental form of knowledge for the organization is of explicit nature, and this is commonly present in the procedures, processes and documentation stored within the enterprise.

Tacit knowledge is the more volatile of the two types since it is carried by staff members who have a dynamic relationship with the firm and can be considered as an unstable asset in the company's future. This issue, which is highlighted as a risk factor in SMB's who rely on this mobile knowledge [21] is ever more challenging to manage inside a micro enterprise.

A. Knowledge Transfer

The four modes of knowledge transfer represented by *Nonaka & Takeuchi* within their SECI model clearly define the states in which these two types of knowledge can exist [16]. They also indentify the continuous spiral process of organizational learning. KM systems work toward keeping this process of organizational learning in motion by means of implementing a KM cycle. Several established KM cycles have been developed such as those by *Wiig, Meyer & Zack, McElroy, and Bukovitz & Williams* and a general consensus is present throughout these models whereby knowledge capture is recognized as the first phase of each of these cycles. This initial phase is also specified to be the most intrusive and administratively time consuming phase of the entire cycle. Due to such an initial hurdle, SME's regularly opt to retaining most of their knowledge in tacit form primarily due to the shortage of time and resources available to converting it into explicit form [1].

B. Qualities that matter to the Micro Enterprise

The limited human and financial resources available to the micro enterprise mandates that any additional resources allocated towards the processing of knowledge is kept to an absolute minimum. Consequently, as highlighted within the limiting factors of Section III, a successful KM model for micro enterprises requires the process of capturing knowledge to be accurate in nature, performed in a transparent and time-efficient manner and require the least amount of incentive. These factors are further elucidated in this section to concretely analyze the manner in which a micro-enterprise can assure compliance to this required framework.

Transparency - The process of capturing explicit knowledge demands different methods to be explored than that of actually transferring tacit into captured explicit knowledge. The one aspect common to capturing both types of knowledge however is the level of transparency involved in the actual capture process. Due to limitations of human resources and the value attributed to time, the micro enterprise is highly sensitive to having processes and procedures loaded with additional tasks to maintain a system which does not provide much incentive to the contributor or any form of immediate return. *Dalkir* emphasizes that the most important challenge of KM success is user incentive [6]. In the absence of the ideal - a way to transfer tacit knowledge directly into explicit knowledge and the importance of approaching as transparent (time-efficient) a method as possible is a mandatory prerequisite to implementing a KM system in the micro enterprise.

Capture Points - Equally important to the transparency requirement, is the point at which the knowledge is actually captured. In contrast to asking the 'knower' to offload his tacit knowledge on a particular topic by interview or other established means, a transparent point to capture and convert this knowledge is when it is in a state of 'transit'.

Bringing *Nonaka's* SECI model [16] into perspective clearly reveals instances of these so-called 'transit' points.

During the internalization and combination phases tacit knowledge is in 'transit' and can be transparently captured. Equally so occurs during the socialization and externalization phases which are much easier to encounter within the familiar context operated by micro enterprises of few employees.

C. Tacit capture methods

Tacit knowledge is the more challenging of the two knowledge types to capture. *Dalkir* explores several methods of individual Knowledge Capture. The first group explored is that proposed by *Parsaye* [18]. These are; Interviewing experts, learning by being told and learning by observation. Each of these three methods involves disruption of the 'knowers' productivity during the knowledge transfer session and requires an additional person to conduct the session and document the 'externalized' knowledge. A process that is clearly unfeasible with the limited resources available to the micro enterprise. The second group of methods explores the Ad hoc sessions, Road maps, Learning

		Qualities that impact the micro enterprise						
		Conducive to productivity?	Self-motivating?	Tacit to Explicit transfer using existing staff?	Relatively accurate?	Perceved as Transparent?	Readily available Capture Points?	SIUTABILITY (YES COUNT)
K - Capture Methods	Interviewing Experts	NO	NO	NO	YES	NO	YES	2
	Learning by being Told	NO	NO	NO	YES	NO	YES	2
	Learning by Observing	NO	NO	NO	YES	NO	YES	2
	Ad hoc sessions	YES	YES	YES	NO	YES	YES	5
	Road maps	NO	YES	NO	YES	YES	YES	4
	Learning histories	NO	NO	NO	YES	NO	YES	2
	Action learning	NO	NO	NO	YES	NO	NO	1
	E-learning	YES	NO	YES	YES	NO	YES	4
	Learning from others (guests)	N0	NO	YES	NO	NO	YES	2

Fig. 1 The Knowledge Capture comparative matrix

histories, Action learning, E-Learning and Learning from others.

These methodologies are compared within a suitability matrix in Fig. 1 which analysis the potential presented by each Knowledge Capture method to adhere to the desirable qualities required by micro enterprises. Each of the qualities has been given equal weighting to maintain clarity and simplicity but should be evaluated on a per case basis upon application. Fig. 1 imminently portrays the fact that despite the accuracy and quality commonly associated with the Action Learning method, this technique is the least suitable for the micro enterprise. This resulted since the method is disruptive to the 'knowers' productivity, incurs additional costs due to the extra staff required to conduct the exercise, relies on some form of incentive being in place and is also not transparent to the usual day-to-day business procedures and processes. The methods of expert interview, learning by being told and learning by observation are nearly equally unsuitable to the micro enterprise since they require staff incentive, are disruptive to productivity and also require additional human resources to conduct the respective sessions [18]. Conversely, the Ad hoc Sessions represents the most favorable method for the micro enterprise. It impacts positively on all aspects and only falls short on the accuracy of the captured knowledge as a result of its 'real-time' recording of the sessions' events. Of particular interest is that this method can be easily adapted to make use of current communication technologies such as email, chat, video conferencing and instant messaging sessions [6]. These technologies also assist the implementation of suitable and non-invasive capture points within the organizations. Moreover, Ad hoc sessions lack a formal structure and thus can be adapted to whatever format is most suited to the knowledge being captured.

D. Adapting the methodology

The adaptability to various technologies and the informal structure inherent in the Ad hoc method provides for tremendous scope in capturing knowledge from several sources automatically, transparently and at minimal additional cost in time. This is useful since each technology demands its own evaluation and analysis of suitability in relation to the nature of the enterprise.

Capturing knowledge from any form of communication session can be considered as an ideal 'transit' capture point in relation to *Nonaka's* SECI model. Using the right technology to tap-in to these transit feeds can serve to efficiently capture vast amounts of knowledge on any topic that is being processed and exchanged by the 'knowers'. Several technologies to convert speech (tacit) to text (explicit) from telephone or other forms of voice conversations are available on the market. These technologies can also serve to index the capture and make it searchable to the organization. Furthermore, the capturing process can also be easily adopted within the organization for utilization in problems of various domains.

Two important considerations to ensure suitability and reliability of the knowledgebase involve the need to be selective about the sources and the reliable categorization of the captured knowledge to avoid corrupt or redundant entries being generated. The latter can be addressed by adding appropriate meta-data to the event [7], which in turn assists the utilization of auto-categorization algorithms. This meta-data can either be keyed-in by the 'knower' or can alternatively be extracted automatically from the content of the event. Categorizing the capture will provide scope for the knowledge and establish a fundamental level of accuracy for further KM cycle processes to utilize.

V. CONCLUSION

This paper has identified the lack of research that exists on the application of Knowledge Management in the micro enterprise. It has defined Corporate Amnesia and recognized its principle causes in the micro enterprise. Following the identification of knowledge capture as the initial and most

significant hurdle in adopting KM, the research conducted has established the need for an optimized method for capturing knowledge. During the discussion stage, particular focus was placed on the evaluation of established methods used for knowledge capture and thus leads to a framework that is optimized for use in a micro enterprise environment to be proposed.

REFERENCES

[1] Alawneh, A. A., Abuali, A. & Almarabeh, Y. T., 2009. The role of Knwoledge Management in Enhancing the Competitiveness of Small and Medium-Sized Enterprises (SMEs). Communications of the IBIMA volume 10, pp. 98-109.

[2] Andrade, J. et al., 2008. Formal Conceptualization as a basis for a more procedural knowledge management. Decision Support Systems, pp. 164-179.

[3] Blair, M. M. & Wallman, S., 2001. Unseen Wealth. Washington D.C.: Brookings Institution Press.

[4] Brossler, P., 1999. Knowledge Management at a Software Engineering Company. Kaiserslautern, Germany, s.n., pp. 163-177.

[5] Bukovitz, W. & Williams, R., 2000. The Knowledge Management Field Book. London: Prentice Hall.

[6] Dalkir, K., 2011. Knowledge Management in Theory and Practice. s.l.:The MIT Press.

[7] Dataware Technologies Inc., 1998. Seven Steps to implementing KM in your Organisation. [Online] Available at: http://www.systems-thinking.org/kmgmt/km7steps.pdf [Accessed 28 4 2012].

[8] Denning, S., 2012. What is knowledge?. [Online] Available at: http://www.stevedenning.com/Knowledge-Management/what-is-knowledge.aspx

[9] EU Parliament, 2012. [Online] Available at: http://europa.eu/legislation_summaries/enterprise/business_environment/n26026_en.htm

[10] Field, A., 2003. Locking up what your employees know. Harvard Business School - Working Knowledge, 12 May.

[11] Gaunt, R., 1998. The Hidden cost of Knowledge Management. Inside Knowledge, 1 9, p. Vol 2 Issue 1.

[12] Kransdorff, A., 1998. Corporate Amnesia. Oxford: Butterworth Heinemann.

[13] McElroy, M., 1999. The Knowledge Life Cycle. Miami Florida, ICM Conference on KM.

[14] Meyer, M. H. & Zack, M. H., 1996. The design and implementation of information products.. Sloan Management Review Vol 37 Issue 3, Spring, pp. 43-59.

[15] Moteleb, A. A., Woodman, M. & Critten, P., 2009. Towards a Practical Guide fir Developing Knowledge Management Systems in Small Organisations. s.l., Middlesex University e-Centre.

[16] Nonaka, I. & Takeuchi, H., 1995. The knowledge-creating company: How japanese companies create the dynamics of innovation.. s.l.:Oxford University Press..

[17] Noria Corporation, 2006. The real cost of Corporate Amnesia. [Online] Available at: http://www.machinerylubrication.com/Articles/print/890

[18] Parsaye, K., 1988. Aquiring and verifying knowledge automatically. AI Expert, 3(5), pp. 48-63.

[19] Schoch, T. P., 2012. Overcoming the Barriers to Implementing Knowledge Management. [Online] Available at: http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202542061249&thepage=1

[20] Stankosky, M., 2008. Keynote Address ICICKM. s.l., International Conference on Intellectual Capital, Knowledge Management and Organisational Learning., pp. 9-10.

[21] Tiwana, A., 2000. In: The Knowledge Management Toolkit. Upper Saddle River: Prentice-Hall.

[22] Wiig, K., 1993. Knowledge management foundations.. Arlington TX: Schema Press

[23] Small Business at a Glance, 2013. [Online] Available at: http://www.entrepreneur.com/sbe/glance/index.html

[24] Small Business Statistics, 2013 [Online] Available at: http://www.fsb.org.uk/stats

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1241–1246

Knowledge conflicts in Business Intelligence systems

Marcin Hernes
Wrocław University of Economics
ul. Komandorska 118/120, 53-345 Wrocław, Poland
Email: marcin.hernes@ue.wroc.pl

Kamal Matouk
Wrocław University of Economics
ul. Komandorska 118/120, 53-345 Wrocław,
Poland
Email: kamal.matouk@ue.wroc.pl

***Abstract*—This document presents a problem of knowledge conflict appearing in Business Intelligence systems. The structure of such class system in context of knowledge creating is presented in the first part of article. Next, the formal definition of knowledge structure of Business Intelligence, which is necessary to comparing these knowledge, was elaborated. The characteristic, sources and examples of knowledge conflicts is presented in the final part of article. The detecting and resolving of this type of conflicts is necessary, because this allows receiving by user, from the system, the proper reports as a results of analyses. On the basis of these reports the user can takes the decision that lead to satisfying benefits.**

I. Introduction

CONTEMPORARY social and economic environment makes quick and accurate decision-making crucial for the competitiveness of a company. Economy forces company managers to make complex operational, tactical, yet most of all, strategic decisions that influence the future of the organization. Those who actually make decisions in a company, are usually exposed to risk and uncertainty, because they cannot foresee the consequences of their decisions or their predictions have very low probability. Therefore, the entire decision-making process is very complicated. Nowadays, decision making processes employ decision-making support computer systems, as well as Business Intelligence (BI) class systems which are being used more and more often. They are used to support business decision making through smart use of data resources already available in companies [8]. The purpose of Business Intelligence systems is to enable easy and safe access to information in a company, operation of its analysis and distribution of reports within the company and among its business partners, which in turn enables quick and flexible decision making. This allows the company to reach a higher level of flexibility and competitiveness. Because of the necessity to fully integrate business processes in a company, BI systems should currently operate within a sub-system of an integrated management computer system. [2].

However, it often occurs that a BI system generates conflicts of different kinds, especially conflicts of knowledge generated from various types of analyses. Conflicts of knowledge result from the fact that the system may offer different analysis results or solutions of a single problem to the user. In other words, conflict of knowledge occurs when the same objects in the world and the features are given different values [7]. This mainly results from using different methods for business processes analysis. If a conflict of knowledge occurs in the system, the system will not be able to generate a satisfactory decision for the user and, consequently, the decision maker will find it hard to conduct the decision-making process properly. The decision maker will then be forced to make a decision with no help from the system, which is time-consuming, requires much work and can lead to a decision that is out-of-date (belated) and made with incomplete information. This situation has obviously negative influence on the work of the entire organization.

Therefore, the key element of BI systems' operation is to detect and, consequently, properly resolve conflicts of knowledge. This article presented a formal definition of knowledge structure in BI system, as well as sources and characteristics of conflicts of knowledge regarding BI class systems

II. The structure of Business intelligence systems

At present, companies incur significant losses due to improper use of knowledge. Losses resulting from incorrect operation of knowledge management processes are very high and often constitute the main reason for companies going bankrupt. Symptoms for improper use of knowledge in a company are as follows [6]:

- overdue reaction to changes in market environment - the company does not keep up with competition and market needs,
- lack of knowledge at each level of organization - when the quality of work decreases,
- slow performance of tasks - occurs when it takes employees too long to locate the necessary knowledge,
- the problem of production quality - when the adaptation of production process do quality requirements drastically extends the process,
- long sales cycles - when the response time of the seller to the customer's needs extends.

The decisive factor that affects the use of knowledge as intangible resource of a company is efficient management. The essential purpose of knowledge resources management

978-83-60810-53-8/$25.00

is to provide information for managers, which is then used for planning, control and decision making. Access to information needed for efficient management should be enabled by Business Intelligence system, because it has the ability to transform 'raw' (not processed) data into useful information that helps make more accurate decisions in a short time, with operational conditions constantly changing for the company in its environment, which evokes high risk and uncertainty.

Bi systems are often offered by various computer system producers as a complete system that supports a certain business area. Currently, the tools used in BI systems usually include the following technologies:

- ETL - Extraction, Transformation and Loading,
- DW - Data Warehouse,
- OLAP - OnLine Analytical Processing (multi-dimensional real-time data analysis tool),
- DM - Data Mining, software for safe presentation of information (analyses, reports) on the net.

All elements mentioned above are aimed at meeting the needs of different groups of users, such as managers that use pre-defined reports on a daily basis or analysts that design reports individually and prepare various business analyses. Most elements presented use Data Warehouse as one of the potential sources of information (see pic. 1).

The main source of data for Business Intelligence systems are transaction systems such as ERP, CRM, SCM or call center. Sometimes data is also extracted from text files, Excel, Access, e-mail software or websites. All data should be gathered in one place (e.g. data warehouse) so that reports and analyses based on the data are complete. Data gathered in warehouses usually come from many different sources that store particular values in different ways and for that reason they must first undergo the process of ETL 'standardization' (Extraction, Transformation and Loading) [12]. ETL programs transform data. The process begins with extraction which consists in selective mining and loading data from transaction systems and other data sets. Next phase is transformation, i.e. necessary modification of data. For instance, transformation of numerical values signs, conversion of dates or currencies (e.g. dates being converted from English format into polish format, PLN currency into EUR). Last phase is upload of 'refined' data into the warehouse.

Fig 1 clearly shows that BI system uses advanced analytical tools for real-time data analysis, including OLAP or data mining.

OLAP is a tool that allows to perform multi-dimensional analyses and display the results in approximately real time. There are two common groups of OLAP whose main difference is the type of server used to build them. The first group includes all solutions based on ROLAP (Relation OLAP) data base, whilst the second group is built based on specialized MOLAP (MultiDimensional OLAP) servers, also known as MD-OLAP.

Both techniques have their own pros and cons. ROLAP solutions are characterized by the ability to store large volumes of data, relatively easy data modification (resulting from the software used and data structure), but they also have their own disadvantages, such as: data structure complexity (resulting from the necessity to represent multi-dimensional relations in a relation-like manner) as well as performance problems the result from lack of adaptation of relation structures to multi-dimensional analysis [2].

Whereas, MOLAP do have much smaller capabilities of data storage and find it difficult to modify data (it often occurs that data modification leads to rebuilding multi-dimensional structure), but they are also characterized by high performance of multi-dimensional analysis and natural representation of multi-dimensional structures.

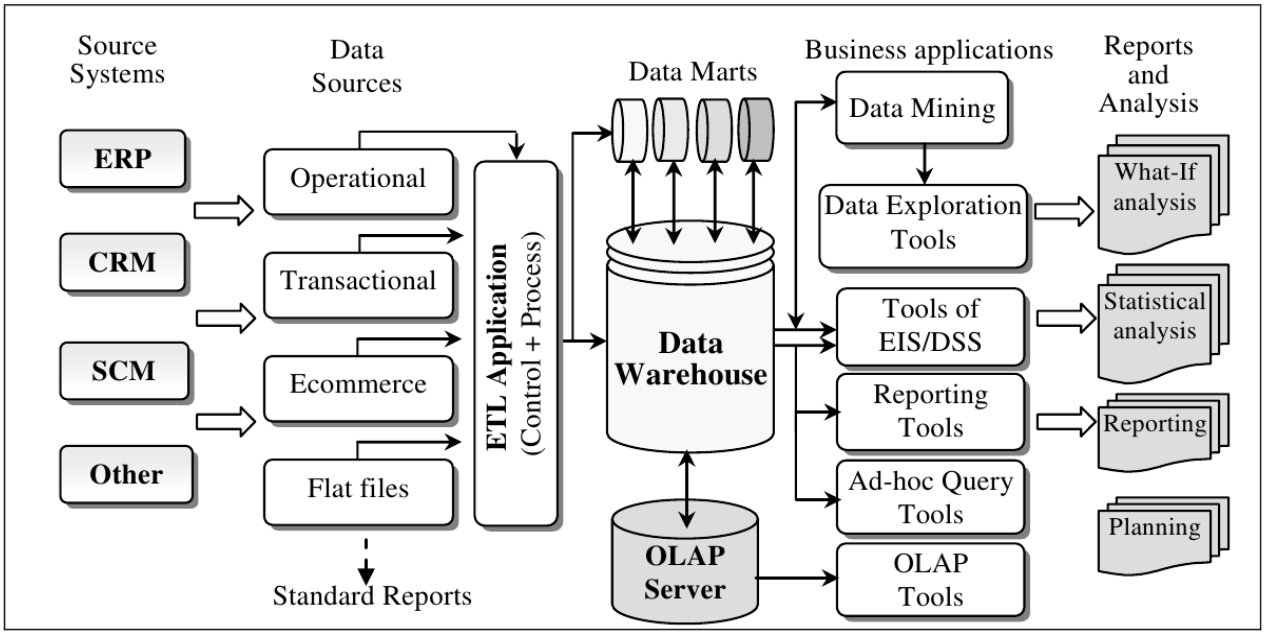


Fig. 1 The overall structure of BI system and its cooperation with other information systems in the enterprise

Source: own work.

A combination of both technologies may ensure a solution that will unite the ability to store large amounts of data and effective multi-dimensional analysis. It will consist in employing a relation data base as data warehouse containing the entire set of elementary data, and MOLAP systems as Data Mart.

A user of OLAP analytical solutions has the ability to perform analysis on available multi-dimensional base (the so-called ROLAP or MOLAP block) or use ready-made reports, defined with the block. Each block has dimensions, also known as perspectives, hierarchies and measures. Data analysis in many dimensions is very intuitive. For instance, while searching for sales figures for a product or a group of products, we are not only interested in general sales figures, but also sales figures categorized by customers and divided into particular periods of time. The area, the customer and time are the dimensions of the analysis and the sales figures are the measures [13].

Reports defined with the OLAP block are being updated while they are generated, therefore they contain current data as accurate as possible until the multi-dimensional block is refreshed again.

ROLAP or MOLAP block analysis consists in performing the following operations: [2]:

- change in detail of data(drill-down, drill-up),
- change in section of analyzed data (slice and dice),
- search for extreme values (exceptions),
- presenting results in the form of graphs,
- contextual switching to detailed data (drill-through).

The 'drill-through' enables the user to proceed within a certain business area and then to switch to other areas with filters engaged beforehand (e.g. time, customer or product dimension). Consequently, it is possible to begin with another multi-dimensional analysis or a pre-defined report or to proceed to 'ad hoc' query environment.

Whereas, Data Mining is an analysis of business data (usually available from data warehouse) in order to detect any rules, relations, patterns and trends contained in them or to set forecasts that may prove useful when making decisions. Therefore, the methods allow to transform data into knowledge.

Data Mining is often considered a contemporary candidate for Artificial Intelligence (Data Mining widely uses neural networks or decision trees). Most often, however, Data Mining employs statistical methods, such as regression, association or classification analysis. The methods are grounds for creating models used to analyze large amounts of data or samples for the existence of certain regularities, hidden relations and similar connections [10]. There are two distinct types of Data Mining [13]:

- hypothesis verification – used when there is a supposition about a significant relation among certain pieces of data and we want to verify it,
- knowledge discovery – used when we want to check if there are connections between pieces of data that man is unable to detect.

The most common use of Data Mining is, for instance, precise segmentation of customers and setting an optimal 'customer basket'. Data Mining allows to know customers better, with their habits, preferences and the risk resulting from customer service. Thus, the company is able to offer proper products or services and gain customers' loyalty.

Using different methods of data analysis enables the system to generate new knowledge, mostly about business processes performed in the company. However, the variety of analyses and the fact that they can be performed with information from heterogeneous sources [8] often leads to a situation where a conflict of generated knowledge occurs. Automated diagnostics and resolving conflicts of such nature by the system is essential, mainly for the sake of proper operation of BI systems, which in turn influences the quickness and accuracy of decisions made by decision makers. However, conducting diagnostic and resolving a conflict of knowledge is only possible when the knowledge is represented as unitary structure, whose definition is will be given in this article.

III. KNOWLEDGE STRUCTURE OF BUSINESS INTELLIGENCE

In order to determine the sources and characteristics of knowledge conflicts in BI it is necessary to formally define the structure of the knowledge gathered in the system. The literature of the subject contains, admittedly, the issues related to this issue, however, they concern only one slice of BI, for example the OLAP cubes [4]. This article presents, the general definition of knowledge structure taking into account the BI as whole.

Assume that the database structure is represented in the form as shown in Figure 2:

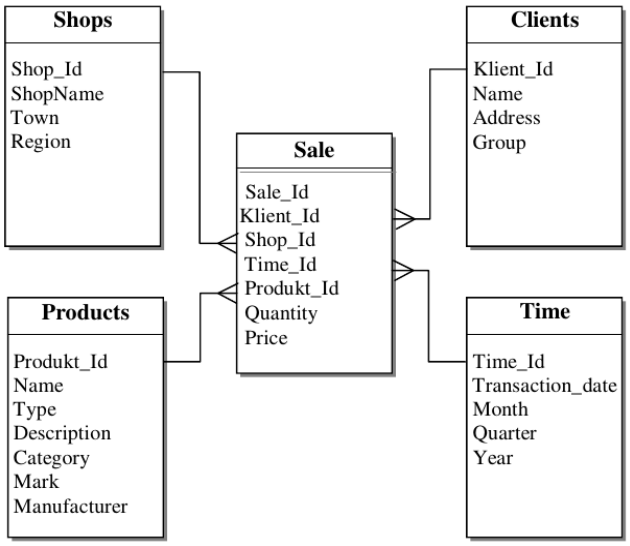


Fig 2. Structure of Data Warehouse

Source: own work.

Accumulation of large amounts of data describing the activities of the company. According to presented the structure of the database, in the long run makes it possible to carry out a detailed analysis of this activity, detection of certain applications, depending on the draw and what further proceedings. For example, company executives may be interested in the implementation of the following analyses:

1. A value for the sales of products by increasing the levels of aggregation: from the city, state and country brand for different time periods.
2. Examine the profit from the sale of goods for individual producers in the designated months. Arrange the months according to the increasing value of the profit.
3. If the deviations contained in the average transaction of individual months are important?

Answering this type of question is made, using the tools to multidimensional data analysis in real time - OLAP. With these types of activities takes over that [9]:

1. Analyzed data resides in databases (operating databases, in data warehouses or data stores), where the volume can be very diverse and reach from mega to multiple terabytes. In maintaining and processing such databases the conventional capabilities of database servers are used.
2. The analytical multidimensional processing, requires the presentation of data using a multidimensional model, where there are terms such as: a table of facts, aggregate functions, dimensions, dimension members, cells, and cell block.
3. The formulation of multidimensional query language-oriented requires multidimensional data query languages. Multidimensional data, as well as the results of the analysis of these data are very clear, when they are visualized graphically.

On the basis of characteristic of BI system, the structure of knowledge is defined as follows:

Definition 1.

The structure of knowledge in BI systems is called following sequence:

$$WBI = \langle \{F\}, \{WYM\}, \{AN\}, \{R\}, \vartheta, SP, DT \rangle$$

where:

1) $F = \langle \left\{ f_1, t_1, m_1^1, m_1^2, \ldots, m_1^n \right\}, \left\{ f_2, t_2, m_2^1, m_2^2, \ldots, m_2^n \right\}, \ldots \ldots \left\{ f_k, t_k, m_k^1, m_k^2, \ldots, m_k^n \right\} \rangle$,

where:

$\left\{ f_1, f_2, \ldots, f_k \right\}$ - denote the set of facts,

$\left\{ t_1, t_2, \ldots, t_k \right\}$ - denote the set of the date of the facts,

$\left\{ m_1^1, m_y^x, \ldots, m_k^k \right\}$ - denote the set of measures assigned another facts $\left\{ f_1, f_2, \ldots, f_k \right\}$ $\left(y \in [1 \ldots k], x \in [1 \ldots n] \right)$,

2) $WYM = \left\{ w_1, w_2, \ldots, w_i \right\}$ - denote set of dimensions,

3) $AN = \left\{ a_1, a_2, \ldots, a_j \right\}$ - denote the set of types of analysis,

4) $R = \left\{ r_1, r_2, \ldots, r_h \right\}$ - denote the set of types of reports,

5) $\vartheta : F \times WYM \times AN \rightarrow R$ - is at least partially a function of knowledge, that mirrors elements of the Cartesian product $F \times WYM \times AN$ in elements of R set. Function ϑ will be partially, when only selected elements of the Cartesian product $F \times WYM \times AN$ will be as its arguments,

6) SP - denote the degree of certainty of reports,

7) DT - denote the date of reports made on the basis of the analysis.

The example of BI structure of knowledge is as follows:

Set of facts (*F*):

Fact1 - On 06-04-2013 sold 10 pieces of the product X – the value of sales 100 EUR.

Fact2: On 07-04-2013 sold 6 pieces of the product Y – value of sales 300 EUR.

Fact3: On 10-04-2013 sold 20 pieces of the product Z – value of sales 200 EUR.

Dimensions (WYM):

Time={06-04-2013 … 10-04-2013},

Product={X, Y, Z}

Territory={Wielkopolskie: Client1, Client2, Dolnośląskie: Client3, Client4},

Client={Client 1: value of sales:100 EUR, Client 2: value of sales: lack of data, Client3: value of sales: lack of data, Client4: value of sales: 200 EUR}.

Set of types of analyses (AN):

Analysis1 - The total value of sales of individual products in the period: 01-04-2013 to 30-04-2013.

Analysis2 - The total value of sales by territory and individual clients in the period: 01-04-2013 to 30-04-2013.

Reports (R):

The report presents the results of the analysis, grouped according to the criteria specified by the with the ability to use pivot tables.

Report1

ϑ(Fact1…Fact*N*,{Time,Product},Analisys1,Report)= Time period=01-04-2013…30-04-2013, Value of sales: ProductX=100;ProductY=300,Product=200, The total value of sales: 600EUR;

Report2

ϑ(Fact1…Fact*N*,{Territory,Client},Analysis2,Report)= Time period=01-04-2013…30-04-2013, Value of sales: Dolnośląskie Voivodeship – Client1=100, Client2=???, Wielkopolskie Voivodeship – Client3=???; Client4=200, The total value of sales: 300 EUR;

There are two reports showing the value of sales grouped according to the given criteria (specified in the function arguments). It can be seen, that although the sales summary should be the same on both reports, however it differ. This may result, for example, of incorrectly entered data or re-

strictions related to the methods of analysis. Therefore, the conflict of knowledge was appeared. It should be clearly pointed out that at the time of the generation of results of analyses as a reports, the user of the system (decision-maker) do not think, why these values differ, because in the turbulently environment decisions must be taken very quickly. It is not a time, for example, to correction by an employee, wrongly entered data (of course, this correction at a later stage should be made, however, this fact may not pause the decision-making process).

So if, as in this example, the structures of knowledge in BI system differ the quantity or the value of the attributes, then the knowledge conflicts appear in this system. These conflicts have been characterized in the later part of the article.

IV. The Knowledge conflicts

Conflicts of knowledge in BI systems result from inconsistency or contradictions in knowledge contained in the system. Inconsistency occurs when one side of the conflict (for instance, one method of analysis) claims that a given attribute (feature) of the world occurs or does not occur in a given period of time, while the other side of the conflict does not have any information or is unable to assess the attribute. Contradiction occurs when one side of the conflict claims that a given attribute of the world occurs in a given period of time, while the other side of the conflict claims that the same attribute does not occur, or the values of the same attribute differ [3,5]. Therefore, conflicts of knowledge when the same objects of the world are given different attributes by different sides of the conflict or the same attributes (features) are given different values by different sides [11]. Obviously, an assumption is made at this point, that the knowledge is represented in a structure, elaborated in this article.

The [8] defined sources of knowledge conflicts, as follows:

1. The fight for managing specific resources. A conflict appears, when first side of the conflict is considered, that the second side of conflict should not has knowledge about a given resource, instead the second side of conflict is considered, that it such knowledge should has.
2. Ideological conflict. It occurs when the parties to the conflict have different beliefs on the subject. These beliefs may arise, for example, with the kind of environment of system works or with adopted algorithms.
3. Requiring the integration of various elements of the system. If there is a need to integrate some elements of the system in one unit, it's naturally a conflict occurs (i.e. different structures of knowledge, different types of knowledge representation).
4. Conflicts resulting from direct knowledge management system. A conflict occurs when each party considers, that it should manage the knowledge accumulated in the system, because it has the current and consistent status of this knowledge.

The last two sources of conflicts of knowledge are most common in BI systems. They are connected with integrating facts, dimensions and analyses into one unit, in order to obtain coherent reports; they are also connected with differences in the system's knowledge represented in structures that differ from one another.

It is worth noting that conflicts of knowledge mainly apply to the difference in amount or value of attributes in structures of knowledge in BI. The situation can be easily illustrated with the following example:

One of the users of the system needs a report on the analysis o sales figures of product X in a given period of time, while another user needs a report on the analysis o sales figures of the same product in the same period of time, categorized by particular characteristics of the product (e.g. the color). It may occur, that the employed method of analysis does not enable performing analysis with categorization by a given attribute (color). Thus, the first user will obtain a specified amount of sales (for instance, 10,000) while the other user will obtain sales at 0. This generates a conflict of knowledge, because the structures of knowledge differ in number of attributes (for example, the 'color' attribute will not occur in the first analysis, but it will occur in the second analysis) and in values of attributes (sales figures will be different in each report).

Another example may be the analysis of settlement of accounts with contractors. We assume that the user needs a report on the analysis of settlement of accounts with customers, categorized by particular products, based on balance of accounts. Next, the user will require a report on the same analysis, but in this case, categorized by particular customers. It may occur that total amounts in both analyses will differ, because, for example, some customers have not been assigned a particular product. This also generates a conflict of knowledge.

A separate problem is generating reports on forecasts by BI system. In this case, structures of knowledge may differ in attributes for many reasons. For instance, different methods of analysis (or different parameters) may generate different figures of predicted sales in the future, even when based on the same range of data. Additionally, even using the same method of analysis, but with detailed categorization by different dimensions, may generate varied values of attributes in each report (for example, if an analysis is performed on the forecasted sales referring to the future, categorized by particular customers and then another analysis is performed in the forecasted sales in the same period of time, but categorized by particular products, then each report may contain a different total amount of forecasted sales).

One must remember that one cannot ignore conflicts of knowledge that occur in BI systems nor uproot them. The conflicts must be located and resolved. Only then the system can perform analyses properly and present reports to the user, and only then can the system do its job.

Knowledge conflict resolving can be carried out using various methods, such as:

a) negotiation methods;

b) deductive-computing methods, based on:

- game theory,
- classical mechanics,
- operational studies,
- behavioral and social sciences,
- choice,
- consensus.

The negotiation methods guarantee the desired compromise, however, this is realized at the expense of increased communication between the nodes of the system, which of course adversely affect its performance. While the methods of deductive-computing group do not affect to a great extent on the speed of operation of the system, they, except the consensus methods, does not guarantee the achievement of a good compromise. Decision-maker, instead, requires a good system performance (often working near the real time) and efficient knowledge conflicts resolving, so the system will effectively support the decision making process.

In order to resolve the knowledge conflicts in BI system, most appropriate methods will be used for the choice or consensus methods, as opposed to other methods, they do not require interference in the internal system statuses (for example you do not need to interfere with the existing programming code). Choice methods rely on the election (on the basis of certain criteria) one of the conflicting knowledge states (represented in the form of the structure presented in this article), which is presented to user. The rest of the states of knowledge are not taken into account in this case, therefore, a high level of risk associated with the choice of the incorrect state of the knowledge. Consensus methods, instead, rely on determining such state of knowledge, that will represent all of conflicting states of knowledge, generated earlier by system. In other words all parties to the conflict will be taken into consideration. This will consequently reduce the level of risk related to choice incorrect state of knowledge. Choice or consensus methods can be implemented as a separate modules of system.

The work on the development of conflict resolving module, with the use of consensus methods, are in progress. The consensus is elaborated in three major stages. In the first stage it is necessary to carefully examine the structure of knowledge. In the second stage it is necessary to define the distance functions among particular structures. The third stage is an elaboration of consensus algorithms that generate a structure, that the distance between this structure (consensus), and the individual structures is minimal (according different criterions).

Algorithms for detecting and resolving the conflicts should be implemented in the system and running, when the structures of knowledge differ. Of course, these algorithms running automatically, without human interaction. Detecting and resolving knowledge conflicts in BI systems allow to certainty of system functioning, in other words, reports generated by system as result of analysis is consistent from the point of view of the criteria defined by the user. Only in this case decision-makers can take full advantage of the system.

V. Conclusion

Conflicts of knowledge occur in virtually every BI system. The system designers should remember that methods for detecting and resolving conflicts should be considered at the first stage of system development. Implementing them after the system has been commissioned may be very difficult due to the need to input additional software code. Proper detection and resolving conflicts is extremely important especially in BI systems, because their operation greatly influences the decisions made by decision makers and, consequently, the operation of the entire organization. Resolving the conflicts is also very important, because only then the system can suggest proper decisions. If the system ignores these aspects, then the user (decision maker) is likely to have problems making a quick and correct decision, because the system may suggest an improper decision, or may suggest several different decisions, forcing the decision maker to spend time selecting one of them, which makes the whole process very time-consuming and impossible to perform in approximately real time.

References

[1] Alsquour M., Matouk K., Owoc M. L. "A survey of data warehouse architectures - preliminary results", *Proceedings of the Federated Conference on Computer Science and Information Systems* (FedCSIS), 2012.

[2] Bytniewski A. (red), *Architektura zintegrowanego systemu informatycznego zarządzania*, Wydawnictwo AE we Wrocławiu, Wrocław 2005.

[3] Hernes M., Nguyen N.T., "Deriving Consensus for Hierarchical Incomplete Ordered Partitions and Coverings" *Journal of Universal Computer Science* 13(2), 317-328, 2007.

[4] Hsu K.C., Li M., *Techniques for finding similarity knowledge in OLAP reports, Expert Systems with Application*s 38 (2011) pp. 3743–3756.

[5] Katarzyniak R., Nguyen N. T., *Model systemu wieloagentowego z procedurami grupowej aktualizacji wiedzy opartymi na metodach teorii konsensusu*, Raport z serii SPR nr 3, ISiTS PWr, Wrocław 2000.

[6] Matouk K., Bytniewski A., *Systemy Business Intelligence w zarządzaniu***,** Prace Naukowe nr 1027, Wyd. AE, Wrocław 2004

[7] Nguyen N.T. , *Metody wyboru consensusu i ich zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych*, Oficyna Wydawnicza Politechniki Wrocławskiej, 2002.

[8] Nycz M., "Business Intelligence as the exemplary modern technology influencing on the development of the enterprise, in: Kubiak B.F., Korowicki A. (ed.), Information Management, Gdansk University Press, Gdańsk 2009r., pp.312-320.

[9] Pankowski T., *Dane wielowymiarowe i język MDX w systemach OLAP*, VI Konferencja PLOUG, Zakopane 2000.

[10] Simon R.A., Shaffer L.S., *Hurtownie danych i systemy informacji gospodarczej*. Oficyna Ekonomiczna. Kraków 2002.

[11] Sobieska-Karpińska J., Hernes M., "Consensus determining algorithm in multiagent decision support system with taking into consideration improving agent's knowledge", *Proceedings of the Federated Conference on Computer Science and Information Systems* (FedCSIS), 2012.

[12] Todman Ch., *Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM),* Wydawnictwa Naukowo-Techniczne, Warszawa 2003.

[13] Wyrębek H., "Znaczenie aplikacji Business Intelligence w zarządzaniu przedsiębiorstwem". in: *Administracja i Zarządzanie*, Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w siedlcach, nr. 88, Siedlec 2011

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1247–1253

One approach to the classification of business knowledge diagrams: practical view

Dmitry Kudryavtsev
Saint Petersburg State Polytechnic University,
Intelligent Computer Technologies Dpt,
ul. Polytechnicheskaya 29, 195251, St. Petersburg,
Russia
Email: dmitry.ku@gmail.com

Tatiana Gavrilova, Irina Leshcheva
Saint Petersburg State University,
Graduate School of Management,
Volkhovsky per. 3, 199004 St. Petersburg, Russia
Email: {gavrilova, leshcheva}@gsom.pu.ru

***Abstract*—Diagrams are an effective and popular tool for visual knowledge structuring. Managers also often use them to acquire and transfer business knowledge. There are many currently available diagrams or visual modeling languages for managerial needs, unfortunately the choice between them is frequently error-prone and inconsistent. This situation raises the next questions. What diagrams/ visual modeling languages are the most suitable for the specific type of business content? What domain-specific diagrams are the most suitable for the visualization of the particular elements of organizational ontology? In order to provide the answers, the paper suggests light-weight specification of diagrams and knowledge content types, which is based on the competency questions and ontology design patterns. The proposed approach provides the classification of qualitative business diagrams.**

I. Introduction

KNOWLEDGE visualization proved to be an effective tool for knowledge creation, acquisition and transfer [5, 6, 13]. Diagrams [2] constitute the basis for visual knowledge representation and elaborated diagrammatic techniques typically form visual modeling languages [17]. In computer science these techniques are reflected in such languages as UML and IDEF. They are also integrated in software engineering methods, e.g. the Structured Analysis and Design Technique (SADT) and are organized by the architecture frameworks, such as the Zachman framework [28].

The focus of this paper is put on the realm of management. Manager also frequently use diagrams in their work [11, 18, 25] but the choice of diagrams is often error-prone and inconsistent [7].

For the effective choice of the visualization method, at least five perspectives should be considered [6]. These perspectives answer five key questions with regard to visualizing knowledge, namely:

1. What type of knowledge is visualized (content)?
2. Why should that knowledge be visualized (purpose, knowledge management process)?
3. For whom is the knowledge visualized (target group)?
4. In which context should it be visualized (communicative situation: participants, place/media)?
5. How can the knowledge be represented (method, format)?

The knowledge type perspective as the focus of the paper, can be used for identifying the type of knowledge with respect to its content. Any complex entity can be represented from several aspects (facets) and at different strata (layers) [13, 28]. The following question-based aspects can be proposed and differentiated [1, 6, 13, 28]:

WHAT-Knowledge: Conceptual representation.
WHAT_FOR-Knowledge: Strategic representation.
HOW_TO-Knowledge: Functional representation.
WHO-Knowledge: Organisational representation.
WHERE-Knowledge: Spatial representation.
WHEN-Knowledge: Temporal representation.
WHY-Knowledge: Causal representation.

Today, there is no validated prescriptive framework that links business diagrams with knowledge types and that offers specific diagram for particular knowledge types. This issue defines the first research question*: What diagrams/ visual modeling languages are the most suitable for the specific type of knowledge (content)?*

The second research question of the paper stems from the task of ontology visualization within different applications. Ontology is a formal, explicit specification of a shared conceptualization. Ontologies and corresponding semantic technologies are actively used for knowledge management, e-commerce, education and semantic web. Currently, each concept of ontology is represented with the same graphical representation independently of its meaning. Graphical representations of ontologies are concerned with the representation of concepts, relations or instances but do not consider a domain specific meaning [21]. Special ontology-based frameworks are developed in order to visualize ontology using domain-specific notations [20, 22, 26]. Some of these frameworks are oriented towards managers and must include knowledge of the currently available popular business diagrams/visual modeling languages with the associated semantics. It defines the second research question: *What diagrams/ visual modeling languages are the most suitable for the visualization of the particular ontology view (elements of ontology)?*

The research was held with the financial support of the Russian Foundation for Basic Research (RFBR), project # 11-07-00140-a, and the Saint Petersburg State University (SPbSU)

II. Related work

Periodic table of visualization methods [23] provides a good top-level diagrams overview for managers. These authors decided that the classification dimensions should be easy to use and have some proven benefits. The organization principles were related to the situation in which the visualization is used (when?), the type of content that is represented (what?) the expected visualization benefits (why?), and the actual visualization format used (how?). As a result, the following five dimensions were suggested:

• *Complexity of Visualization:* Low to High, referring to the number of rules applied for use and/or the number of interdependences of the elements to be visualized.

• *Main Application or Content Area [how?,what?]:* Data, Information, Concept, Metaphor, Strategy, Compound Knowledge.

• *Point of View [when?]:* Detail (highlighting individual Items), Overview (big picture), Detail and Overview (both at the same time).

• *Type of Thinking Aid [why?]:* Convergent (reducing complexity) vs. Divergent (adding complexity).

• *Type of Representation [what?]:* Process (stepwise cyclical in time and/or continuous sequential), Structure (i.e., hierarchy or causal networks)

The authors organized these dimensions in the specific table of visualization methods. But we may conclude that while it is a very impressive result the values for these dimensions are rather general, overlapping and are specified insufficiently.

Lohse et al. [24] reported a structural classification of visual representations. These authors identified 11 major clusters of visual representations: graphs, tables, graphical tables; time charts; networks; structure diagrams; process diagrams; maps; cartograms; icons; pictures. Criteria for classification were represented using 10 anchor-point phrases: spacial-nonspacial; temporal-nontemporal; hard to understand-easy to understand; concrete-abstract; continuous-discrete; attractive-unattractive; emphasize whole-emphasizes parts; numeric-nonnumeric; static structure-dynamic process; convey a lot of information-convey little information. We may conclude that this classification mostly works with structural dimension. Semantic dimension of diagrams is not covered.

Some of the diagramming tools provide its own classifications of the templates. Visio 2010 (http://office.microsoft.com/en-us/visio/) provides the following 8 embedded categories: Business; Engineering; Flowchart; General; Maps and floor plans; Network; Schedule; Software and Database. Visio 2010 Online library (http://visiotoolbox.com/2010/templates.aspx): Application Architecture; Asset Management; Business Analysis; Business; Capacity Planning; Database Planning; Educational; Facilities; Financial; Human Resource templates et al. 25 categories totally. Smart Draw (http://www.smartdraw.com/): Charts: Flowcharts, Project, Org; Education; Engineering; Forms; Mind Maps; Presentations; Timelines; Decision Trees; Cause & Effect Diagrams; Marketing Charts; Strategy & Planning et al. 29 categories totally. Our general conclusion is that Visio embedded categories do not cover all the knowledge types and have rather inconsistent classification criteria. Smart Draw categories are extremely overlapping, have different level of abstraction and also use inconsistent classification criteria.

Also there exist several enterprise architecture based classifications, e.g. Archimate [19], MEMO [10], IBM Enterprise framework or populated Zachman Framework (http://publib.boulder.ibm.com/infocenter/rsysarch/v11/topic/com.ibm.sa.bpr.doc/topics/r_IBM_Enterprise_fmwk.html). But these classifications and frameworks do not include all the types of diagrams used by managers and in general such taxonomies cover mostly IT-oriented diagrams and proprietary diagrams.

We also would like to mention some independent conceptual specifications for the popular business diagrams / visual languages [3, 14]. Unfortunately these descriptions do not involve all the popular business diagrams / visual languages. Also the existing specifications mostly incorporate the area of business processes, while the other areas are insufficiently specified.

III. Methodology And Results

We suggest to use ontology-based specifications for knowledge types and diagrams/visual modeling languages. Alignment between these two specifications will enable managers to choose diagrams for the particular knowledge type. Additionally it will provide opportunity to select the diagram for the specific competency question and for the visualization of the particular ontology view (elements of ontology).

In order to describe informally the knowledge types and to take a step towards the ontology-based specification we suggest to use competency questions technique [16].

Ontology-based knowledge types specification consists of a set of Ontology Design Patterns (ODP) [12]. ODP — a modeling solution to solve a recurrent ontology design problem. It is a template that represents a schema for specific design solutions. An ODP consists of a set of "prototypical" ontology entities that constitute the "abstract form" of a pattern, and of a set of metadata about its use cases, motivations, provenance, the pros and cons of its application, the links to other patterns, etc. Design solutions based on ODPs encode ontology entities that apply, specialize, or instantiate the prototypical entities defined by the schema. Some of the popular ready-made ODPs are represented at http://ontologydesignpatterns.org/. The other ODPs can be extracted from enterprise-related ontologies [4, 9, 27].

The suggested ideas are integrated in the method of business knowledge diagrams classification (Table 1).

Ontology-based diagram specification is based on the ideas of [15], but we suggest to use "light-weight" ontology-based specifications. They do not require the complete ontological model for every diagram, but conceptualize just the core elements of each diagram. The incompleteness of the specifications is justified by the purpose of the specification — the classification and the choice of modeling language.

Alignment between the two ontology-based specifications can be provided by means of ontology mapping/matching techniques and tools [8].

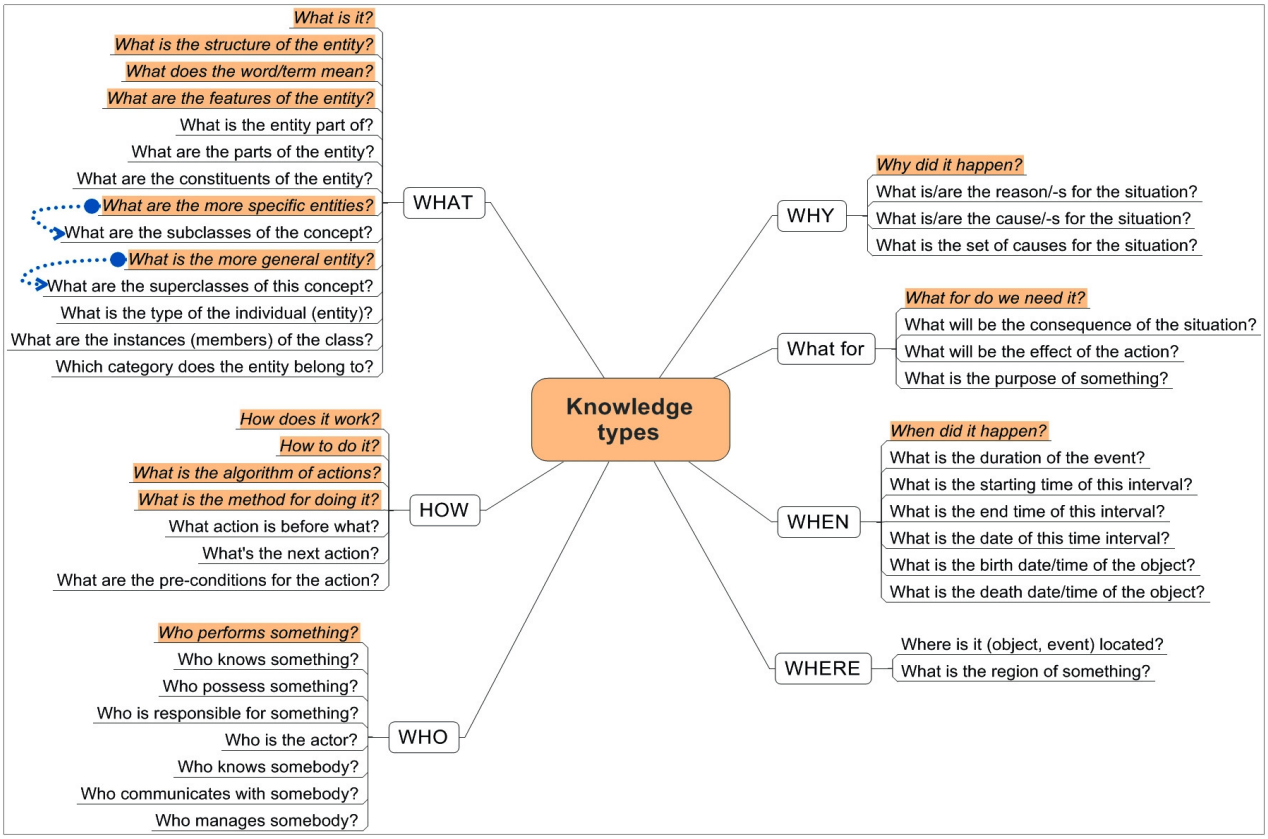
IV. Usage Scenarios

We can introduce three possible scenarios of results usage.

Scenario A (answering the first research question). The user choose the diagrams based on the competency ques-

Table I.
Method of business knowledge diagrams classification

№	Steps	Results
1	Define and describe the knowledge types using competency questions.	Informal description of the knowledge types is represented in Fig. 1.
2	Specify the chosen knowledge types using ODPs — each type of knowledge answering the concrete managerial question may be specified by ontology patterns	ODP specification of knowledge types is based on the Content ODP annotation schema and include the following elements: Pattern name, Intent, Competency questions, Diagram, Elements and examples, Source, Reusable OWL file, Submitted by. The incomplete list of the ODPs for several knowledge types can be found in Table II. Table III shows an example of the ODP specification. Knowledge types descriptions in terms of concepts and relationships can be developed based on the ODP specifications — see Fig. 2.
3	Identify diagrams, which will potentially correspond to the suggested knowledge types, e.g. from Visio, SmartDraw, [23] and provide ontology-based specifications of these diagrams.	Ontology-based specification of diagrams include: diagram name, thumbnail, brief description/purpose, Conceptual model (classes and properties), Conceptual model diagram. Table IV shows an example of the diagram specification.
4	Align ontology-based specifications of knowledge types and diagrams. The alignment is provided using the ontology-based specifications (see steps 2 and 3).	Example alignment between ontology-based specifications of knowledge type and diagram is shown in Table V.
5	Classify diagrams according to knowledge types based on the ODP alignment (from step 4).	The above-proposed approach helps us to work out the classification which may be useful for the practitioners in selecting the appropriate type of business diagram (Fig. 3).



*non-specific competency questions are highlighted (won't be directly relate to ODPs)

Fig. 1. Knowledge types description using competency questions

tions only. If the competency question is non-specific ("voice of the customer") and doesn't directly relates to ODPs, then he/she selects all the diagrams associated with the knowledge type (which is associated with the chosen competency question. The choice among the suggested diagrams is based on the supported ODPs.

Scenario B (answer for the first research question). The advanced user may choose the diagrams using ODPs and the competency questions can be used for preliminary filtering.

Scenario C (answer for the second research question). The user or service wants to represent his/her ontology or ontology view using domain-specific visual language. Then

TABLE II.
THE LIST OF THE ODPS FOR THE KNOWLEDGE TYPES (INCOMPLETE)

Knowledge type	Ontology Design Patterns
WHAT-knowledge	"Part of", "Classification" *
	"Subclass", "Type" **
HOW-knowledge	"Action sequence" (Action + Sequence), "Controlflow" *, "Action pre-condition" (Source: [27])
WHO-knowledge	"Role-task", "AgentRole" *
WHAT-FOR-knowledge	"Help achieve" ODP (Source: [27])
WHEN-knowledge	"TimeInterval", "TimeIndexedSituation" *
WHERE-knowledge	"Place" *

Sources: * - http://ontologydesignpatterns.org/,
** - http://www.w3.org/TR/2004/REC-owl-features-20040210/

Fig. 2. The list of concepts and relationships for the knowledge types

is

TABLE III.
EXAMPLE ODP SPECIFICATION: "PART OF" ODP

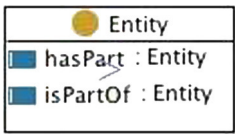
Pattern name: PART OF	
Intent: To represents entities and their parts	
Competency questions: What is this entity part of? What are the parts of this entity?	
Diagram:  	Elements and examples: *Entity* (owl:Class) Anything: real, possible, or imaginary, which some modeller wants to talk about for some purpose. *hasPart* (owl:ObjectProperty) A transitive relation expressing parthood between any entities, e.g. the human body has a brain as part. *isPartOf* (owl:ObjectProperty) A transitive relation expressing parthood between any entities, e.g. brain is a part of the human body. Example: Brain and heart are parts of the human body
Source: http://ontologydesignpatterns.org/wiki/Submissions:PartOf	
Reusable OWL file: http://www.ontologydesignpatterns.org/cp/owl/partof.owl	
Submitted by: ValentinaPresutti	

TABLE IV.
EXAMPLE DIAGRAM SPECIFICATION: ORGANIZATIONAL CHART

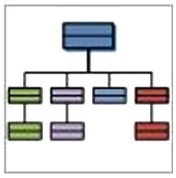
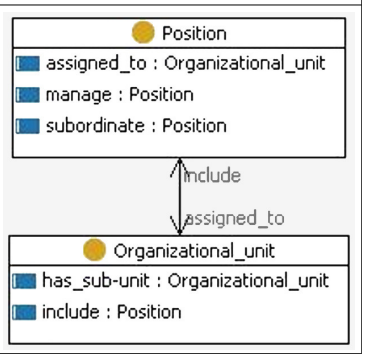
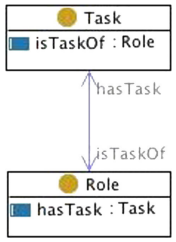
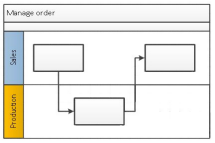
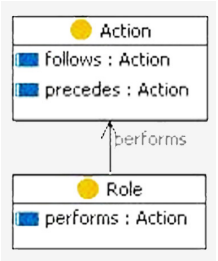
Name, Thumbnail	**Definition**	**Conceptual model**	
		Core elements	**Diagram**
Organizational chart 	A diagram that shows the structure of an organization and the relationships and relative ranks of its parts and positions/jobs.	Organizational unit, Position, Manage (EO) / subordinate relations, include/ assigned to has sub-unit	

TABLE V.
EXAMPLE ALIGNMENT BETWEEN WHO-KNOWLEDGE AND SWIM-LANE DIAGRAM SPECIFICATIONS

Knowledge type	**Competency question/-s**	**ODP**	**Diagram**	**Conceptual model**
WHO	Who performs smth? (informal) What roles are this task (action) of?	"Role task" ODP  	Swim-lane diagram  	

user aligns ontology which must be represented, with ontology-based descriptions of diagrams and then selects the appropriate diagrams for the ontology or ontology view based on the alignment.

V. DISCUSSION AND CONCLUSION

The main novelty of our approach is the mapping between knowledge types and popular business diagram types, which grounded on ontological specifications. Such the mapping together with the suggested informal descriptions of knowledge types can support managers, while working with visual models. Our novel classification is only the attempt as the list of diagrams for knowledge types is incomplete. Creation of the extended catalogue/repository for diagrams should be a collaborative effort. The suggested method of business knowledge diagrams classification can be used within this effort. ODP–based diagram classification method is also a

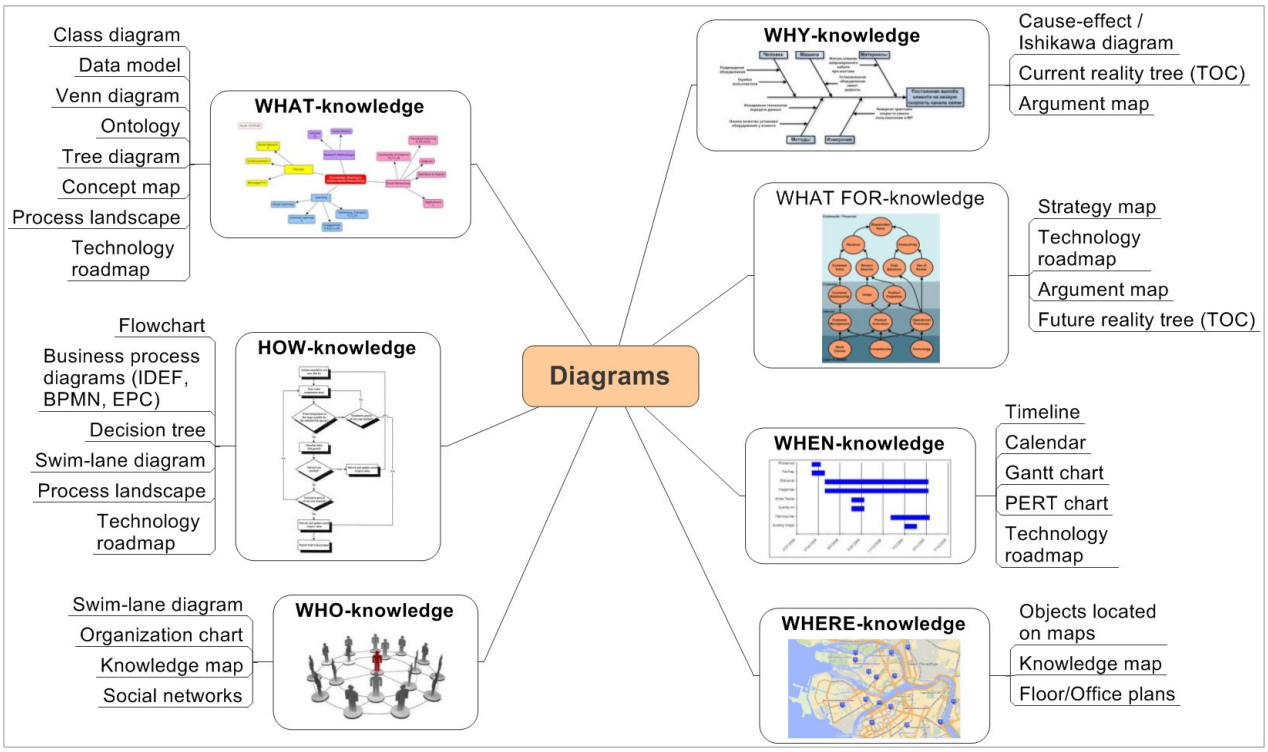


Fig. 3. Diagrams vs. knowledge types

contribution of the paper. Thesaurus based descriptions (synonyms) for ODPs and ontology-based diagram specifications can be a useful appendix (see WordNet). The suggested diagrams can be typically considered as diagram types, which may have a lot of variations and particular notations. We've tried to extract the most generic or prototypical inherent elements of diagram / visual modeling language. Additionally, informal description of knowledge types provides new classification the existing ODPs.

Such pattern-based approach can be considered as the first step towards ontologically founded usage of diagrams among managers. Business diagrams are typically describes some components of enterprise architecture. So according to the "Maturity Model" for Enterprise Architecture Representations [29] adhoc visual models of enterprise architecture correspond to the 1st level of maturity. This approach to enterprise architecture, though a natural, common and easy place to start, does not scale well. Any sizeable organization generally has more than one person or a single group doing enterprise. The ultimate goal is the design of a consistent organizational ontology or ontology network behind a collection of diagrams. This will allow organizations to have ontology-based knowledge repository with consistent domain-specific visual views.

References

[1] Alavi, M., & Leidner, D. (2001). Knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Quarterly, 25(1), 107-136.

[2] Blackwell, A., & Engelhardt, Y. (2002). A Meta-Taxonomy for Diagram Research. In M. Anderson, B. Meyer, & P. Olivier (Eds.), Diagrammatic representation and reasoning (p. 584). Springer.

[3] Cabral, L., Filipowska, A., Grenon, P., Nitzsche, J., Norton, B., Pedrinaci, C., et al. (2009). Process Ontology Stack, Evolved Version, Deliverable 1.5 of the SUPER project.

[4] Dietz, J. L. G. (2006). Enteprise Ontology Theory and Methodology. Springer.

[5] Eisenstadt, M., Domingue, J., Rajan, T., & Motta, E. (1990). Visual knowledge engineering. IEEE Transactions on Software Engineering, 16(10), 1164-1177.

[6] Eppler, M., & Burkhard, R. (2007). Visual representations in knowledge management: framework and cases. Journal of Knowledge Management, 11(4), 112-122.

[7] Eppler, M., & Jianxin, G. (2008). Communicating with Diagrams: How Intuitive and Cross-cultural are Business Graphics? Euro Asia Journal of Management, 18(35), 3-22.

[8] Euzenat, J., & Shvaiko, P. (2007). Ontology matching. Springer.

[9] Filipowska, A., Hepp, M., Kaczmarek, M., & Markovic, I. (2009). Organisational Ontology Framework for Semantic Business Process Management. In W. Abramowicz (Ed.), Business Information Systems (Vol. 21, pp. 1-12). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Frank, U. (1999). MEMO: Visual Languages for enterprise modeling.

[11] Galloway, D. (1994). Mapping work processes (p. 89). ASQ Quality Press.

[12] Gangemi, A., & Presutti, V. (2009). Ontology Design Patterns. In Steffen Staab & Rudi Studer (Eds.), Handbook on Ontologies (pp. 221-243-243). Berlin, Heidelberg: Springer Berlin Heidelberg.

[13] Gavrilova, T., & Voinov, A. (1998). Work in progress: Visual specification of knowledge bases. In A. Pasqual del Pobil, J. Mira, & M. Ali (Eds.), Tasks and Methods in Applied Artificial Intelligence (Vol. 1416, pp. 717-726). Berlin/Heidelberg: Springer-Verlag.

[14] Giannoulis, C., Petit, M., & Zdravkovic, J. (2010). Towards a Unified Business Strategy Language: A Meta-model of Strategy Maps. In P. Bommel, Stijn Hoppenbrouwers, S. Overbeek, Erik Proper, & J. Barjis (Eds.), The Practice of Enterprise Modeling (Vol. 68, pp. 205-216). Berlin, Heidelberg: Springer Berlin Heidelberg.

[15] Guizzardi, G., Pires, L. F., & Sinderen, M. (2006). Ontology-Based Evaluation and Design of Domain-Specific Visual Modeling Languages. In A. G. Nilsson, R. Gustas, W. Wojtkowski, W. G. Wojtkowski, S. Wrycza, & J. Zupančič (Eds.), Advances in Information Systems Development (pp. 217-228). Boston, MA: Springer US.

[16] Gómez-Pérez, A., Suárez de Figueroa Baonza, M. C., & Villazón, B. (2008). NeOn Methodology for Building Ontology Networks: Ontology Specification.
[17] Harel, D., & Rumpe, B. (2000). Modeling Languages: Syntax, Semantics and All That Stuff, Part I: The Basic Stuff.
[18] Hodgkinson, G. P., Maule, A. J., & Bown, N. J. (2004). Causal Cognitive Mapping in the Organizational Strategy Field: A Comparison of Alternative Elicitation Procedures. Organizational Research Methods, 7(1), 3-26.
[19] Jonkers, H., Burren, R. van, Arbab, F., Boer, F. de, Bonsangue, M., Bosma, H., et al. (2003). Towards a language for coherent enterprise architecture descriptions. Seventh IEEE International Enterprise Distributed Object Computing Conference, 2003. Proceedings. (pp. 28-37). IEEE Comput. Soc.
[20] Karagiannis, D., & Kühn, H. (2002). Metamodelling Platforms. Proceedings of the Third International Conference on E-Commerce and Web Technologies (p. 182–). London: Springer-Verlag.
[21] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., & Giannopoulou, E. (2007). Ontology visualization methods - a survey. ACM Computing Surveys, 39(4), 10-es.
[22] Kudryavtsev, D., & Grigoriev, L. (2011). The technology for the ontology-based business architecture engineering. Accepted paper for The 10th International Conference on Intelligent Software Methodologies, Tools and Techniques, September 28-30, 2011.
[23] Lengler, R., & Eppler, M. (2007). Towards a Periodic Table of Visualization Methods for Management. Proc. of the Conference on Graphics and Visualization in Engineering, 2007 (pp. 1-6).
[24] Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. Communications of the ACM, 37(12), 36-49.
[25] Meyer, J. (1997). The acceptance of visual information in management. Information & Management, 32(6), 275-287. Retrieved May 7, 2011, from http://dx.doi.org/10.1016/S0378-7206(97)00032-3.
[26] Plexousakis (Eds.), D. (2009). Deliverable 3.1 Next Generation Modelling Methodology, plugIT project (p. 87).
[27] [27] Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1998). The Enterprise Ontology. The Knowledge Engineering Review, 13(1), 31-89. Retrieved March 14, 2011, from http://portal.acm.org/citation.cfm?id=976223.976226.
[28] Zachman, J. (2003). The Zachman Framework for Enterprise Architecture: A Primer for Enterprise Engineering and Manufacturing.
[29] Polikoff, I., & Coyne, R. F. (2005). Towards executable enterprise models: Ontology and semantic web meet enterprise architecture. Journal of Enterprise Architecture, Fawcette Publications. Retrieved May 7, 2013, http://www.topquadrant.com/docs/whitepapers/WP-BuildingSemanticEASolutions-withTopBraid.pdf

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1255–1260

Knowledge Management as Foundation of Smart University

Mieczysław Owoc
Wroclaw University of Economics Komandorska
118/120, 53-345 Wrocław, Poland
Email: mieczyslaw.owoc@ue.wroc.pl

Katarzyna Marciniak
Wroclaw University of Economics Komandorska
118/120, 53-345 Wrocław, Poland
Email: katarzyna.marciniak@ue.wroc.pl

***Abstract*—Functioning in an era of knowledge is forcing organizations to manage this valuable resource in exact way. Very frequently activities of organizations are dependent on the application of knowledge, sometimes even means "to be or not to be" for enterprise. Nowadays, to fulfill business goals of institutions it becomes fundamental for them to use intelligent systems supporting comprehensive management of the organization. Such support allows to increase efficiency and better effectiveness of the running businesses. As we are living in age of international integration, where world economy is tending to reach type of knowledge-based economy (KBE), universities are forced also to change way of their functioning. It is important for modern universities to be not only education centers but mainly the successfully prospering organizations-based-on-knowledge. Such approach is going to provide higher competitiveness of particular institution and will make its functioning more useful for economy of the region. Implementing a comprehensive and intelligent IT solution within a university and providing educational services, which are personalized to the needs of the market, will allow universities to reach a type of institution called "smart". The aim of the paper is explanation why university centers should evolve in a type of institution based on knowledge. The paper is managed as follows. After short introduction concerning research context the discussed concepts of Knowledge Management and Smart Universities are presented. In the main section real examples of Knowledge Management Systems implementation and examples of Smart Universities are investigated in order to identify and describe roles of Knowledge Management Systems in this area. It allows for formulation conclusions on intersection of two investigated approaches.**

I. Introduction

IT IS noticeable, that particular word countries are realizing implementation process of knowledge-based economy concept within its own economies. Such step is treated from scientific point of view as a fundamental step for the implementation and maintenance of balanced development of the country. Today, where the management is based on knowledge, it is crucial to use new technologies as a support of decision making process in organization. Currently main factors of entrepreneurship are: highly qualified human capital, universities, along with research centres, IT infrastructure and legal environment conducive to the construction and development of selected sectors of the economy. Accordingly, key elements of the knowledge-based economy in the modern management model becomes important to generate rapid pace of intangible capital, the implementation of innovation in every possible aspect of the business environment and the generation of high-quality knowledge.[24]

Terms of the knowledge economy is crucial for the functioning of businesses. Even Peter Drucker - recognized in the world as the father of modern concepts of management - management expert said that traditional enterprise resources such as land, labour, capital, constitute a bigger barrier than the driving force behind the development of the company. According to Drucker, a mission-critical determinant of creativity in all aspects of the knowledge.[8] Following Drucker, knowledge is a key factor determining the existence of the company and gives the public a new, unique character.[22] In addition, it is important to pay attention to the fact that it is impossible to say about the delivery of the knowledge-based economy, while acting within state organizations - including the institutions do not operate on the basis of the knowledge management process. Hence adopted to say that organizations which take into account aspects related to the management of knowledge management are called knowledge-based businesses. For the purposes of this article, for such an undertaking means an organization that "actively create knowledge and can use it in their daily activities"[14]. In addition, it is needed to note that the creation of knowledge is not only a task of companies involved in the provision of purely market knowledge but should manifest itself in the functioning of each operator - which in turn implies KBE.

How organizations generate and manage knowledge, which information systems are used to support such process, why universities can be classified as a corporation and for what reason authors tried to nominate them as a "smart" is going to be describe in this paper. Article will therefore include the characteristics of the company knowledge, a description of the key concepts of knowledge management, including meaning of information in order to present the essence of having intelligent solutions within the organization. In next sections it will be also explained why universities should be seen as the company providing the market "creative people" (called "smart people"), through the provision of educational services, functioning on the basis of intelligent information solutions.

978-83-60810-53-8/$25.00

II. Company knowledge

Knowledge is not only one of the key resources of the enterprise, but also is the foundation, the starting point for determining the company's strategy, particularly for the implementation of management information systems. Knowledge of the computer science is defined, in part, dependent on the data and information. The data does set of facts, measurements, statistics. Information, in turn, is nothing more than structured data. Thus, in this sense, knowledge is a collection of information that can be used in practice. However, the organization cannot exist without human capital, so to be able to say that the company has the knowledge, it is needed to take into account in addition to possession of selected information skills, experience and qualifications of specialists in selected fields. The combination of these two elements is a complete understanding of each company. Pay attention to the continuity of the process of converting data into information and information into knowledge. [2]

Because of the persistence of the two currents are still classifying knowledge: as an information resource and as an element of human capital. Each company needs both of these resources to function properly, so it is reasonable to say that knowledge is nothing else than "a combination of everything: facts, phenomena and relationships between them, which is consciously perceived and recorded (in any way saved as real entities or conceptual) and can give to others, according to the intention of having knowledge in specific conditions and circumstances to arouse certain behaviours". [6]

This is how the company manages knowledge depends on how knowledge in the enterprise is defined and what type it is preferred to carry out their activities. In theory, knowledge management is distinguished by personalized knowledge, which consists of the explicit knowledge and discreet. The formal knowledge are recognized as a formal document prosperity of the organization. However, informal knowledge - discreet, is nothing else but the skills, qualifications and experience of the employees. This knowledge is nowhere recorded, is set in their minds, so there is a barrier to capture the sensitive knowledge of that. [19] However, there is no longer a barrier to overcome. The knowledge of a discreet, distinguished by the so-called core, which allows the conversion of that knowledge in a standardized, codified form and vice versa. Due to the prevailing organizational culture within the organization and information system properly prepared can be personalized to gather knowledge and to process it and share as needed to other members of the organization. Although, it is depended on employee what part of knowledge is he/she going to "give" the company, it is also possible by using suitably generated mechanisms of organizational learning. Supporting such process, company is able to obtain a well-established knowledge that remains within its borders regardless of the availability of human capital.[16]

III. Knowledge Management

Being aware of all possible kinds of company's knowledge resources and knowing value of this crucial factor, it automatically refers to the essence of knowledge management, treated as a priority in the strategic management techniques. Knowledge management, as a young field of management encompasses the latest methods and techniques that are designed to provide most spectacular use of knowledge.[15] To make it more detailed, "knowledge management system is a modern concept, involving the effective use of knowledge and transforming the company into a lasting value for customers and employees of the organization"[20]. Moving forward, "knowledge management is clearly defined and systematic management of vital knowledge for the organization and its associated processes of creating, gathering, organizing, diffusion, use and exploitation of knowledge, carried out in pursuit of the objectives of the organization"[19]. Knowledge management can also be treated as a specially designed "system that helps organizations to acquire, analyse the use (re-use) of knowledge in order to make faster, smarter and better decisions, so that they can achieve a competitive advantage"[10]. In order to obtain a complete picture of knowledge management need to mention two aspects. The first talking about the fact that knowledge management is "management of information, knowledge and expertise available within the organization, i.e. the creation, collection, storage, sharing and use, to ensure the organization's future development of existing resources"[12]. Second, where knowledge management is regarded as a "deliberate business strategy, which selects, distils, stores, organizes, packs and provides information relevant to the company's business in a way that improves staff efficiency and competitiveness"[5].

As we can see, authors of the above definition also emphasizes the two items related to knowledge management. Both accessible to, the information, experience, staff and their expertise and the technological, where the focus is on codifying knowledge, its acquisition, collection, analysis, storing and sharing at any time, by a specific user. The logical also is the fact that the development of knowledge takes place through the exchange of experiences, analysis, opinion, finding new sources of information, where the information systems are the basis to allow all of the actions.

IV. Knowledge management tools

As mentioned earlier, knowledge management is to create added value for the company and its environment proximal and distal. Due to the ongoing intensive technological development has now become a common use of communication tools allowing the use of accumulated knowledge and its proper share respectively of all company employees. According to prevailing quality standards for functional knowledge management systems, it is desirable that the software is compatible with other environments that

utilizes the company or its partners in the supply chain. It is true that currently supplied management information systems do not focus on the separation of the component dedicated to the management of knowledge, but rather allow it to manage in the "background", without limiting in any way the access to knowledge or to give support key company processes.

Due to integrated information system, it should be understood as a modularly organized system, supporting all areas of its business, from marketing and planning and procurement, through the technical preparation of production and the control, distribution, sale, management of repair work financially - accounting and management human resources. [1]

The integrated system is a system in which data or information is entered only once, and they are available to all users of the various processes in the enterprise. The most common, the most distinguished units in integrated systems are:

- an organizational chart - taking into account the structure: vertical and horizontal - user login to the system,
- input - screen formats for data entry,
- order - the products / services - short / long-term, temporary of contractors - implementation schedules - time and value of the contracts short / long term of complaints - the frequency,
- planning - sales of short / long-term production capacity, about forecasting in demand, prices, the scheduling of purchases - material procurement automation, the distribution needs - transport, forwarding
- production / services - technological resources, location, staff, materials, semi-finished products
- staff - directory management employees, managers, employees, executive, administrative, a database of periodical co-workers,
- materials management - information on materials (eg, prices)a database of suppliers, about purchase history, the timing of purchases
- fixed assets: production by division, assets, condition of assets,
- control - of finishing orders of jobs, areas of control, the amount of control about the control points, reports,
- book sales - a record audience, issued documents: the amounts received and receivable of sales analysis, implementation plans, short / long term
- module obligations - register suppliers of necessary goods and their normal consumption of goods purchased,
- master register - maintaining the organization's financial data, simulations of possible states
- module financial - accounting - analysis of energy and raw materials, wages, load workstations of finance - budget, bank, payment procedure. [11]

Carefully selected an integrated management information system now enables efficient management of the entire organization, carrying out the functions of the enterprise environment taking into account both internal and external, and therefore, appropriate management of enterprise knowledge. "The selection of specific solutions and technologies in the field of knowledge management should be based on the specific nature of the company, its profile, personal economic situation, its strategy and approach to knowledge management. Any company or institution should be considered as characteristic for the organization of the organizational culture, creative workers, rules and norms prevailing within it and look at it from the perspective of ongoing business processes. In addition, each organization must be aware that a comprehensive knowledge management system cannot be based only on an appropriately selected technology"[19].

V. UNIVERSITY=ORGANIZATION

To be able to say that higher education institutions can be considered for the organization, in terms of management, one must start from the definition of the firm. Companies in Polish law is defined as "an organized set of tangible and intangible assets intended for business" [7]. Institution (according to management sciences), as defined by T. Pszczołowski is the "organization, which is a team of people interacting with each other and properly resourced" [17]. Continuing, J. Zieleniewski states that the term "institution" is a social creation, called. "Thing organized," which factors are creative human capital and tooling needed. [21] In the literature it is also possible to find the phenomenon, where the concept of the organization is used interchangeably with the term institution.[4] It can be concluded, that the concept of institutions is synonymous with the concept of the organization in terms of factual. Knowing that Polish universities are institutions, we can consider the above classification that should be treated as an organization that can be managed in a comprehensive manner.

Taking into account the objectives of Economy Based of Knowledge and National Development Strategy, as well as other documents and regulations speak of the need to develop and adapt to the new market situation, taking into account the assumptions of the concept of sustainable development, it can be concluded about the need to implement appropriate IT tools within the "forward-thinking organization". As colleges and universities that provide comprehensive solution of the organization, have their own resources, a strict hierarchy and organizational structure, culture, and provide specific services, they are forced to face this present demographic projections take steps to ensure the survival of a heavier period and in a short time to adapt to new trends in the

labour market, in order to refocus the education directly under the existing demand.

Knowing that every company is made up of individual stocks, processes and the environment in which it operates, it is necessary that each learning organization, have implemented an integrated information system. Information systems are nowadays an integral part of most organizations. As previously shown, information systems affect the basic structure and design of the organization. The main purpose of information systems is to satisfy all existing information needs generated by the organization in order to be able to make the right decisions for appropriate decision-making bodies.[3] It is also known that the larger the organization, the knowledge management system using the information becomes more difficult. However, due to technical capabilities, integrated management systems support mainly horizontal communication, without compromising on the priority of vertical communication. Properly designed information system should also allow for quick and easy communication with the environment of the organization, such as customers, and in the case of higher education such as students.

Why universities should have within its structure an integrated information system? The answer is very simple. Universities are being developed organization providing educational services primarily with the use of highly skilled human capital, the use of modern technologies, and other essential resources needed to carry out business activities, providing the market of people defined by psychologists as "creative". Because the university is the responsible of the obligation to seek to achieve the objectives of the KBE and the National Development Strategy, it is necessary to establish their business as a company - put an integrated information system that will not only improve operational efficiency by almost immediately verify the responsibility of individual positions, structured different levels of management, but also through better organization of business processes occurring inside and outside the university, along with the improvement of communication between internal customers and external institutions.

What is an integrated information system? It is nothing other than a comprehensive system designed to optimize business processes both inside and outside the company, using tools to automate the exchange of information between clients throughout the logistics chain. Considered as the most effective integrated information systems are Enterprise Resource Planning systems (ERP). ERP represent a group of integrated computer systems, such as modular structured enterprise information systems. They gather in one coherent system of all the traditional functions of management (related to financial and management accounting, finance, human resources and payroll, technical preparation of production and its control, procurement, warehouse management, planning and execution of sales and logistics, quality management)"[23]. Their main objective is to integrate internal and external environment of organization. This process is supported by the latest information technology solutions, such as multidimensional data analysis in data warehouses. The modular ERP also provides comprehensive knowledge management in organization. Confirmation of this view can be found even in the same characteristic features of ERP systems:[11]

- functional complexity - covers all spheres of business activity,
- integration of data and processes - including the exchange of data and information between modules, internal organization and its environment (EDI),
- functional and structural flexibility - is designed to provide maximum adaptability, customization and hardware solutions - software to the needs of the individual modules,
- open-enables you to extend the system with another, new modules
- substantive progress - provides full support for the processes of information - decision-making using the full data extraction and aggregation, as well as practical support system for logistics strategies such as JIT, MRP II and TQM
- technological advancement - ensures compliance with current standards of hardware - the programming with the ability to migrate to new platforms,
- compliance with law such as the law on accounting.

VI. Smart University

Based on the assumptions, known in the world concepts of smart city, smart business, or even smart building, it is possible to say that colleges and universities can also get a smart domain name. Based on the assumptions of the Research Group of IBM Specialists, which attempts to outline the concept of smart institution as a smart city, where it is defined as "the integration of infrastructure: physical, social, business and IT"[9], by analogy we can try to concretize basis of smart university. Figure 1 shows five segments, which authors assume as relevant to the concept of smart.

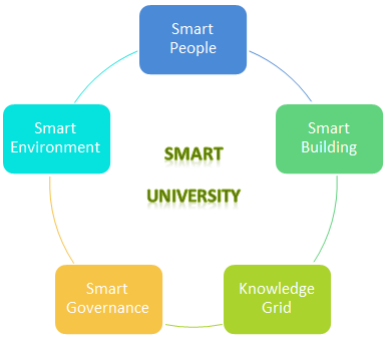


Figure 1. Components of smart university.

To be able to say that the university achieves type of smart, it must be managed in an intelligent way. It means, the authorities during making decision have to include all

five university contractual determinants: human and social capital (called smart people), available physical infrastructure (called smart building), an integrated information infrastructure (called knowledge grid), strategic decision-making processes (called smart governance) and aspects related to the protection of the environment (called smart environment). In order to manage institution according to such approach, decision-makers must consider institution as a whole, as a single organism. Therefore authorities of universities should be aware that making changes in one specified area has influence in second one. Considering such obvious mechanism it becoming needed to manage all segments of the university in a coordinated manner through the use of available information solutions enabling consolidation of knowledge and efficient management. Hence the necessity of implementation properly prepared, dedicated, fully-integrated management information system enhanced by intelligent modules, allowing to carry out a number of advanced business analytics. Computerization of communication between university-customers will help to facilitate the collection of information and improve communication between system users. In addition, the accumulation of large amounts of information, both in the form of reports, studies, statistics, or in the form of answers to questions, contribute to the development of the wisdom of the institution. Operation of university decision-making bodies in terms of sufficient information can therefore be involved in creating the best conditions for the development of the institution, as well as serve the needs of the labour market and the customers themselves universities. Improving the quality of services provided by universities can be guaranteed only taking into consideration the resource information from its environment. Therefore, it is important that the participants had the opportunity to free, independent, re-use of public resources, tools and information.

In addition, smart university should have:

- complex interaction (called comprehensive contact), including intelligent management of resources, equipment and utilities, allowing to specify the location of objects in real time using ICT infrastructure
- full integration (called fully-integrated), assuming that the basic scheme of the system infrastructure supporting university heterogeneous data are fully integrated together,
- incentive for innovation (called encouragement for innovation), covering all activities of the university intended mainly for its interior and exterior public institutions in order to spread the use of its new technology as a way to ensure the development of their own, as well as regional
- group work (called collaborative operation), based on intelligent infrastructure, critical systems and cooperating users - employees of the university, both administrative and scientific, which will help to improve the efficiency of the university.

A perfect example of higher education in Poland, which aims to provide smart can be a School of Social Sciences (pl. SWPS), based in Warsaw, together with affiliated divisions in Poznan, Wroclaw, Katowice, Sopot have deployed dedicated to the needs of the institution integrated management information system ERP SIMPLE.EDU [25]. The use of a comprehensive IT solution by School of Social Sciences allowed for more efficient management of the institution and the proper management of its resources, mainly knowledge organization.

VII. Conclusion

The discussion conducted in the pages of this article can be concluded that the company's success is conditioned by the precise knowledge management company. It also means that organizational success depends on access to high-quality information, appropriately implemented IT solutions and business culture of institutions. “Managing organization requires efficient management of knowledge and human capital, treated as assets that are purchased, maintains, develops, evaluates and monitors. To enable the organization to be fully competitive in the global and local market should meet two conditions: have adequate knowledge and be able to take advantage of their knowledge"[13]. As can be seen in the pages of this article, knowledge is a sensitive factor in determining business value.

Additionally, caring about knowledge management at every level of organization enables creating opportunities of organization growth. It allows also to implement innovation, conduct studies on the effectiveness of organization considering processes inside and outside the organization. Such strategy may improve functioning of the supply chain, in which the institution operates. Also, implementation of customized information system supporting managing allows improving communication within the organization. Ensuring integration within organization supported by a comprehensive intelligent information system management is the starting point for achieving the objectives contained in the concepts such as Sustainable Development Strategy and Knowledge-Based Economy. Meeting the objectives of computerization of enterprises, which is crucial for the use of highly skilled human capital, business management in a holistic way, where decisions to solve one problem take into account potential changes in other aspects of company is none other than the foundation of modern management concepts, where the business seeks to type named “smart”.

References

[1] Adamczewski P., Zintegrowane Systemy Informatyczne w Praktyce, MIKOM, Warszawa 2004

[2] Baltzan P., Phillips A., Business Driven Information Systems, second edition, McGraw-Hill Irwin, New York 2009

[3] Banaszak Z., Kłos S., Mleczko J., Zintegrowane systemy zarządzania, Polskie Wydawnictwo Ekonomiczne, Warszawa 2011

[4] Bednarski A., Zarys teorii organizacji i zarządzania, Dom Organizatora TNOiK, Toruń 2001,
[5] Bergeron B.m Essentials of Knowledge Management, John Wiley & Sons, New Jersey 2003
[6] Błaszczuk A., Brdulak J.J., Guzik M., Pawluczuk A., Zarządzanie wiedzą w polskich przedsiębiorstwach,Szkoła Główna Handlowa, Warszawa 2004
[7] Buczna M. (red.), Kodeks cywilny, art. 55, stan prawny na 1 września 2007
[8] Grudzewski W.M., Hejduk I, Zarządzanie wiedzą w organizacjach, E-mentor Dwumiesięcznik Szkoły Głównej Handlowej w Warszawie, Warszawa 2005, nr 1 (8) 2005
[9] Harrison, C.; Eckman, B.; Hamilton, R.; Hartswick, P.; Kalagnanam, J.; Paraszczak, J.; Williams, P., Foundations for Smarter Cities, IBM Journal of Research and Development , 2010, 54 (4)
[10] Jakubczyc J., Mercier-Laurent E., Owoc M.L : What is Knowledge Management? Materiały konferencyjne "Pozyskiwanie wiedzy z baz danych", Szklarska Poręba 1999 , Baborski A. (red.). Prace Naukowe AE Wrocław nr 815, Wrocław 1999Jaszkiewicz A., Inżynieria oprogramowania, Helion, 1997
[11] Karamalla-Gaiballa E., Matouk K., Zarządzanie wedzą w przedsiębiorstwie a jego potencjał ludzki: http://www.swo.ae.katowice.pl/_pdf/352.pdf [2013-05-15]
[12] Kisielnicki J., Kierunki i tendencje zastosowań informatyki we współczesnym zarządzaniu, [w:] Materiały seminaryjne z konferencji „Polskie autorytety naukowe o komputerowych systemach wspomagania zarządzania", Centrum Promocji Informatyki, Warszawa 17.06.2004
[13] Klincewicz K., Cele zarządzania wiedzą, Zarządzanie wiedzą, Wydawnictwo Akademickie i Profesjonalne, Warszawa 2008
[14] Kotwica A., Owoc M.L.: Knowledge Organizing for Small and Medium Size Enterprises. Towards Nature Imitation. World Computer Congress, Toulouse 2004
[15] Mikuła B., Geneza, przesłanki i istota zarządzania wiedzą, w: Zarządzanie wiedzą w przedsiębiorstwie, red. K. Perechuda, Wydawnictwo Naukowe PWN, Warszawa 2005
[16] Pszczołowski T., Mała encyklopedia prakseologii i teorii organizacji, Ossolineum, Wrocław 1978
[17] Skyrme D.J., Knowledge Networking. Creating the Collaborative Enterprise, Butter-worth-Heinemann, Oxford 1999
[18] Turban E., Leidner D., Mclean E., Wetherbe J., Information Technology for Management Transofrming Organizations in Digital Economy, 6th edition, John Wiley & Sons, 2008
[19] Wrycz S. (red.), Informatyka Ekonomiczna Podręcznik Akademicki, Polskie Wydawnictwo Ekonomiczne, Warszawa 2010
[20] Ziencik P., Wiedza w przedsiębiorstwie, „Ekonomika I Organizacja Przedsiębiorstw" 2003, nr 3
[21] Zieleniewski J., Organizacja zespołów ludzkich, PWN, Warszawa 1976
[22] http://decyzje-it.pl/centrum-wiedzy/erp.html#definicja [2013-05-20]
[23] http://www.europejskiportal.eu/id03.html [2013-04-25]
[24] http://www.systemyerp.com.pl/home/1059-swps-wybra a-system-erp-dla-uczelni.html [2013-03-01]

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1261–1267

Scalable Web Monitoring System

Andrzej Opalinski, Wojciech Turek and Krzysztof Cetnarowicz
AGH University of Science and Technology
Krakow, Poland
Emails: andrzej.opalinski@agh.edu.pl, wojciech.turek@agh.edu.pl, cetnar@agh.edu.pl

***Abstract*—Publicly available Web search engines suffer from several limitations, which significantly reduce usability in particular cases. The most important limitations are out-of-date information, very simple query language and limited number of results. In many cases, users of the Internet are interested in finding new information which appear in the particular Web portal. In this paper, a system for monitoring of Web sites is presented. The system can continuously analyze the content of specified Web pages using advanced text processing algorithms. It actively notifies the user when required information is found in newly-added content. It can be deployed on a single PC as well as on a cluster of computers, providing good scalability. The paper presents an abstract architecture of the system, details of the implementation and real-life experiments results.**

I. Introduction and related works

THE GROWTH of the World Wide Web, which has been observed over last years, has resulted in the greatest base of electronic data. It is hard to even estimate real size of the Web. The WorldWideWebSize.com portal claims that the most popular search services index more than 50 billion Web pages [1]. Four years ago Google published an information, that the indexer found 1 trillion unique addresses [3]. These estimates definitely do not show the real size of the Web because the indexers deliberately ignore particular fragments, like content generators, link farms or pages with illegal content.

The features of the Web pose huge challenges for searching systems. The size itself creates significant scalability and performance issues. What is more, it is very hard to acquire information about what a user is really looking for and detect pages containing information needed by the user.

Publicly available Web search services offer access to very simple and fast ways of finding pages. The services use Web crawlers to visit as many pages as possible and build an inverted index of all processed content. The indexes make it possible to find Web pages which contain specified words in few milliseconds. This method of finding information in the Web is used every day by each Web user. However, several significant drawbacks and limitations of the approach do exist:

- If several pages contain all specified words, ordering of results is imposed by the search engine. Sorting is typically based on popularity. This feature connected with the limit in the number of found pages results in inability of finding some pages.
- The query language is typically very simple. It is impossible to express advanced patterns concerning sentences or use synonyms. It is even impossible to specify rules specifying letters casing, distance between words or words ordering.
- Low frequency of crawling causes outdated results. The searchers often find pages which contain different content or no longer exist.
- High popularity of search engines results in an interesting feature of the Web: if a page cannot be found using search services it is considered nonexistent.

These limitations encourage researchers to continue work on different ways of finding valuable information in the Web. The subject of focused crawling has received significant attention over last few years. The idea of a crawler which can select pages relevant to a specified topic [7] has been implemented using various techniques [8], [10]. Most obvious application of a focused crawler is a topic-specific search service, which can provide more accurate results.

Use of index-based search engines can successfully direct a user to potentially interesting Web sites. However when the content of the Web sites changes fast and the information must be detected as soon as possible after it is published, indexing-based methods becomes insufficient. When a user knows where to look for results, but it is impossible to watch the Web sites continuously, a different approach to the problem of searching the Web is needed.

Some complex solutions in this area were also presented. Liu et al. [4][5] proposed a system that monitors web resources and reports changes of web content by sending messages for system's users. Although, those solution does not focus on various methods of information pattern detection, as it's presented in this article. The another system - WebMon [9] - is also a tool for web information monitoring, and could be applied to monitor date, keywords or links. It's intended for multiuser access and provides a useful functionalities, but it also doesn't support various pattern detection mechanism. There are also publicly available solutions as Corona tool [6], which is easily scalable decentralized system available for multiple subscribers, but it's detection pattern flexibility is limited.

In this paper a system for monitoring selected fragments of the Web is presented. It provides a service, which can monitor precisely specified fragments of the Web and actively report when a particular pattern is found in a newly-published content. The various pattern detection methods were tested and compared. Also crawl performance and pattern detection is

978-83-60810-53-8/$25.00

tested and presented, in comparison to standard Google search results. Proposed system could be deployed on a PC, server or cluster - based architecture, to fulfill required performance. Possible applications of the system include monitoring auctions services or job advertisements. It can also be used by law enforcement services for detecting illegal content quickly.

II. Architecture of the Web Monitoring System

The architecture of the system is inspired by a Java-based general purpose Web crawler with indexer presented in [11], [12]. The crawler uses a cluster of computers for parallel processing of different Web sites. It provides a distributed inverted index of all words found on visited pages. The general architecture of the Web monitoring system is presented in a Figure 1.

The Crawler component is a single processing thread. It contains a queue of URLs to download and analyze. It is responsible for performing all operations needed to process a Web Page – details on processing algorithms will be presented in the next section. The most important result of the processing is URLs detection – the URLs are returned to the Smith Component.

The Smith component controls multiple Crawler threads. It starts specified number of Crawlers, manages URLs queues, receives found URLs and communicates with the Node Manager.

Each node used by a system has a single Node Manager is responsible for communication with global System Manager. Each Node Manager provides administration interface for monitoring and management. It also provides a service interface, which is used for executing search queries.

The System Manager is responsible for controlling nodes. It collects and distributes found URLs, performs distributed search and provides access to management interface of every node. It also provides a Web Service interface for clients of the system.

The Client application uses provided Web Service interface. It is implemented in a different technology and provides convenient graphical user interface.

The system can be deployed in three different ways:

1) PC-based,
2) server-based,
3) cluster-based.

The smallest configuration can be executed on a single modern PC. In this configuration all components are running on the same computer. This configuration uses particular settings, which significantly limit required system resources. It does not need expensive hardware however, it can be used only for monitoring several small Web sites.

In the configuration using a single server, the MySQL database server and the JBoss application server are executed on a powerful machine, which is working constantly. User application can be started from time to time in order to verify searching progress. In this configuration several users can use the same server. Tests showed that a single server can process around 100 000 Web pages every hour.

The most advanced configuration uses a cluster of servers. The performance of processing in this configuration can be easily increased by adding new servers to the cluster.

III. Resources Processing Algorithm

The most important part of the system is implemented by the Crawler component. It performs processing of Web pages content downloaded from the Internet. A diagram of steps performed by the Crawler is shown in a Figure 2.

The process of crawling is controlled by the Manager, which stores a queue of URLs to process. All URLs found by the node are stored in the Urls database. The Manager continuously executes the processing sequence, which consists of the following steps:

- Resource downloading, which results in HTML source stored in a memory buffer. This step includes filtering unsupported file formats, HTTP servers error handling and maximum source length verification.
- HTML parsing by the Lexer. This is the most complex and time-consuming step of the processing which builds document model.
- Changes detection, which results in selecting fragments of a Web page content that have never been processed.
- Content processing by various plugins operating on the document model created by the Lexer. One of the plugins returns a list of URLs found in the processed content, that are added to the queue of the Manager.

This sequence is being executed by every single URL which appears on the list of the Manager. The following sections provide more details on the processing algorithms.

A. Content Parsing and Resource Model Building

The Lexer converts the HTML source into a resource model. The model represents a tree structure built of segments. Each segment represents a selected structural element of the Web page (tables, paragraphs and lists). Segments can contain other segments or they can constitute leafs of the tree containing lists of words, special characters and HTML tags.

Each element of the HTML source is converted to an element of the tree structure or to a token. There are three basic types of tokens:

1) words,
2) tags,
3) special characters.

Each token has its unique identifier – an eight byte integer. Selected ranges of the identifiers are reserved for tags and special characters. The rest is being dynamically assigned to new words found in the content. This approach converts the content of each leaf segment into a list of identifiers, making following processing very efficient.

The dictionary of words is a very large data structure. Average Web page contains several thousand words, however typically very few are new words. Nevertheless the size of

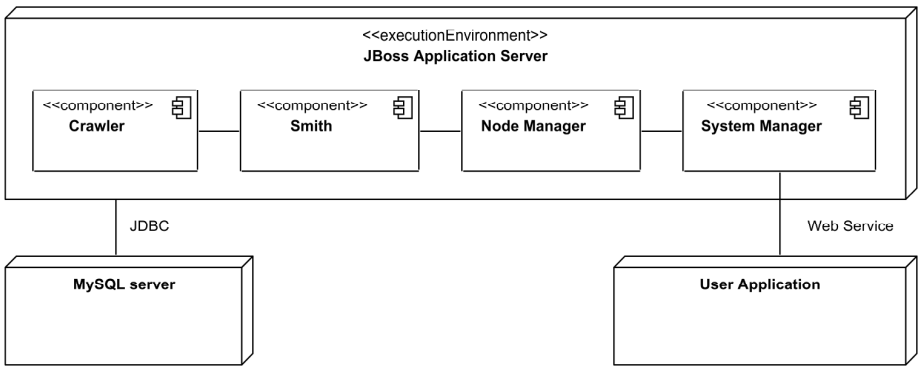


Fig. 1. Abstract architecture of the Web Monitoring System

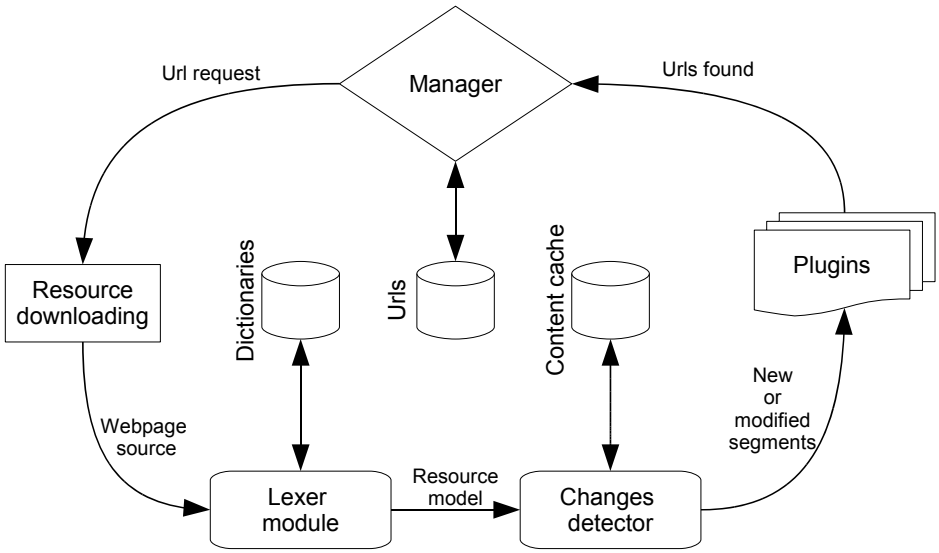


Fig. 2. Processing performed by a single Crawler component.

the words dictionary can reach millions of entries after a few days of crawling. Therefore, the implementation uses large in-memory caches based on hash maps to make the word-to-identifier conversion as fast as possible.

B. Changes Detection Algorithm

The changes detection algorithm is based on hash codes calculated for analyzed content. The hash code for a segment seg containing n tokens is calculated using tokens' ids in the following way:

$$hash(seg) = \sum_{i=0}^{n-1} seg[i] \cdot 31^{n-1-i} \quad (1)$$

where $seg[i]$ is the identifier of the i^{th} token in the segment. The algorithm is very similar to the one used by Java String class implementation.

The hash codes are calculated for every leaf segment. If the hash code has been found in any previous processing of the same Web page, the segment is considered unchanged and is not processed any further. Theoretically, two different segments could have the same hash code, however, using 64 bit identifiers and 64 bit hash codes makes it almost impossible.

To determine what values of hash codes have been already processed, the Content cache database is used. It stores all hash codes of leaf segments found in a page content. Typical Web page contains between 10 and 100 leaf segments.

C. Content Processing

Leaf segments that are considered new or modified, are processed by all enabled plugins. A plugin is a component which provides a common interface – it accepts resource model or its parts.

There is one plugin which is mandatory for proper functioning of the system. The URL detector plugin must be enabled to continue crawling process. It finds URLs in the content of provided segments, searching for anchor HTML tags.

The Web Monitoring System provides several other plug-ins that are used for finding Web pages containing patterns specified by a user. The plugins provide several methods for defining the patterns.

a) List of words: – a user provides a list of words in particular form. A segment will match the pattern if all words are present in the segment.

b) List of stems: – a user provides a list of words in any form. A segment will match the pattern if any form of each word is present in the segment. To implement the functionality a plugin uses a stemmer component which can convert any word to its basic form and provide all possible forms for a given word. The stemmer typically requires a language-specific dictionary of all words and possible forms.

c) List of close words: – a user provides a list of words and maximum distance (in words) between all the words. A segment will match the pattern if all words are present in the segment and the distance between the most distant words is less than the value provided. The order of words in the list is ignored. This plugin can also use stems instead of words.

d) List of words in a sentence: – a user provides a list of words. A segment will match the pattern if all words are present in a single sentence in analyzed the segment. This plugin can also use stems instead of words.

e) List of optional words: – a user provides a list of words and required threshold. A segment will match the pattern if the number of specified words found in the segment exceeds the threshold. This plugin can also use stems instead of words.

The plugins provide several convenient ways of defining precise patterns which a user is looking for. They provide much more flexibility than the query languages provided by the most popular publicly available search services. Particular examples of the patterns and found content will be provided in the next section.

IV. Tests

Special set of tests of the system has been performed after implementing proposed solutions. For testing purposes, four most popular news portals, according to Alexa [2] ranking, were selected :

- interia.pl [1],
- gazeta.pl [2],
- onet.pl [3],
- wp.pl [4].

For the crawl process, five detectors based on the methods described in previous chapter were configured. All of detectors searched for patterns in Polish language, due the fact that in the implementation of the stemmer algorithm and its database was available only for Polish language. It could be easily adapted to other languages by implementing the stemmer algorithm and its database in other languages.

[1]www.fakty.interia.pl
[2]www.wiadomosci.gazeta.pl
[3]www.wiadomosci.onet.pl
[4]www.wiadomosci.wp.pl

Detectors that were used to search for results, are:

- *Simple* – based on *List of words* method, parametrized by words ”virus” and ”flu”,
- *Stem* – based on *List of stems* method, parametrized by words ”virus” and ”flu” and it’s stems,
- *Distance* – based on *List of close words* method, parametrized by words ”virus” and ”flu” and it’s stems and distance between first and last word equal 5,
- *InPhrase* – based on *List of words in sentence* method, parametrized by words ”virus” and ”flu” and it’s stems,
- *Percentage* – based on *List of optional words* method, parametrized by four words: ”virus”, ”flu”, ”AH1N1” (which is special kind of flu widespread in Poland in December 2012), ”disease” and it’s stems, with minimum 50% percent of threshold.

Detectors search for defined patterns within the processed web page body and return the results if all searched criteria are fulfilled. The big advantage of the proposed system is the results memory mechanism. It allows to return the result only if it appears on processed web page for the first time, or if the surrounding content has changed since the previous processing.

TABLE I
Crawl performance

Domain	NoC/FC	GS	AFU/APU	ACT	EFF	UM/PM
interia.pl	17/11	3,8mln	3320/3150	1h 45m	0,5	0/0
gazeta.pl	17/12	3,6mln	4375/3951	1h 52m	0,58	24/283
onet.pl	15/10	2,4mln	6112/5950	3h 10m	0,52	119/286
wp.pl	17/8	6,9mln	2602/2520	1h 45m	0,4	258/581

- NoC - number of crawls
- FC - number of full crawl processes
- GS - number of pages returned by Google ”site:” query
- AFU - average number of urls found on domain during single crawl
- APU - average number of urls processed on domain during single crawl
- ACT - average total domain crawl time
- EFF - average crawl efficiency [urls/second]
- UM - number of urls with with any patterns matched
- PM - number of patterns matched within the domain

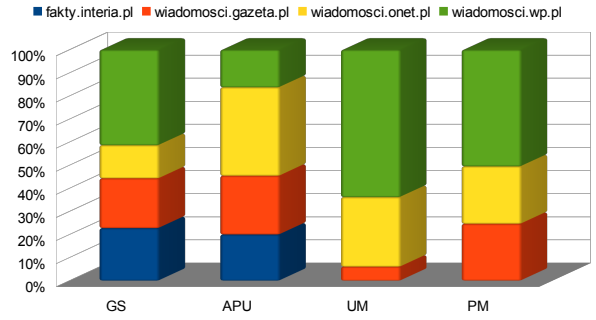


Fig. 3. Crawl performance diagram

The test period last 48 hours – it started on Saturday at 10am and it finished on Monday at 10am. Crawl performance

recorded during the test are shown in Table I and Figure 3. There could be observed some interesting facts:

- Number of urls within the domains returned by the Google "site:" search query varied from 2,4 to almost 7 millions. The number is huge, because it includes archive urls from the Google storage. Performed crawl returned at most about 6000 urls on single crawled domain. This is less than 0.3% of urls declared by the Google search engine, but those urls are all currently available urls, accessible by links spread from top domain url.
- Time of the single domain crawl process varied from 1h30min to 4 hours per domain. Average domain crawl efficiency was comparable for all the domains, about 0,5 processed url per second.
- Not every crawl process succeeded with full url list being processed. Some crawls has been terminated after crawling just a few percent of urls from domain. This is probably a result of temporary ban for IP of the crawling system.
- The number of returned results is neither related to the estimate (returned by Google "site:" query), nor real urls number of the domain. Most results were found on the smallest of crawled domains. The biggest domain was on the second position regarding number of unique web pages with matched pattern.
- One of the domains (fakty.interia.pl) did not contain any results for whole test crawl period.
- About 95% (average) of urls found during crawl process are the documents in textual (text or html) format and they are processed by the detectors.

Crawl and detection effects are presented in Table II. Results are grouped into two hour time units. First two rows of table indicate hour of test ("HoT") and related hour of day ("HoD") of crawl process. "NoU" row represents a number of unique urls found by all detectors within every domain crawled during the test. "NoP" represents a number of patterns found on web pages by all detectors. Below, there are statistics for all detectors separately, displayed as "NoU" and "NoP" values.

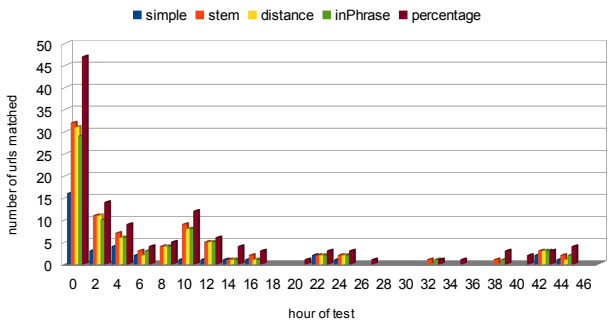


Fig. 4. Number of web pages found during crawl

Interesting remarks that could be observed on a base of test results are:

- The detector with biggest number of urls returned is the *Percentage* one – based on *List of optional words*. It is also result of its search criteria - 2 of of 4 words were required to report as pattern matched.
- Detectors based on word's stems (Stem/Distance/InPhrase) return very similar number of results (about 90% of average common results).
- Detector Simple (without word's stems) returns about 40% of stem-based detector results and matches about 30% of patterns found by stem-based detectors.
- In the figure 4, there is clearly shown, that most of the results (above 51%) are found during first crawl process (in first 4 hours of test).
- There could be observed some tendencies and periodicity of new information appearance. New patterns are found every morning, between 22 and 26 hour of test (8-12am), and also between 42 and 46 hour of test (4-8am). It differs slightly but it is probably caused by crawl density process. Also Saturday afternoon is the time, when the information peak could be observed. Both of this remarks could be result of:
  - new articles published (morning news),
  - increased activities of users commenting articles (evenings).
- an average number of detected pattens observed in the Figure 5 correspond to the trend of the number of unique urls containing pattern. Although there can be observed some deviations. On the 6 and the 16th hour of crawl, there are conspicuous peek of number of detected patterns, which is not related to an adequate incrementation of number of unique pages containing results (8 pages with 79 results and 14 pages with 210 results). Such result was caused by crawling the portal's search engine webpages, containing set of queries and query results related to search topic.

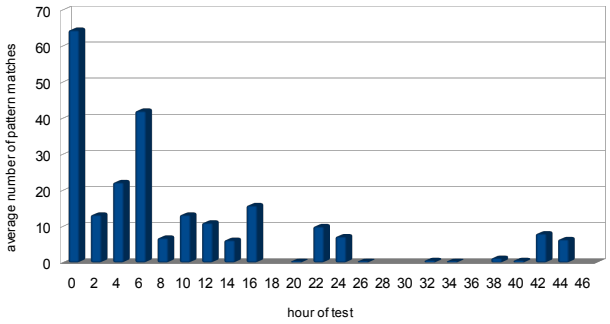


Fig. 5. Number of patterns matched on web pages during crawl

Table III presents statistics about number of patterns found on single web page. The remarks based on those results are:

TABLE II
DETECTOR RESULTS

HoT	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46
HoD	10	12	14	16	18	20	22	00	02	04	06	08	10	12	14	16	18	20	22	00	02	04	06	08
NoU-Simple	16	3	4	2	0	1	1	1	1	0	0	2	1	0	0	0	0	0	0	0	0	2	1	0
NoP-Simple	24	3	11	11	0	4	4	4	3	0	0	7	4	0	0	0	0	0	0	0	0	5	4	0
NoU-Stem	32	11	7	3	4	9	5	1	2	0	0	2	2	0	0	0	1	0	0	1	0	3	2	0
NoP-Stem	67	15	25	59	8	16	13	7	31	0	0	10	8	0	0	0	1	0	0	1	0	10	8	0
NoU-Dist	31	11	6	2	4	8	5	1	1	0	0	2	2	0	0	0	0	0	0	0	0	3	1	0
NoP-Dist	63	15	19	36	8	11	10	4	3	0	0	8	7	0	0	0	0	0	0	0	0	7	4	0
NoU-InPhr	29	10	6	3	4	8	5	1	1	0	0	2	2	0	0	0	1	0	0	1	0	3	2	0
NoP-InPhr	52	13	19	36	8	11	10	4	3	0	0	8	7	0	0	0	1	0	0	1	0	7	5	0
NoU-Perc	47	14	9	4	5	12	6	4	3	0	1	3	3	1	0	0	1	1	0	3	2	3	4	0
NoP-Perc	116	20	37	68	10	24	18	12	39	0	2	17	10	2	0	0	1	2	0	4	3	11	11	0
NoU	155	49	32	14	17	38	22	8	8	0	1	11	10	1	0	0	3	1	0	5	2	14	10	0
NoP	322	66	111	210	34	66	55	31	79	0	2	50	36	2	0	0	3	2	0	6	3	40	32	0

- HoT - hour of a test (0 - 47)
- HoD - hour of a day (0 - 23)
- NoU - number of unique urls with results (for all detectors)
- NoP - number of pattern occurrences (for all detectors)
- NoU-X, NoP-X - NoU or NoP for particular detector

TABLE III
NUMBER OF PATTERNS FOUND ON WEB PAGE

Number of patterns on page	1	2	3	4	5	6	7	8	9	10	11	12	13	14	28	34	35	57	65
Number of occurrences	203	93	28	30	8	6	10	10	2	1	1	1	1	1	1	2	1	1	1

- above 50% contained only one pattern within its body,
- almost 25% contained 2 pattern matches within its body,
- about 15% of web pages contained 3 or 4 pattern matches,
- 9% of results contained 5 to 8 results within its body,
- less than 3% of results contained more than 8 pattern occurrences.

The tests showed clearly, that the solutions used in the presented system works correctly. The performance was satisfactory during whole two-day long experiment. The changes detection algorithm was able to find fragments of Web pages, which have actually changed between visits. The detectors using the stemmer algorithm have demonstrated advantages over a simple, word-based query language.

V. CONCLUSIONS AND FURTHER WORK

The system presented in this paper is a promising base for further works and development, in the area of information retrieval from WEB resources. After certains improvements it could provide useful functionalities, that can definitely find applications in real-life scenarios. Created method for detecting modifications in the webpages' content provides an efficient way of finding only newly-added information.

The proposed solution may also have a variety of applications in open source intelligence analysis. The crawler can be used to build a knowledge base that contains information about objects, events and relations between them (e.g. companies and involved people). This can be further integrated with other analytical tools, such as LINK platform, which supports mainly criminal analysis [13]. This way the crawler can be used as valuable data source for performing various analyses (for example searching relations between people and companies involved in fraud).

Further research on the approach will include performance improvements and development of more advanced methods for defining content patterns. Moreover, tags-avoiding methods are planned, in order to optimize results quality. A possibility of defining semantic meaning of a content or of a similarity to a given text would be more useful than specifying a list of words. This could be achieved by methods of machine learning [14] for building classifiers for particular types of content

ACKNOWLEDGMENT

The research leading to these results has received funding from the research project No. O ROB 0008 01 "Advanced IT techniques supporting data processing in criminal analysis", funded by the Polish National Centre for Research and Development.

REFERENCES

[1] M. Kunder, *WorldWideWebSize.com*, 12.2012
[2] Alexa – provider of global web metrics, *http://www.alexa.com/*, 01.2013.
[3] J. Alpert, N. Hajaj, *We knew the web was big...*, http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html, 25.07.2008
[4] L. Liu, W. Tang, D. Buttler, C. Pu. *Information Monitoring on the Web: A Scalable Solution* World Wide Web, 2002, Volume 5, Issue 4, pp 263-304
[5] Liu, L., Pu, C., Tang, W.; *WebCQ-detecting and delivering information changes on the web.* In Proceedings of the ninth international conference on Information and knowledge management (pp. 512-519). ACM, 2000
[6] V. Ramasubramanian, R. Peterson, E.G. Sirer, *Corona: A high performance publish-subscribe system for the world wide web.* Proceedings of Networked System Design and Implementation (NSDI). 2006
[7] F. Menczer, R.K. Belew. *Adaptive Information Agents in Distributed Textual Environments*, Proceedings of the 2nd International Conference on Autonomous Agents, ACM Press, 1998, p. 157-164.

[8] H. Dong, F.K. Hussain, E. Chang. *State of the Art in Semantic Focused Crawlers*. Computational Science and Its Applications ICCSA 2009, International Conference, Seoul, Korea, 2009, p. 910–924.

[9] B. Tan, S. Foo, S. C. Hui; *Web information monitoring for competitive intelligence* Cybernetics and Systems, Vol. 33, Iss. 3, 2002

[10] K. Dorosz, M. Korzycki. *Latent Semantic Analysis Evaluation of Conceptual Dependency Driven Focused Crawling*. Multimedia Communications, Services and Security, 5th International Conference, MCSS 2012, Krakow, Poland, 2012, p. 77–84.

[11] W. Turek, A. Opaliński, M. Kisiel-Dorohinicki *Extensible Web Crawler – Towards Multimedia Material Analysis*, Multimedia Communications, Services and Security, 5th International Conference, MCSS 2011, Krakow, Poland, 2011, p. 183–190.

[12] K. Wilaszek, T. Wjcik, A. Opaliński, W. Turek. *Internet Identity Analysis and Similarities Detection*, Multimedia Communications, Services and Security, 5th International Conference, MCSS 2012, Krakow, Poland, 2012, p. 369–379.

[13] R. Debski, M. Kisiel-Dorohinicki, T. Milos, K. Pietak, *LINK: a decision-support system for criminal analysis*, MCSS 2010: Multimedia Communications, Services and Security: IEEE International Conference, Springer, 2010, p.110-115

[14] B. Śnieżyński, *Resource Management in a Multi-agent System by Means of Reinforcement Learning and Supervised Rule Learning*, Springer Lecture Notes in Computer Science vol.4488, 2007, p. 864–871.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1269–1271

Business Intelligence as a service in a cloud environment

Maciej Pondel
Wroclaw University of Economics
ul. Komandorska 118/120,
53-345 Wrocław, Poland
Email: maciej.pondel@ue.wroc.pl

Abstract— Business Intellligence should be described as a way of managing our company more than a set of functionalities in a computer software. Acquiring a real profit requires enterprise management to understand the value of the data and the way data describe business processes. Being aware of the business and measuring its performance we are able to improve the processes and make whole the business more effective. To achieve business improvement we require efficient Business Intelligence system as a combination of a software, hardware, the communication infrastructure and services regarding data preparation, integration and delivery to the system.

In this paper author considers if the service oriented approach and cloud computing can make BI implementation more efficient.

I. Introduction

BUSINESS INTELLIGENCE for sure is the way for the enterprise to make the business more efficient. BI systems are fed by a big volumes of data coming from the transactional systems working inside company. That is why the BI systems are most often on-premise software which mean they are hosted on the servers inside the building the company is located.

Another way of using software is called software as a service or cloud computing which can bring the significant benefits for a company. Are these benefits strong enough to make BI implementation in cloud environment more efficient than on-premise?

II. BI Implementation

Different definitions show that the components of business intelligence software are inter ailia[1]:

- Data warehousing
- Multidimensional analysis, for example OLAP
- Data mining
- Business analysis
- Visualization
- Querying, reporting and charting (including just-in-time and agent-based alerts)
- Geospatial analysis

The software is supplied by the following categories of data:

- Operational data (regarding financial, logistics, sales, orders, personnel, billing)
- Private data (mainly spreadsheets prepared by business analysts, knowledge workers, statisticians and managers regarding analysis and reports)
- External data (purchased from vendors specializing in collecting industry-specific information such as: Health care statistics, customer profile information, customer credit reports, trends, currency fluctuations, stock prices demographic data and many more)

Operational data in most cases come from the transactional systems hosted inside the company. They are the biggest data volumes from all mentioned categories. Operational data is probably the most important data category because the data describe directly our business.

In a current situation most of BI systems are implemented basing on the following assumptions:

- BI system is located in the same LAN as the main transactional systems, because it collects the huge amount of data from transactional systems,
- Data loading is performed once a day at the time when users do not access the systems,
- The internal structure of the data warehouse is unique for every implementation (because every enterprise has unique combination of transactional systems that they use).
- The user interfaces are usually typical for business analytics, reporting, visualizations, charting and manager dashboards
- BI system needs a lot of resources, mainly: disk storage because of the volumes of data, network capacity during data loading, processor and memory during data processing.

We have to admit, that BI is evolving. The new possibilities of use appear what implies new challenges for Business Intelligence. We can find among them [3],[8]:

- Big Data analytics

978-83-60810-53-8/$25.00

- Mobile BI
- In memory BI
- Self service BI
- Consumerization of Enterprise Software

III. SERVICE ORIENTED ARCHITECTURE

Service oriented architecture of software is not a new technology [4]. It is rather the idea of creating software as a set of modules called services that collectively provide the complete functionality of a large software application. The services should cooperate by exchanging data and information with other services without any human interaction. The services should be treated as black boxes with precisely defined input parameters and output results. For the architect of enterprise application consisted of services the algorithms implemented inside the service are not so important. The services can have various types. We can have:

- business service – simple IT component responsible for a part of business,
- Web Service – as a method of communication between 2 programs based on World Wide Web protocol
- IT Service – as a business process of supplying benefits to the recipient by a supplier

In order to build efficient SOA solutions the services must meet following requirements:

- Interoperability between different programming languages, systems that allow integration of services
- Federation of resources that allows transparently mapping multiple autonomous resources to be treated by users as one federated resource

Cloud is becoming more and more popular environment for hosting business applications. All greatest IT vendors in the world provide their software as a service available in the cloud. It means the systems are hosted in the vendor's environment on vendor's severs and the systems are available by Internet connection. To the most popular IT systems provided in cloud belong:

- Enterprise email systems together with tools regarding workgroup like shared calendars, resources reservation and applications improving employees' productivity,
- Document management systems,
- Content management systems,
- Databases (like Google Cloud SQL or Windows Azure SQL),
- Business applications like CRM, ERP,
- Application hosting services with programmers API allowing customer to build his own software and host it in the cloud.

The main assumption of cloud services according to Forrester's definition is supplying standardized IT capacity over Internet in a pay-per-use and self-service way [7]. We can understand, that it is suitable way of using IT for all systems that works in a standard way. It we have to use dedicated software the cloud can be too limited for us. According to the Internet technologies – cloud is supporting the standards prevalent in mobile devices. Most of the standardized cloud services can be run at any devices with any operating systems.

What are the main benefits or using cloud services in business[6]?

- It lowers the costs of entry for smaller companies who are trying to use the same software and technologies as the big corporations.
- If the IT solution needs a large amount of computing power for relatively short time the cloud can provide us the resources dynamically what is financially available even for small companies.
- It provides immediate access to hardware resources without any upfront investments. It shortens the time to market for many IT solutions.
- Even for large enterprises cloud can allow to scale their applications in a simpler way. The company has an easy access to the new computer resources whenever then are necessary.
- Cloud computing requires payments only for the resources the company really utilizes. In many cases it is more effective financially than investing the money on the start.

When we compare the characteristics of the cloud environment and the benefits with the assumptions of BI implementations and challenges of BI we can observe that they match each other. Many of those benefits concerning cloud are suitable for Business Intelligence solutions. The features of BI Systems and cloud are summarized in table 1. As we can observe there are some features that do not fit each other. The basic problem concern the huge data transfer from transactional systems hosted mainly on-premise to the solution kept in cloud. The second issue is about dedicated integration procedures and unique data warehouse structure what do not match the standardization of the cloud solutions. We have to admit that those issues will not occur when we have transactional systems hosted also in the cloud – the cloud BI system is a very natural solution because:

- data during the loading stay in the same environment and do not overload the network connection
- the procedures of integration will be typical because the cloud transactional systems must be implemented in a common way

TABLE I.
BI SYSTEMS AND CLOUD CHARACTERISTICS

BI Characteristics and challenges\ cloud features	Standardized solutions	Internet technologies	Pay-for-use	Dynamic allocation of resources	Hosting in the vendor environment
Huge data transfer from transactional systems					-
Data loading once a day				+	
Unique data warehouse structure and unique integration	-				
Typical user interfaces	+				
Needs a lot of resources in various part of time			+	+	
Big data analysis					
Mobile BI	+	+			
In memory BI					
Self service BI	+				

If we have most common architecture of transactional systems and they are hosted on-premise we have to elaborate the solution for the indicated issues.

One of the solutions may be an adoption of a hybrid BI environment consisting of:

- the data integration layer which is stored on premise
- the business logic layer stored in the cloud

IV. CONCLUSION

This paper presents the idea of cloud BI system. There are some cases that may inhibit such model of BI software functioning, but author presents how to manage them and what advantages gives combining features of a cloud environment with user expectations regarding BI systems.

REFERENCES

[1] Moss L., Atre S. Business Intelligence Roadmap, Pearson Education, Boston 2003
[2] Wu L., Barash G., Bartolini C. „A Service-oriented Architecture for Business Intelligence", Service-Oriented Computing and Applications, 2007. SOCA '07. IEEE Computer Society, Los Alamitos 2007
[3] http://bi.pl/publications/art/99-trendy-w-business-intelligence-z-czym-to-sie-je-jak-czytac-menu-i-co-zamowic#section_section-375
[4] Łagowski L, „SOA – Ideologia nie technologia", XV Konferencja PLOUG, Kościelisko, 2009
[5] Velte, Anthony T. (2010). Cloud Computing: A Practical Approach. McGraw Hill. ISBN 978-0-07-162694-1.
[6] Marston S, Li Z, Bandyopadhyay S, Zhang J, Ghalsasi A. Cloud computing: the business perspective. Decis Support Syst. 2011
[7] Lecznar M., Patig S. Cloud Computing providers: Characteristics and recommendations in: Babin G, E-Technologies: Transformation in a Connected World: 5th International Conference, MCETECH 2011, Springer Berlin-Heidelberg 2011
[8] Howson C., 7 Top Business Intelligence Trends For 2013: http://www.informationweek.com/software/business-intelligence/7-top-business-intelligence-trends-for-2/240146994

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1273–1278

Knowledge Acquisition for New Product Development with the Use of an ERP Database

Marcin Relich
University of Zielona Gora,
ul. Licealna 9, 65-216 Zielona Gora, Poland
Email: m.relich@wez.uz.zgora.pl

***Abstract*—Nowadays, a considerable number of enterprises develop new products using an Enterprise Resource Planning (ERP) system. One of the modules of a typical ERP system concerns project management. Functionalities of this module consist of defining resources, company calendars, sequence of project tasks, task duration etc. in order to obtain a project schedule. These parameters can be defined by the employees according to their knowledge, or they can be connected with data from previous completed projects. The paper investigates using an ERP database to identify critical factors, i.e. variables that significantly influence on new product development. Project duration and cost is estimated by a fuzzy neural system that uses data of completed projects stored in an ERP system.**

I. Introduction

THE present information and communication technologies have become one of the most important factors, conditions and chances of the firm development. These technologies enable the collection, presentation, transfer, access and using of enormous amount of data. The data are a potential source of information that in connection with manager skills and experience may influence on the choice of the correct decision. ERP systems help to collect, operate, and store data concerning daily activities of an enterprise (e.g. client orders), as well as the results of previous projects (development of products).

Project success or failure depends on many critical factors, such as the kind of project, access to resources, methods of project management, and environment [1], [2]. The reasons for project failure can be generally considered as a lack of accessibility of resources (e.g. human, financial, raw materials) and changeability of the external environment. Moreover, unstable requirements, lack of well-defined scope, quality of management, and skill of the employees can cause project failure. To reduce project overruns, there are two ways to approach the problem. The first way is to increase the accuracy of the estimates through a better estimation process and the second, to increase the project control.

It is unrealistic to expect very accurate estimates of project effort because of the inherent uncertainty in development projects, and the complex and dynamic interaction of factors that influence on its development. However, even small improvements will be valuable, especially by large-scale projects. More accurate forecasting supports the project managers in planning and monitoring the project, for instance, in project cost, resource allocation, and schedule arrangement.

New product development is connected with uncertainty that includes both internal (e.g. communication in project team, planning techniques, cash flow) and external environments (e.g. social, economic, political, technological conditions). Sources of uncertainty are wide ranging and have a fundamental effect on projects and project management [3]. Uncertainty is an important issue in the support of any decision-making in the process of new product development. Since most companies should estimate project parameters, there is a need to develop an approach that takes into account the imprecise character of data and copes with enormous amount of data. The description of project management and knowledge management in the context of an ERP system, as well as a fuzzy neural system combining the ability for learning and processing inaccurate data is presented in the next section.

II. Background

A. Project management in an ERP system

In recent years, the advancement of information technology in business management processes has placed ERP system as one of the most widely implemented business software in various enterprises. ERP software promises significant benefits to organizations. Some of these benefits include lowering costs, reducing inventories, increasing productivity [4], improving operational efficiency [5], [6], attaining competitive advantage [7], and bettering the reorganization of internal resources [8], [9].

The goal of an ERP based integrated information system is to make the system effective, efficient and user friendly. The performance of software depends on the interaction between the software and users. The primary task of an integrated system is to maintain the data flow of an organization and to reduce the redundancy [10]. ERP is a system for the seamless integration of all the information flowing through the company such as finances, accounting, human resources, supply chain, and customer information [11]. One of the functionalities of an ERP system also includes project management that company can use to develop new products.

The project management functionality comprises the definition of master files and obtaining a project schedule. The definition of master files includes resources that are use in a project, resource calendars, company calendars, time models (blocks), as well as project activities, estimates of activity duration, sequencing (order of activities), milestones. This

978-83-60810-53-8/$25.00

data is the bases to obtain resource allocation, planned cost, network chart, and analyses concerning e.g. original and actual cost.

B. ERP systems from a knowledge perspective

A variety of knowledge management and knowledge integrated manufacturing models have emerged one after another in recent years [12]. These models can be considered in the context of manufacturing or knowledge aspect [13]. The role of manufacturing knowledge is a key strategic resource and it can be presented as the interactive process between manufacturing knowledge and cross-functional activities [14], [15], [16]. Knowledge management may be incorporated into ERP implementation with the use, for instance, a self-sufficient model [17]. In turn, a base to characterize product development and knowledge evolution process can be an integrated knowledge reference system [18].

In the implementation phase, ERP systems have an impact on organizational knowledge (stock of knowledge, distribution, learning processes) [19], as well as on efficiency and flexibility of a knowledge management system [20]. Moreover, business knowledge in the ERP software package to the adopting organization (types of knowledge transferred, resolution of conflicts with existing organizational knowledge, changed knowledge structure) is transferred and internalized from consultants to clients [21], [22]. Knowledge-related ERP systems research mainly concerns the knowledge issues encountered during the system implementation phase or the 'shake-out' period immediately following implementation [23]. The use of an ERP database in the post-implementation phase of ERP system lifecycle is usually obscured. This provides the motivation to develop the approach dedicated for knowledge acquisition in the context of new product development in an enterprise that uses an ERP system.

C. Description of fuzzy neural system

Knowledge acquisition requires some techniques that cope with the description of relationships among data and that solve the problems connected with e.g. classification, regression, and clustering. These techniques include neural networks, fuzzy sets, rough sets, time series analysis, Bayesian networks, decision trees, evolutionary programming and genetic algorithms, Markov modeling, etc.

Fuzzy logic and artificial neural networks are complementary technologies and powerful design techniques that have their strengths and weaknesses [24]. Table I shows a comparison of the properties of these two technologies.

The fuzzy neural system has the advantages of both neural networks (e.g. learning abilities, optimization abilities and connectionist structures) and fuzzy systems (simplicity of incorporating expert knowledge). As a result, it is possible to bring the low-level learning and computational power of neural networks into fuzzy systems and also high-level human like IF-THEN thinking and reasoning of fuzzy systems into neural networks. The fuzzy neural method is rather a way to create a fuzzy model from data by some kind of learning method that is motivated by learning procedures used in neural networks. This substantially reduces development time and cost while improving the accuracy of the resulting fuzzy model. Being able to utilize a neural learning algorithm implies that a fuzzy system with linguistic information in its rule base can be updated or adapted using numerical information to gain an even greater advantage over a neural network that cannot make use of linguistic information and behaves as a black box [25].

The behaviour of a fuzzy neural system can be represented by a set of humanly understandable rules or by a combination of localized basis functions associated with local models, making them an ideal framework to perform nonlinear predictive modelling. Nevertheless, one important consequence of this hybridization between the representational aspect of fuzzy models and the learning mechanism of neural networks is the contrast between readability and performance of the resulting model [25]. The combination of fuzzy systems and neural networks has recently become a popular approach in engineering fields for solving problems in control, identification, prediction, pattern recognition, etc [26], [27], [28]. One well-known structure is the adaptive neuro-fuzzy inference system (ANFIS). ANFIS model is a universal approximator which has the non-linear modelling and forecasting function.

III. Method for estimating project duration and cost

The proposed method is dedicated for new product development in an enterprise that uses an ERP system. New product development is often connected with the superficial changes in design and/or functionality of past products. Thus, data of completed projects can be used to identify relationships between the parameters of past projects and their durations and costs. The method consists of the following stages:

1) extracting data (parameters of past projects) from an ERP system;
2) identification of critical factors that significantly influence on new product development;
3) learning ANFIS in order to obtain rule base;
4) estimating duration and cost of new product development;
5) loading data (estimate of project duration and cost) to an ERP system (module project management).

The presented methodology concerns the estimation of project duration and cost in the different phases of new product development (see Fig. 1). In each of these phases, the critical factors (parameters of an ERP database) that significantly influence on new product development are sought.

Database of an ERP system comprises an enormous number of parameters that can be considered as potential variables to identify the duration and cost of project phases. The second stage in the above-presented procedure concerns the identification of critical factors that influence on the project duration and cost, and indirect on new product development. If the relationship between a variable and the project duration and cost is significant (greater than a level defined by the user), then the variable is considered as the critical factor. The

TABLE I
PROPERTIES OF NEURAL NETWORKS AND FUZZY SYSTEMS

Skills	Type	Fuzzy Systems	Neural Networks
Knowledge acquisition	Inputs	Human experts	Sample sets
	Tools	Interaction	Algorithms
Uncertainty	Information	Quantitative and qualitative	Quantitative
Reasoning	Cognition	Heuristic approach	Perception
	Mechanism	Low	Parallel Computation
	Speed	Low	High
Adaption	Fault-tolerance	Low	Very high
	Learning	Induction	Adjusting weights
Natural language	Implementation	Explicit	Implicit
	Flexibility	High	Low

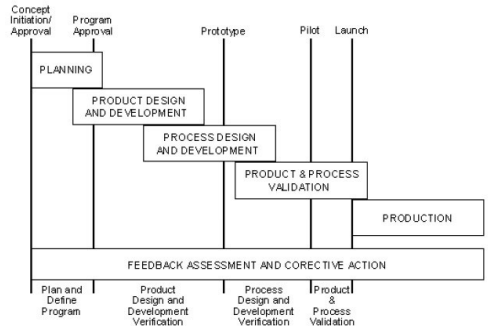


Fig. 1. New product planning phases
Source: [29]

variables are chosen according to the user's experience and can be as follows: a number of human resource (in person-hour), machine-hour, raw material, and activities in a project phase, as well as financial means, delay in client's payment and material delivery by suppliers, quality of material (number of complaints), time of machine inspection, absenteeism during project implementation, project team members, and project manager.

A large number of independent variables in a large data set can present two major problems. Firstly, too many variables result in long training times when the model is built. Secondly, large number of observations and variables tend to retain redundant information through multicollinearity leading to unreliable models. Some of the variables present in historical data are needed for some problems and some variables for others. Often, different variables may carry the same information [30].

A variable reduction method can be based on principal component analysis that is used as a dimension reduction technique for linearly mapping high dimensional data onto a lower dimension with minimal loss of information. The variable reduction is not the main issue in this research and it is not further considered.

The third stage in the proposed methodology concerns obtaining rule base with the use of ANFIS. The identification of rules and the initial parameters of membership function of fuzzy sets are obtained with the use of e.g. grid partition, fuzzy c-mean, or subtractive clustering. The learning stage requires the declaration of optimisation weights method (e.g. backpropagation algorithm) and stop criterion (e.g. error value or the number of iteration). After learning phase, the testing data are led to input of system to compare the results with target. Next section presents an example concerning the use of the above procedure.

IV. EXAMPLE OF PROJECT DURATION AND COST ASSESSMENT

The output variables contain the duration (d_i in months) and cost (c_i in monetary unit - m.u.) of the *j*-th phase in project *i*. In turn, the input variables include:

- h_{ji} - number of human resource in the *j*-th phase of project *i* (person-hour);
- a_{ji} - number of activities in the *j*-th phase of project *i*;
- s_{ji} - number of subcontractors in the *j*-th phase of project *i*;
- tm_{ji} - number of project team members in the *j*-th phase of project *i*.

Table II presents data of eight past projects (development of products) for product design phase that has been applied to the proposed approach.

Calculation has been generated with the use of ANFIS tool that is Matlab software. The application of fuzzy-neural system requires the declaration of input variables and parameters connected with ANFIS, e.g. defuzzification method. Figure 2 presents two ANFIS, for the duration and cost of project phase, respectively.

After the declaration of input variables in fuzzy neural system, the initial parameters of membership functions of fuzzy sets are estimated. As a result, the structure of fuzzy neural system is determined. In next stage, the fuzzy neural system is learnt according to e.g. backpropagation algorithm, and consequently, the shape of membership function is determined (Fig. 3).

The rules can be presented for decision maker in descriptive form. The example of fuzzy rules for the duration and cost is presented in Fig. 4.

To eliminate too strictly function adjustment to data and to increase the estimation quality, the data set is divided into

TABLE II
PROJECT VARIABLES FOR PRODUCT DESIGN PHASE

Variable	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
Human resource	500	400	350	450	600	400	650	500
Number of activities	25	22	20	25	28	22	25	20
Number of subcontractors	3	4	4	5	6	4	5	5
Number of team members	8	7	6	8	12	10	12	10
Duration	14	12	10	15	16	15	15	15
Cost	380	320	320	400	500	420	650	600

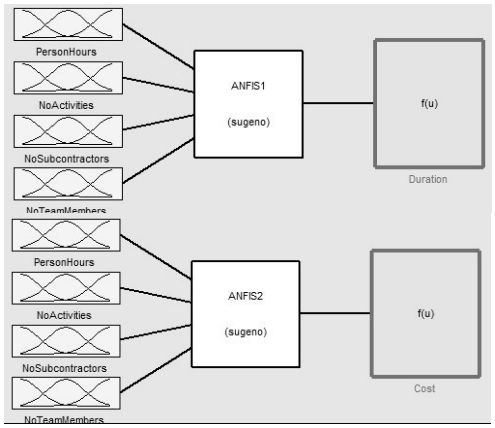


Fig. 2. Specification of fuzzy neural system

learning (P_1-P_6) and testing set (P_7-P_8). The learning phase requires the declaration of method of weights optimisation, and stop criterion (e.g. error value or the number of iteration). After learning phase, the testing data are led to input of system to compare the error between different models. Mean square error (MSE) for various models are presented in Table III. It is noteworthy that the least error in testing set for the duration has been generated with the use of average. It can be connected with a low level of variance for the duration of project design phase. In turn, the least error for the cost has been generated with the use of ANFIS with subtractive clustering method.

The membership functions and rules are a basis to evaluate the duration and cost of an actual project. Let us assume that for the actual project are considered the following values: number of person hours equal 475, number of activities equal 24, number of subcontractors equal 11, and number of team members equal 9. Thus, the duration equals 16.4 months and cost of project phase equals 440 m.u. (see Fig. 5).

There is also possibility to conduct what-if analysis. For instance, if a number of subcontractors will be increased to 20, then the project will be decrease to 13.7 months (see Fig. 6).

The above-presented analysis is conducted for each phase of project and the obtained estimates can be used to evaluate cash flow, working capital, financial reserves, product launch, and other critical factors of an enterprise activity.

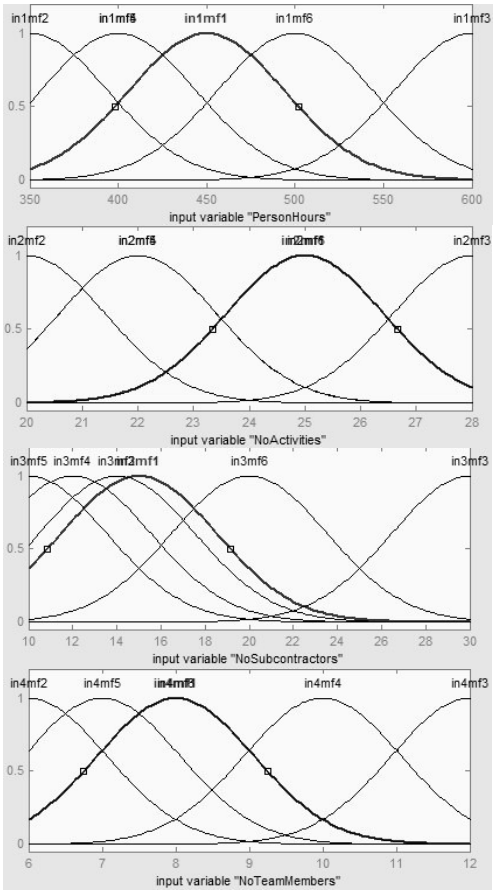


Fig. 3. Membership function for input variables

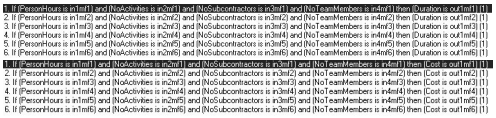


Fig. 4. Fuzzy rules generated by fuzzy neural system

TABLE III
COMPARISON OF MSE FOR DIFFERENT MODELS

Model	Duration	Cost
Average	1.78	236.32
Linear model	4.16	309.85
ANFIS - grid partition	12.87	564.63
ANFIS - subtractive clustering	2.99	96.10

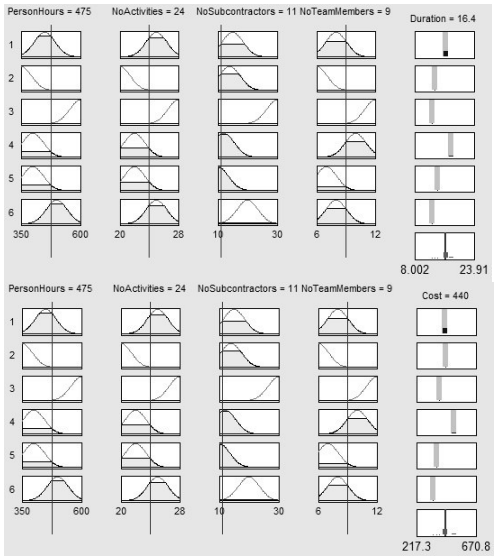


Fig. 5. Estimation of project duration and cost

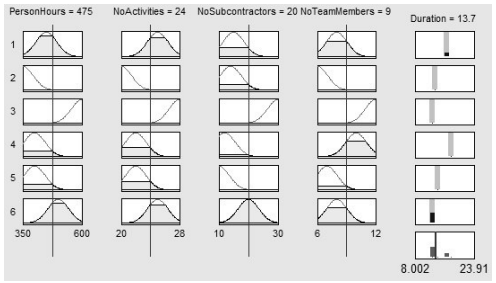


Fig. 6. Project duration for additional subcontractors

V. CONCLUSION

The capabilities of an enterprise to create, share and utilize knowledge effectively are today regarded as one of the key drivers of competitive advantage for industrial enterprises. Competition in quality, design, cost of new products, and time their launching into the market has increased with new competitors having established segments and, in some cases, with change in competitive tools. This forces more frequent and larger-scale changes in contemporary companies, also changes in the use of new information technologies. One of the technologies concerns a fuzzy neural system that is used in this paper to evaluate the project duration and cost.

More exact identification of project duration and cost enables more precision of cash flow planning and finally, decreases the risk of lack of liquidity. If in the enterprise is a database of past projects, then there is the possibility to gather additional information in the form of conditional rules. The application of the proposed approach encounters some difficulties, among other things, by the collecting enough amounts of data of the past similar projects. Moreover, the lack of uniform rules that concern the development of fuzzy neural systems may cause an acceptance problem for the decision-makers. However, the presented approach seems to have the promising properties for acquiring information from an ERP system.

Further research focuses on the development of the presented approach towards searching a set of key performance indicators according to their influence on the success of completed projects. Moreover, future research will be aimed at verifying the proposed approach in a real world to test its practicality.

REFERENCES

[1] P. Nitithamyong and M. J. Skibniewski, "Success/failure factors and performance measures of web-based construction project management systems: professionals' viewpoint," *Journal of Construction Engineering and Management,* vol. 132, 2006, pp. 80–87.
[2] S. Robertson and T. Williams, "Understanding project failure: using cognitive mapping in an insurance project," *Project Management Journal,* vol. 37, 2006, pp. 55–71.
[3] R. Atkinson, L. Crawford, and S. Ward, "Fundamental uncertainties in projects and the scope of project management," *International Journal of Project Management,* vol. 24, 2006, pp. 687–698.
[4] D. Olson, *Managerial Issues of Enterprise Resource Planning Systems,* McGrawHill/Irwin, Boston, 2004.
[5] J. Benders, R. Batenburg, and H. Van der Blonk, "Sticking to standards; technical and other isomorphic pressures in deploying ERP systems," *Information & Management*, vol. 43, 2006, pp. 194–203.
[6] L. Häkkinen and O. Hilmola, "ERP evaluation during the shakedown phase: lessons from an after-sales division," *Information Systems Journal,* vol. 18(1), 2008, pp. 73–100.
[7] J. Beard and M. Summer, "Seeking strategic advantage in the post-net era: viewing ERP systems from resource-based perspective," *Journal of Strategic Information Systems,* vol. 13, 2004, pp. 129–150.
[8] J. Stratman, "Realizing benefits from enterprise resource planning: does strategic focus matter?" *Production and Operations Management,* vol. 16(2), 2007, pp. 203–216.
[9] J. May, G. Dhillon, and M. Caldeira, "Defining value-based objectives for ERP systems planning," *Decision Support Systems,* vol. 55, 2013, pp. 98–109.
[10] A. Imtiaz and M. G. Kibria, "Modules to optimize the performance of an ERP based integrated information system," IEEE International Conference on Informatics, Electronics & Vision, 2012, pp. 598–601.
[11] T. Davenport, "Putting the enterprise into the enterprise system," *Harvard Business Review,* July-August, 1998, pp. 121–131.
[12] A. Bernard and S. Tichkiewitch, *Methods and tools for effective knowledge life-cycle-management,* Springer, Berlin, 2008.
[13] Y. Xu and A. Bernard, "Quantifying the value of knowledge within the context of product development," *Knowledge-Based Systems,* vol. 24, 2011, pp. 166–175.

[14] E. L. Paiva, A. V. Roth, and J. E. Fensterseifer, "Organizational knowledge and the manufacturing strategy process: a resource-based view analysis," *Journal of Operations Management,* vol. 26(1), 2008, pp. 115–132.
[15] M. Cambal, D. Caganova, and J. Sujanova, "The industrial enterprise performance increase through the competency model application," Proceedings of the 4th European Conferences on Intellectual Capital, Helsinki, Finland, 2012, pp. 118–126.
[16] J. Sujanova, P. Gabris, M. Licko, P. Pavlenda, and R. Stasiak-Betlejewska, "Aspects of Knowledge Management in Slovak Industrial Enterprises," Proceedings of the 13th European Conference on Knowledge Management, Cartagena, Spain, 2012, pp. 1135–1144.
[17] T. C. McGinnis and Z. Huang, "Rethinking ERP success: a new perspective from knowledge management and continuous improvement," *Information & Management,* vol. 44(7), 2007, pp. 626–634.
[18] A. Bernard and Y. Xu, "An integrated knowledge reference system for product development," *CIRP Annals,*, vol. 58(1), 2009, pp. 119–122.
[19] R. Baskerville, S. Pawlowski, and E. McLean, "Enterprise resource planning and organizational knowledge: patterns of convergence and divergence," Proceedings of the 21st International Conference on Information Systems, Brisbane, Australia, 2000, pp. 396–406.
[20] J. C. Huang, S. Newell, R. D. Galliers, and S. L. Pan, "ERP and knowledge management systems: managerial panaceas or synergetic solutions?" Proceedings of the Seventh Americas Conference on Information Systems, 2001, pp. 1136–1141.
[21] D. G. Ko, L. J. Kirsch, and W. R. King, "Antecedents of knowledge transfer from consultants to clients in enterprise system implementation," *MIS Quarterly,* vol. 9(1), 2005, pp. 59–85.
[22] Z. Lee and J. Lee "An ERP implementation case study from a knowledge transfer perspective," *Journal of Information Technology,* vol. 15(4), 2000, pp. 281–288.
[23] T. Srivardhana and S. D. Pawlowski, "ERP systems as an enabler of sustained business process innovation: a knowledge-based view," *Journal of Strategic Information Systems,* vol. 16, 2007, pp. 51–69
[24] R. Fuller, "Introduction to neuro-fuzzy systems," Advances in Soft Computing Series, Springer-Verlag, Berlin/Heildelberg, 2000.
[25] A. T. Azar, *Adaptive neuro-fuzzy systems,* In: Fuzzy systems. InTech, 2010.
[26] J. Zeng, M. An, and N. J. Smith, "Application of a fuzzy based decision making methodology to construction project risk assessment," *International Journal of Project Management,* vol. 25, 2007, pp. 589–600.
[27] M. Y. Cheng , H. C. Tsai, and E. Sudjono, "Evolutionary fuzzy hybrid neural network for project cash flow control," *Engineering Applications of Artificial Intelligence,* vol. 23, 2010, pp. 604–613.
[28] S. C. Chien , T. Y. Wang, and S. L. Lin, "Application of neuro-fuzzy networks to forecast innovation performance," *Expert Systems with Applications,* vol. 37, 2010, pp. 1086–1095.
[29] K. Sanongpong, "Automotive process-based new product development: A review of key performance metrics," Proceedings of the World Congress on Engineering, London, U.K., 2009.
[30] G. L. Colmenares and R. Perez, "A data reduction method to train, test, and validate neural networks," Proceedings of IEEE, Southeastcon, 1998, pp. 277–280.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1279–1285

Preliminaries for Dynamic Competence Management System building

Przemysław Różewski
West Pomeranian University
of Technology, ul. Żołnierska 49
71-210 Szczecin, Poland
Email: prozewski@wi.zut.edu.pl

Bartłomiej Małachowski
West Pomeranian University
of Technology, ul. Żołnierska 49
71-210 Szczecin, Poland
Email: bmalachowski@wi.zut.edu.pl

Jarosław Jankowski
West Pomeranian University
of Technology, ul. Żołnierska 49
71-210 Szczecin, Poland
Email: jjankowski@wi.zut.edu.pl

Marcin Prys
Inspeo Sp. z o.o.,
ul. Łużycka 87, Gryfino, Poland
http://www.inspeo.com/
Email: marcinprys@mindflow.pl

Piotr Dańczura
West Pomeranian University
of Technology ul. Żołnierska 49
71-210 Szczecin, Poland
Email: piotrdanczura@gmail.com

***Abstract*—Competence management systems are an important addition to knowledge management systems. Competencies can be processed, during the identification, assessment and acquisition processes, because there is a certain set of tools used to test competencies and estimate their levels. In this paper, we focused on the analysis of the concept of Dynamic Competence Management System. The system takes into account competence changes caused by the efflux of time and competence diffusion process in project group.**

I. Introduction

THE management and control of knowledge and skills, and more recently the management of firms' competencies have turned out to be essential factors of industrial processes' performance and part of a strategic objective of human capital management [2]. In studies related to knowledge management, managing competencies is becoming a crucial research problem [22]. On one, hand it is being analysed from the point of view of an educational organisation in which we focus on transparent description of a student's achievements in a form of their competencies and their levels described [15]. On the other hand, we analyse competencies in companies. According to [26] *intellectual capital=competence x commitment*. Competence profile of an employee should allow us to make the right decisions regarding training, assigning to a project or even recruiting. Moreover, the knowledge about the competencies is produced and transformed by identification, assessment and acquisition processes [25]. Competencies can be processed because there is a certain set of tools used to test competencies and estimate their levels (e.g. http://www.inspeo.com, http://www.matchinglab.com, http://www.actonomy.com/) and methodology to competence assessment [14]. Moreover, the process of competence computing should be understood as enabling the use of competence databases for inference and combination of competencies for different functions and processes, not as a reductionist account of competencies to numeric models [6].

Large enterprises are characteristic of several features that cause difficulties in managing competencies of their employees. The first feature is a high personnel rotation on different positions. Second one is related to the lack of standardised system approach regarding saving and storing information about competencies. The competencies themselves change in a dynamic way, employees thanks to trainings or being members of some projects are developing and/or achieving new competencies. Additionally, we have to factor in a constant obsolescence of knowledge, which affects the competencies achieved by certain employee. Luckily this process is being avoided by implementing Life Long Learning policy. Organisations constantly approach the question of what employees are needed for certain projects, so that its rate success will be as high as possible thanks to certain set of competencies.

This paper will describe the concept of Dynamic Competencies Management System (DCMS) which will help in better processing of the dynamic nature of competencies. There are several reasons to create and maintain the DCMS in the organization [4]: (1) they can provide identification of the skills, knowledge, behaviours and capabilities, needed to meet current and future personnel selection needs, (2) they can focus the individual and group development plans to eliminate the gap between the competencies requested by a project, job role, or enterprise strategy and those available. The existing systems to not support dynamic aspects of competence management enough. Systems focused on providing the tools for recording competence profiles and use them to select the employees to the project.

The first part will include an overview of literature related with competencies management in a typical organisation. Next, the components and functions of the DCMS will be defined. After that we will clear up the concepts related to the DCMS and its relation. The next part will showcase dif-

The part of this work was supported by Project „Platforma Informatyczna TEWI", nr. POIG.02.03.00-00-028/09,

ferent approaches regarding modelling the dynamic nature of competencies. At the end, the social network approach to dynamic nature of competencies modelling will be discussed.

II. Competence in the Organization

According to [10] the competence is a observable or measurable ability of an actor to perform a necessary action(s) in given a context(s) to achieve a specific outcome(s). The competence information (competence profile) is a data about a competence that may be aggregated for communication among individuals, organizations, and public administrations. The more detail discussion about competence notion can be found in [5] or [16]. The history and background of standardization in this area and research project are covered in [7].

Generally speaking the competencies in the organization are placed on unit (organizational), collective (team) or individual level [17]. The typical related competencies are presented on Tabele I.

TABLE I. TYPICAL COMPETENCIES IN ORGANISATION (BASED ON [13])

Unit Competence	Collective Competence	Individual Competence
- Knowledge Landscape - Knowledge Assets - Information Sharing - Push/Pull Power Balance - Synergy Creation	- Knowledge Sharing - Cultural Integration - Resources Utilization - Innovation - Management/ Leadership	- Result Orientation - Role Commitment - Continuous Learning - Networking - Creativity

Based on the [8] we can defined what qualities and capabilities the competent person or team require:

1. The domain knowledge empirical, scientific or a blend of both;
2. The experience of application (knowing what works) in different contexts;
3. The drive and motivation to achieve the goals and strive for betterment/excellence;
4. The ability to adapt to changing circumstances and demands by creating new know-how;
5. The ability to perform the requisite tasks efficiently and minimise wastage of physical and virtual resources;
6. The ability to sense what is desired and consistently deliver that at a high quality to the satisfaction of the end client.

While creating a IT system that will allow for competencies processing we must take into account the nature of competencies. The main medium for competencies caring is human, its work and personal development affects the parameters of given competence. Competencies are considered as an union of different components. Thanks to literature analysis (e.g. [7], [20], [21]) we can distinguish some components like: knowledge, skills, experience, etc. The important issue is the question of whether competence is a binary quality or not. According to [7] in natural language, and in other domains such as law and biology, competence is seen as binary, someone is either competent or not. In the educational domain, however, the competence can be graded on a scale, and that it can have degrees or "dimensions".

Competence can get gradually stronger, in a situation where surroundings affect and stimulate its components. For example, we acquire new skills in a training session or while working (e.g. software developers programming everyday). Competence (its level) can also degrade. The most common reason for it is not using the given competence in everyday work. The other is thanks to technology progress which makes the components of competence outdated. We can distinguish different relations between competencies which affect the interaction between them. Increasing competence in a certain competence group (e.g. communication) can affect the increase of other competencies (e.g. sales of products). Next issues regarding competence processing in an organisation start to show up when we take a look from a company's perspective. From the company's point of view, certain competencies are created only by combining the competencies of a greater number of employees. The complexity of these combined competencies is too great for a single person to obtain this kind of competence.

III. Dynamic Competencies Management System Definition

Based on literature analysis we can specify the following functions of the Dynamic Competence Management System (DCMS): employment planning, recruitment, trainings, raising work efficiency, personal development, managing key competencies. The literature current thinking is that competence management can be organized according to four kinds of mutual related processes [1]: competence identification, competence assessment, competence acquisition, competence knowledge usage.

Currently in Competence Management Systems, the analysis of changes happening in competencies by time is much confined. Employee, while working in a project group, develops his/her soft competencies and acquires experience and knowledge that later result in hard competencies. At the same time, some competencies (when not used) can decline. This process leads to a rapid obsolescence of employee's competencies profile. From the point of view of organisation, the process of managing competencies is dynamic because indicators comprising certain competencies constantly change thanks to employee's and its surroundings actions in the organisation. Building the DCMS is a complex process. The problem lies in a complex nature of competencies. Just like knowledge, competencies are created in human mind and are manifested by doing intellectual or manual operations correctly.

The core components of typical DCMS are [4]: competence model management, departments management, job roles management, learning object management, employees management, projects management. This functions allow to the competence knowledge model manipulation for insert-

ing, updating and deleting ontology data – knowledge model mapping on competence. The system also support functions for creating, updating and deleting a relationship between two competencies, a job assignment and an association between a learning materials and a competence [4].

Typical tasks of module of competencies processing system include (based on [4]):

- Finding competencies gaps between the existing and the desirable employee's competencies according to the position that he/she occupies, the related function is skill gap analysis for a future position (succession planning),
- Aggregation of individual competencies to the group's level - it is allows to estimated the average organization's/department's/team's proficiency level. This function calculates the average level of proficiency possessed by the employees for all competencies in the competence model for a specific department.
- Estimating the costs of acquiring competencies - it gives employee possessed proficiency level and the corresponding cots (e.g. time duration) for every competence since he/she was hired. The function 'optimizing the process of competencies transfer' is looking for minimal cost. In some cases we need to invest in employee by using function find learning materials for employee/project.
- Building projects groups - is a function that finds an employee who is an expert for a given competence on required level related to the project requirements (find best fit employees for project' function). If there is still no employee, then the system searches for an expert for other competencies (and comparing employees abilities), which are related somehow to base competence.

Practical problems to solve:

1. Discussion and analysis of competencies nature to determine the methods of analysing and estimating them for a certain employee.
2. Proposal methods for manipulating the structures of competencies.
3. Proposal methods for building the competencies profile for given employee.

Corresponding research problems:

1. Mathematical methods of processing and describing competencies.
2. Describing knowledge with a model (e.g. OWL ontology) regarding the acquisition of competencies.
3. Modelling the environment of competencies acquisition (social networks, viruses, network game theory).
4. Modelling and optimising network models of competencies.

IV. System nomenclature

To define the Dynamic Competencies Management System (DCMS) we need to establish axioms related to building the system:

A. Organisation

Has a global objective which is maintaining a position on the market. It is possible only by developing owned core competencies. Core competencies are competencies which are unique for the company and help to build a competitive advantage. Currently, in a dynamically changing environment, competitive advantage is not only decided by the fact of owning this kind of competencies but also their high levels. In a knowledge-based economy, employees (and their competencies) and intangible resources of the organisation (e.g. patents) are the main elements of its assets. The key to correct operation of organisation is to effectively manage the process of transferring knowledge which will use its assets in the most effective way.

B. HR Department

A department in organisation responsible for competencies audit and management. The work of HR department is based on using dynamic competencies management system.

C. Knowledge worker

Knowledge worker thanks to intelligent operations takes part in project's tasks assigned to him/her. Knowledge worker enhances his/her competencies by taking part in projects and cooperating with other employees (which are willing to share knowledge and they have higher competencies), attending training and self-study. Knowledge worker is described by competence profile which includes: possessed knowledge, competencies and their levels. Moreover, every knowledge worker is also described by his/her individual objectives, cognitive and social characteristics.

Worker has different roles assigned in different projects. Assigning a role requires to possess a certain set of competencies (set of competencies) on a certain level. By assigning a certain position to a worker his minimal competencies are being described.

D. Project

Conducted by knowledge workers. Each worker has a role in the project. Project is being created to solve a certain task, which is composed from sub-tasks. The condition for conducting a task (sub-task) is to have the required competencies on a certain level. Project always involves working in a group with other knowledge workers.

E. Competencies bank

It is distinguishing for every company. It contains a set of all competencies in an organisation (both possessed competencies and those planned to obtain). All competencies are related to each other and they form up a competencies network in a company. Each competence has a certain accumulated level for an organisation. The strategic purpose of organisation's operation is to achieve a desired level for each competence placed in competencies network. Every competency is described by description, and related descriptors (see Table II.)

TABLE II. EXAMPLE OF EVALUATION MATRIX FOR *COMMUNICATIVENESS* COMPETENCE (BASED ON INSPEO.COM SYSTEM)

Competence name	Communicativeness
Definition	Transfer to other information in a clear and understandable way, as well as listening and clarification to what others are saying to us.
Descriptors (1-5 Likert scale):	A. He/she speaks in an understandable way A.1: He/she is expressed in a vague and difficult to understand. No form of communication adapted to the situation and audience whatsoever. A.2: Sometimes he/she has problems with the formulation of clear and concise expression, even in standard situations. A.3: He/she speaks in a clear, concise, and keeps the topic of conversation. A.4: Even in the case of complex issues and in new or challenging situations (time pressure, challenging the audience), is expressed clearly and precisely. Adjusts the way of expression to the situation and audience. A.5: He/she is expressed in a understood manner, even on specialized topics. Communicate to the other their knowledge on how to communicate effectively. Creates and implements the rules of good communication practices. B. He/she ensures that the message was understood by the public … C. Encourages others to share their opinions … D He listens to speeches of his callers / listeners … E Knows the rules of proper written communication … F. It strengthens and validates their content through body language (posture, gestures, facial expressions, distance) …

F. Competencies network

A graph structure showing relations between competencies. Adding new competence to the network requires to make a relation with a competence already existing in the network. Usually it may involve creating new competencies to maintain/create relations with competencies that are already in existence. All competencies in the network must be connected.

G. Competencies catalogue

Description of positions, including competencies and their minimum level required to work on certain positions.

H. Competencies audit

A process that focuses on establishing what competencies and on what level, from the competencies bank, a certain employee possesses. Competencies audit is conducted by completing different psychometric tests, chats with a qualified assessor or an outside certificate that can confirm competencies.

I. Process of acquiring competencies

Process of acquiring competencies consists of different kinds and types of knowledge (ability) transfer to achieve a certain knowledge/experience for an employee and thus allows for achieving good results and reactions regarding certain competence. The transfer occurs between employees or between employees and dedicated systems.

J. Knowledge status

Set of certain areas of knowledge, measured with special equipment, that are included in the certain competence.

K. Core competencies

A set of competencies which are essential for an organisation to work.

L. Organisation graph

Shows relations between workers that occur in certain organisation. It changes over time.

Typical organization is composed of many different departments. In the context of competency management HR department is the most important, because it is responsible for storing information about the competences of employees in the competencies bank in form of the competence profiles. It also performs regular audits of competence, which provides the information necessary to update the employee's competency profile. Audit examines different areas of competence of the employee's knowledge and determine their constitution (knowledge status). All profiles are stored in the bank's responsibility. The structure of competence bank is a graph structure mapping network of competencies. HR department also manages a catalogue of competencies, which is used when hiring new employees.

When the project is coming up the HR helps is project stuff based on an analysis of the project required competencies and the competencies already possessed by the individual employees. When deciding on the allocation of staff to the project , the one must also take into account the organizational structure of the company, expressed as a organisation graph.

In addition, the HR department supports and manages the process of competencies acquiring. In this process, the key issue is to ensure an adequate level of the core competencies in the organization and the desire to cover the competence of the entire competencies network.

V. DISCUSSION OF THE DYNAMIC COMPETENCIES MODEL

In previous sections of the article it was discussed that the competence of a person changes and is subjected to many factors. Thus, the competence should be considered as a dynamic system that depends on many factors, mainly related to time. The competence of a person can be acquired in the process of training or strengthen during work which requires using this competence. Moreover, the competence can be transferred from others while working collaboratively. The competence of a person can also decline while he or she is not actively using it for certain period of time. The pace in which the competence is acquired, strengthen, transferred or declining usually is non-linear and dependent on many fac-

tors, like the nature of competence, its structure, context, current state and individual qualities of a person.

Like any dynamic system the competence can be represented by the set of its states variables values. State variables represent different pieces of a person's competence, like pieces of knowledge, information and skills, thus the competence can be seen as the function of several time-based arguments, such as:

- Time of training (acquiring).
- Time of working actively using competence (strengthening).
- Time of inactivity (decline).
- Time of team work/problem solving (transfer).

The proposed model should reflecting the structure of the competence and represent it as a set containing skills and knowledge existing in a certain domain. The model should be aware of context of competence and be able to reflect relations existing between them (ex. composition, similarity, etc.). There are many researches on this subject that propose well elaborated models [3], [29]. These approaches usually focus on providing exact models of different professional domains for human resource description or training planning. The intention of this model is to extend regular competence model by adding fuzzy measures describing the level at which a person mastered certain competence. The description of personal competence will use fuzzy representation of set to precisely show the "strength" of every element of competence by setting quantitative value of degree of membership for every element of the competency set. This approach will allow performing quantitative analysis on personal competences (ex. level of meeting competence requirements by a person, comparison of competence of two people for staffing purposes etc.). There are several works representing this "fuzzy" approach to competence modelling (e.g. [17] ,[27]) but, in turn, they lack the possibility to map complex relationships between competences.

The next step of the work is to elaborate the method for evaluation of the level of competence basing on analysis of a person portfolio (analysis of training history and professional experience in order to assess the value of the level of competence). This analysis will take into account the phenomenon of "learning curves", which assumes non-linear pace of knowledge and skill increase during work and learning process. On the last stage the model for group competence will be elaborated. The model will allow representing competence of the whole organization consisting of many individual professionals. This method will cover topics such as: aggregation of the level of competence, complementarity, competence domain coverage etc.

The competence model assumes fuzzy representation of set to precisely show the "strength" of every element of competence by setting quantitative value of degree of membership for every element of the competency set [17] ,[27]. The model will focus on modelling the process of competence development, which occurs during training and professional work. In the case of the fuzzy competence set this process reflects in rising the degree of membership for the element representing the competence under development [19].

Known studies on this subject assume linear relationship between the increase in the competence strength and the time spent on training. On the other hand, studies in the domain of cognitive science show that the learning process is non-linear and goes according to different "learning curves". Thus, the main goal of this work will be to develop the formal competence extension model, which will take into account the idea of "learning curves" in order to reflect non-linear relationship between time of training and competence strengthening. The formal model of non-linear competence extension will be then used to develop the method for competence extension cost analysis. This method bases on the assumption that extension of personal competence requires effort that takes some time, which can be translated into financial cost by introducing cost factors.

VI. Social Network Approach to Dynamic Competencies Management

Within the organization social networks can be identified, which are related to the flow of information and exchange of knowledge. Studying structures within an organization can be considered in terms of improving the information flow mechanisms and the identification of key members of the organization. Social network analysis methods can be used that make it possible to determine the quantitative parameters of the network, identifying community and relationships within the network structures [28]. This type of research involves both static properties of networks and their evolution over time.

Social networking is also used for measuring competence, which may relate to specific network segments, competence of separate groups and members of the community. While the subject matter of competence is usually considered both static and focused on the analysis of individual units, in this article it relates to the recognition of dynamic problems which can include movement of competence in the organization. That concept includes flow modeling competence within social network using diffusion models derived from the field of epidemiological studies which in recent years have been developed in the direction of social networking and viral marketing [30]. This is used to identify trends related to the prediction the range of diffusion, diffusion models and the identification of network nodes, which should be contacted by an initial infection [9]. The analysis of diffusion processes is based on in epidemiological models, linear model threshold or independent cascades model and branching processes [12].

In the area of competence-oriented modelling, a number of areas can be identified where the diffusion processes take place when acquiring competence. It is therefore a possible

link between the methods of social network analysis to identify key members of the team, whose competence is worth investing. Taking into account the mechanisms of diffusion a flow of competence in an environment can be specified where there is exchange of knowledge among team members and use the multidimensional approach [11]. Competence can be treated as information diffusion processes and can move between team members. The number of contacts between team members is conducive to the flow of competence and invests in workers for whom the identified high communication activity can promote the exchange of information to a greater degree than improving the skills of workers whose social activity is reduced.

The proposed concept of using methods of social network analysis and modelling of adaptation mechanisms of diffusion processes can find a number of applications in the analysis of competences and is reflected in both the theoretical and practical areas related to this topic.

VII. Conclusion

The construction of the system, with implementing the new approach to competence management, requires an in-depth conceptual phase. This article was devoted to this goal. The dynamic aspects of competencies level fluctuating have to be implemented with the proposed models, data structures and system components. The most important observations from the paper are:

- Typical tasks for competencies processing include: finding competencies gaps, aggregation of individual competencies to the group's level, estimating the costs of acquiring competencies, building projects groups
- Competencies bank, catalogue and network are the main elements that store information about competencies of employees in the organization.
- The competence level is changing in time during following situations: training (competence acquiring), active using (competence strengthening), inactivity (competence decline), team work/group problem solving (competence transfer).
- Transfer of competence in the group through the realization of tasks within a project can be analysed using the tools and methods of social network analysis.

Acknowledgment

The authors would like to thanks for close cooperation with *Inspeo.com* company - Polish frontrunner in tools development for competence audit and assessment.

References

[1] G. Berio, M. Harzallah, “Knowledge management for competence management, ” Journal of Universal Knowledge Management, vol. 0, no. 1, pp. 21-28, 2005.

[2] X. Boucher, E. Bonjour, N. Matta, “Competence management in industrial processes. Editorial,” Computers in Industry, vol. 58, no. 1, pp. 95-97, 2007.

[3] T.Y. Chen, “Using competence sets to analyze the consumer decision problem,” European Journal of Operational Research, vol.128, no. 1, pp. 98-118, 2001.

[4] F. Draganidis, P. Chamopoulou, G. Mentzas, (2008),"A semantic web architecture for integrating competence management and learning paths", Journal of Knowledge Management, vol. 12, no. 6, pp. 121-136, 2008.

[5] F. Draganidis, G. Mentzas, "Competency based management: a review of systems and approaches," Information Management & Computer Security, vol. 14, no. 1, pp. 51-64, 2006.

[6] E. García-Barriocanal , M.-A. Sicilia, S. Sánchez-Alonso, “Computing with competencies: Modelling organizational capacities, ” Expert Systems with Applications, vol. 39, no. 16, pp. 12310–12318, November 2012.

[7] S. Grant, R. Young, “Concepts and Standardization in Areas Relating to Competence” International Journal of IT Standards and Standardization Research, vol. 8, no. 2, pp. 29-44, 2010.

[8] G. Hessami, M. Moore, “Manage Competence Not Knowledge,” in Integrated Systems, Design and Technology 2010, Berlin Heidelberg: Springer-Verlag, 2011, pp. 227-242.

[9] O. Hinz, B. Skiera, Ch. Barrot, and J.U. Becker, “Seeding Strategies for Viral Marketing: An Empirical Comparison,” Journal of Marketing, vol. 75, no. 6, pp. 55-71, January 2012.

[10] ISO/IEC TR 24763 Information technology -- Learning, education and training -- Conceptual Reference Model for Competency Information and Related Objects, ISO standard, 2011.

[11] J. Jankowski, S. Ciuberek, A. Zbieg, R. Michalski, “Studying Paths of Participation in Viral Diffusion Process,” in Proc. of 4th International Conference on Social Informatics, SocInfo 2012, LNCS, 7710, pp. 503-516, 2012.

[12] J. Jankowski, R. Michalski, P. Kazienko, ”The Multidimensional Study of Viral Campaigns as Branching Processes,” in Proc. of 4th International Conference on Social Informatics, SocInfo 2012, LNCS, 7710, pp. 462-474, 2012.

[13] G. Kayakutlu, G. Büyüközkan, "Effective supply value chain based on competence success", Supply Chain Management: An International Journal, vol. 15 no. 2, pp. 129 -138, 2010.

[14] K. Koeppen, J. Hartig, E. Klieme, and D. Leutner, “Current Issues in Competence Modeling and Assessment,” Zeitschrift für Psychologie / Journal of Psychology, vol. 216, no. 2, pp. 61-73, 2008.

[15] E. Kushtina, O. Zaikin, P. Różewski, B. Małachowski, “Cost estimation algorithm and decision-making model for curriculum modification in educational organization,” European Journal of Operational Research, vol. 197, no. 2, pp. 752-763, 2008.

[16] F. D. Le Deist, J. Winterton, “What Is Competence?,” Human Resource Development International, vol. 8, no. 2, pp. 27-46, March 2005.

[17] G. Pépiot, N. Cheikhrouhou, J.-M. Fürbringer, R. Glardon, “A fuzzy approach for the evaluation of competences,” International Journal of Production Economics, vol. 112, no. 1, pp. 336-353, 2008.

[18] P. Rauffet, C. Da Cunha, A. Bernard, “Conceptual model and IT system for organizational capability management,” Computers in Industry, vol. 63, no. 7, pp. 706-722, 2012.

[19] P. Różewski, B. Małachowski , “Competence Management In Knowledge-Based Organisation: Case Study Based On Higher Education Organisation,” in: D. Karagiannis and Z. Jin (Eds.): KSEM 2009, LNAI 5914, pp. 358—369, 2009.

[20] D. Sampson, D. Fytros, “Competence Models in Technology-Enhanced Competence-Based Learning,” In: H.H. Adelsberger et al. (eds.): Handbook on Information Technologies for Education and Training, 2nd edition, Springer-Verlag, Heidelberg, pp. 155–177, 2008.

[21] R. Sanchez, “Understanding competence-based management: Identifying and managing five modes of competence,” Journal of Business Research, vol. 57, no. 5, pp. 518–532, 2004.

[22] J. Sandberg, “Understanding of Work: The Basis for Competence Development, ” in International Perspectives on Competence in the Workplace, C.R. Velde (ed.), Berlin Heidelberg: Springer-Verlag, 2009, pp. 57-85.

[23] G. Szulanski, R.J. Jensen, “Presumptive adaptation and the effectiveness of knowledge transfer,” Strategic Management Journal, vol. 27, pp. 937–957, 2006.

[24] V. Tarasov, "Ontology-based Approach to Competence Profile Management," Journal of Universal Computer Science, vol. 18, no. 20, pp. 2893-2919, 2012.
[25] V. Tarasov, T. Albertsen, A. Kashevnik, K. Sandkuhl, N. Shilov, A. Smirnov, "Ontology-Based Competence Management for Team Configuration", in Proc. of HoloMAS 2007, LNCS, Springer, Vol. 4659, 2007, pp 401-410.
[26] D. Ulrich, "Intellectual capital = competence x commitment," Sloan Manage. Rev., vol. 39, no. 2, pp. 15-26, Winter 1998.
[27] H-F. Wang, C.H. Wang, "Modeling of optimal expansion of a fuzzy competence set," International Transactions in Operational Research, vol. 5, no. 5 , pp. 413–424, 1995.
[28] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge University Press, 1994.
[29] P.L. Yu, D. Zhang, "Optimal expansion of competence sets and decision support," Information Systems and Operational Research, vol. 30, no. 2, pp. 68–85, 1992.
[30] N. Zekri, J. Clerc, "Statistical and Dynamical Study of Disease Propagation in a Small World Network," Phys Rev E., vol. 64, pp. 056115, 2001.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1287–1291

Outsourcing of knowledge in change and renewal processes

Małgorzata Sobińska, PhD
Department of Information and Knowledge Management Wroclaw University of Economics
E-mail: malgorzata.sobinska@ue.wroc.pl

Jakub Mierzyński, MSc
Department of Information and Knowledge Management
Wroclaw University of Economics
E-mail:
mierzynski@poczta.fm

***Abstract*–This paper is an attempt to present the concept of organizational transformation with the use of the strategic renewal theory oriented towards organizational learning. It is indicated that enterprise renewal processes constitute the basis for organizational changes enabling evolutionary development towards implementing enterprise learning mechanisms. A new aspect occurring in this presentation is the discussion of the benefits which might be brought for the organization by the use of external resources (outsourcing tools) for the purpose of enhancing renewal processes.**

I. Introduction

IN THE last decades, the dynamics of the turbulence in environmental changes occurs simultaneously with the development of the latest technologies (and in particular along with the considerable acceleration in the development of the digital technologies). Therefore, entrepreneurs are forced to focus their activities on various types of transformations primarily aimed at the organization's keeping pace with the changing environment. One of interesting forms of such transformations of the recent years is the strategic renewal concept [4], [3]. Cognitive components of the strategic renewal concept based on which enterprises strive for transforming into learning organizations are described in the study [14]. The authors additionally point to the possibility to intensify renewal processes by using the knowledge coming from organization's external resources through implementing tools such as outsourcing, and in particular Knowledge Process Outsourcing.

According to Prahalad and Krishnan, "in the new innovation era, it is the capability of introducing and improving flexible transparent business processes allowing continuous changes in selecting resources R=G in the interest N=1 that will determine the advantage of companies" [12, p. 34] (these principles will be discussed in the further part of the presentation).

For the sake of this idea, organizations should give up storing all needed resources and initiate the implementation of programs for accessing specialized global suppliers, and one of the cheap and effective manners of sharing/using resources is outsourcing.

II. Strategic Renewal in the Context of the Classical Theory of Change

Strategic renewal is an interesting concept of evolutionary development of organizations of the recent years, which refers to various aspects of the enterprise's operation. The foundation of renewal is the theory of change but the renewal itself has not been unambiguously defined to date. It is frequently identified with the theory of change in the relevant literature [3, p. 13]. Among the first who made an attempt at distinguishing change from renewal were R. Agarwal and C. Helfat [1, p. 281]. They defined the notion of renewal as "the process, content and effect of transformation or exchange of the characteristics of organization which have a significant impact on long-term operation perspectives." Following this interpretation, it should be assumed that renewal is a much broader phenomenon than change and refers to the foundations of survival or development of the organization. Change, in turn, is a mere component of renewal [4, p. 33]. Such a viewpoint was adopted also by J. Skalik, who defines strategic renewal as a fundamental change with a broad scope, one being a form of response to fluctuations of the environment and all phenomena inside the organization that decrease its effectiveness level [15, p. 18].

The common plane of change and renewal is their process. While in the case of change its process nature does not arouse doubts, in the case renewal the process aspect is related with elaborating new attributes the significance of which is strategic for the organization [4, p. 35]. However, renewal ought to be understood as a process whereby changes generating qualitatively new bases for implementing modern enterprise development concepts occur.

Both change and renewal are based on a common plane, that is the process. While in the case of change its process nature does not arouse doubts, in the case of renewal the process aspect is related with elaborating new attributes the significance of which is strategic for the organization [4, p. 35]. Renewal ought to be understood as a process whereby changes generating qualitatively new bases for implementing modern enterprise development concepts and innovation occur.

It is indicated that renewal processes cannot be one-time processes in the moment of an organizational crisis but they

978-83-60810-53-8/$25.00

should be continuous processes which systematically renew the enterprise by bringing it to the state of equilibrium and at the same time ensuring its development.

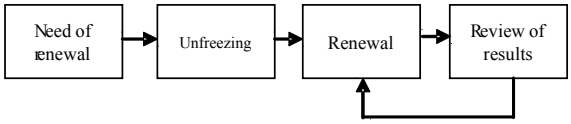


Fig. 1. The model of the infinite renewal process with feedback
Source: own work

In the ideal renewal model presented in Fig. 1, the once initiated renewal process should last infinitely in the organization. After diagnosing the need of renewal (internal impulse) in the enterprise, the permanent unfreezing stage occurs only once and then the organization becomes a plastic body in the constant unfreezing state. The infinite renewal process begins in this stadium, being interrupted only with the review of results.

A significant complementation of the above model is the aspect of renewal initiation. It is maintained that the initiator of the renewal process should be an impulse coming from inside the organization. Only such a stimulus can be strong and effective enough to commence an effective renewal process. The probability of renewal initiation effectiveness by forces coming from the environment of the organization seems to be negligible since each of such actions will be treated as an obligation rather than as one's own action, which has an advantage in the form of significantly weaker resistance forces than in the case of external initiation from the very beginning.

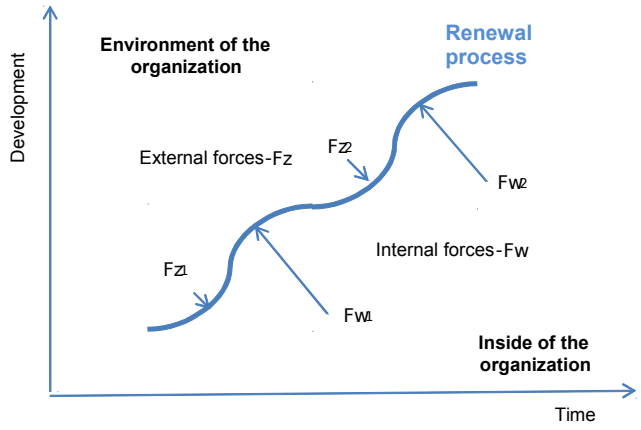


Fig. 2. The renewal process against forces from the inside and the outside of the organization
Source: own work

When discussing the component of the renewal process initiation from inside the organization, it needs to be indicated that such an impulse is possible only when the organization has reached an appropriate organizational maturity level. In enterprises with no or a low organizational maturity level, the renewal impulse will never occur.

Strategic renewal in the presented model (Figure 1 and 2) is a method of continuous and systematic introduction of changes and of formation of the organization without the necessity to defeat change resistance forces every time. At the same time, it introduces a new quality of the organization's operation. Renewal imposed by an internal impulse becomes the property of the organization's members. Introducing innovative solutions seems to be difficult without renewal, and even impossible in numerous cases.

III. Mechanisms of Organizational Learning of Enterprises as the Renewal Process

Information and knowledge are currently the basic resources in the micro- and macro-economic approach. The speed of obtaining information and the ability to learn fast are among the main factors ensuring competitive advantage. The above principles are used by the learning organization concept. Learning organizations broaden their creative capacities such that they can create their future effectively. Working in such organizations involves the continuous knowledge improvement process and the use of experiences rather than mere performance of tasks [14]. A common term for this type of enterprises is an intelligent organization. The basis for the operation of this type of companies is building a community of specialists (knowledge workers) who communicate well with each other and are capable of continuous transformation of the enterprise, its products and themselves for the purpose of satisfying market requirements and challenges formulated by the society. Intangible resources, including in particular knowledge, are most important in such an organization [10, p. 103].

The learning organization concept is a response to the continuously changing environment, technological changes and growth of employee competences and requirements; at the same time, it is closely related to the enterprise innovation strategy, which is possible owing to a properly pursued enterprise renewal process.

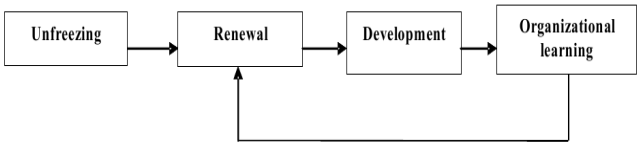


Fig. 3. The renewal process model towards organizational learning.
Source: own work

The learning organization concept and, literally, organizational learning is a natural consequence of strategic renewal.

Innovativeness is a creative process and, like every process of this kind, it begins with an idea, is followed by incubation, and ends with illumination, that is a qualitatively new idea that solves the formulated problem. From the viewpoint of learning organizations, such a process begins with implementing organizational changes, is followed by strategic enterprise renewal, and ends with implementing structural mechanisms of learning organizations (Fig. 4) [6].

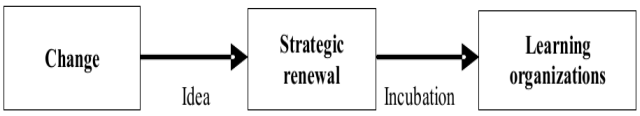


Fig. 4. The implementation model of the organizational learning process – from changes to strategic enterprise renewal. Source: [6]

It needs to be indicated that implementing structural changes in the organization that facilitate the enterprise development towards organizational learning is a component of strategic renewal. Strategic enterprise renewal is a phase of "incubating" organizational development, the effect of which should be qualitative improvement of organizational effectiveness of the enterprise that facilitates information flow management in the organization called synergy or supercompensation effect. The enterprise capacity to implement renewal processes is a necessary condition for the enterprise to be able to build its organizational potential towards organizational learning [6].

IV. Outsourcing Tools – Current Applications and Trends

According to Prahalad and Krishnan [12, pp. 31-31], "the principal significance for firms to be ahead of their competitors is a sense of the achievable and the use of innovations coming from laboratories of institutions operating on the global scale and small new firms… The nature of financial, human and technological resources outgrows the firm and its legal boundaries. Today, resources are global. Attention needs to be focused on access and influence rather than on ownership and control." Prahalad and Krishnan provide the following principles to be followed by the organizations which wish to build competitive advantage:

N = 1 and

R = G.

N=1 states that "value is based on unique, personalized experiences of consumers." That is, even companies serving 100 million consumers need to focus on individuals.

R=G, meanwhile, argues that since no company can hope to satisfy the varied expectations of so many consumers, it must diversify how it operates. "All firms will access resources from a wide variety of other big and small firms—a global ecosystem," write Prahalad and Krishnan. In other words, companies' internal focus should be on gaining access to resources, not necessarily owning them [12, pp. 27-28].

Globalization provides enterprises with easier access to more numerous labor force resources and employment of a greater number of employees with a more specialized education, which contributes to increasing the quality of the provided work and the number of innovations in many cases. Outsourcing and offshoring development results from the fact that more and more organizations strive for improving their competitiveness by relocation of goods and services. Outsourcing and offshoring more and more frequently concern knowledge-based processes aimed at supporting innovative operation of the organization [5, p. 48].

Oshri, Kotlarsky and Willcocks, who have been observing the outsourcing market for years now, notice that various new types of global supply models have been emerging. The major difference between the models lies in:

- whether the function is used by a business unit dependent on the parent company or an external supplier (or jointly by both entities), and
- whether the function is used by the enterprise on-site (in the registered office) or off-site (outside the registered office – in the country where the organization is based (onshore), in a neighboring country (nearshore) or in a distant location (offshore) [9, p. 25].

The ongoing growth in the outsourcing market has major implications for management. Organizations have to develop new capabilities supporting the ever-changing business models in their sourcing engagements. Understanding how and where value is created in sourcing engagements becomes another challenge. Dependency on external partners has increased. Providers of outsourcing services are becoming more aware of clients' growing demand to realize innovation and transformation from outsourcing engagements and they are refocusing their efforts to deliver value to clients by improving their performance management systems and by seeking to extend their offerings [17, pp. 1-2].

One can observe a trend towards outsourcing relationships becoming increasingly managed and leveraged as strategic assets. New forms of sourcing deals are required for collaborative innovation to succeed.

In today's knowledge-based economy, one of the major sources of competitive advantage has been the ability of the firm to transfer external knowledge efficiently and effectively. Knowledge transfer can be defined by as activities of exchanging explicit or tacit knowledge between two agents, during which one agent receive and apply the knowledge provided by the other agent. The agents could be an individual, team/department or an organization. In the literature, knowledge transfer has been given various but related labels such as knowledge sharing, knowledge flows, knowledge acquisition and knowledge mobilization [2, p. 1].

One of the knowledge acquisition options by taking advantage of external resources is using such centers as Knowledge Process Outsourcing. KPO centers are established dynamically worldwide. A trend proving the increasing comprehensiveness of business processes served from Poland is also noticeable. At the same time specialization of many centers increases – consisting in serving more advanced processes (market research, business analyses, etc.). Thus, a "subsector" called knowledge process outsourcing is being slowly formed in Poland. On the global scale, it is characterized by a significantly faster growth than the entire BPO (Business Process Outsourcing) sector, which means a chance for Poland to develop in this area mainly due to the held resources of qualified employees, appropriate communication infrastructure and political stability. Owing to the qualified staff and good academic centers, and not only low labor costs, Poland has been becoming an attractive place for locating this type of projects for a few years now [13, p. 46].

V. Outsourcing of Knowledge as an Accelerator of Changes Towards Enterprise Renewal

Outsourcing as a strategic component of renewal should facilitate organizational learning, and for this purpose the organization deciding to outsource must set clear goals for outsourcing collaboration. Collaboration in a strategic sourcing context is pro-active working together and risk sharing, in flexible integrated ways, to achieve high performance on longer, mutually rewarding commercial goals [17, p. 129].

Whitley and Willcocks suggest four fundamental practices behind effective collaborative innovation [based on 17, pp. 143-144]:

- Leading. Leadership shapes and conditions the environment in which requisite contracting, organizing and behaving can occur. Leadership also changes the approach to risk in order to share and manage down risk and manage in opportunity.
- Contracting. New forms of contracting are required to secure successful collaborative innovation. Such contracts share risk and reward in ways that provide incentives for innovation, collaboration and high performance to achieve common goals.
- Organizing. Organizing for innovation requires more co-managed governance structures and greater multi-functional team working across those organization and people responsible for delivering results. Team working now requires the ability to collaborate within a client organization, between client and supplier and between suppliers in multi-supplier environments.
- Performing. Leading, contracting and organizing in these ways provides incentives to change existing modes of performing and enables collective delivery of superior business outcomes. Collaborative innovation is most effective when it generates high personal, competence-based and motivational trust among the parties. High trust is a key component and shaper of the collaborative, open, learning, adaptive, flexible and interdependent performance style required and open communication that help in knowledge diffusion.

These four elements have a temporal sequence as shown in Figure 5.

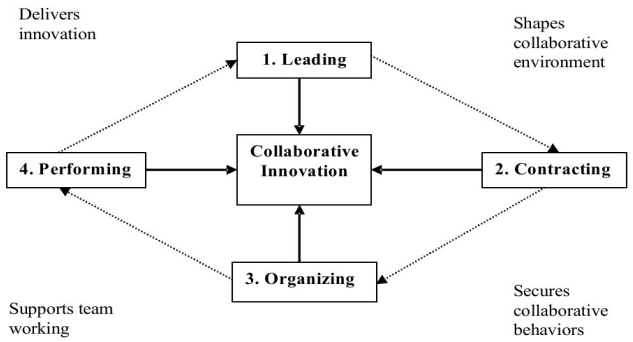


Figure 5. The process of collaborative innovation
Source: [17, p. 144].

Knowledge management in IT outsourcing relationship should enable creation and utilization of intellectual capital, that, in Willcocks opinion, should be generated through the interplay between such essential elements as [9, pp. 111-112]:

- structural capital (that refers to the codified bodies of semipermanent knowledge that can be transferred and the tools that augment the body of knowledge by bringing relevant data or expertise to people),
- human capital (which represents the capabilities of individuals to provide solutions to customers),
- customers capital (that is linked to shared knowledge, or the value of an organization's relationships with the people with whom it does business)
- and social capital (for example: trust, loyalty and reciprocity within a community - the values created from social networks), which helps bring these elements together.

As outsourcing often disrupts and reduces social capital by disembedding people, systems, and institutional knowledge from the client organization more attention should be paid to cultivating social capital. Social capital can have a considerable impact on effective knowledge transfer between outsourcing parties. It allows outsourcing partners reduce cultural barriers, understand common goals and strengthen network stability and ties.

According to the authors, organizations should take advantage of outsourcing such that is supports renewal processes and organizational learning to the highest possible extent. For this purpose, on the one hand, the organization ought to ensure easy access to the required information to both parties to the contract, and on the other hand it should not burden employees with excessive formalization but it needs to stimulate creativity and involvement in activities related to achieving the goals of the outsourcing project/contract.

Concepts such as outsourcing can accelerate renewal process since:

- they contribute to innovation growth when they concern processes related to innovation activities;
- they facilitate the improvement of the quality and effectiveness of processes and services owing to access to better/cheaper resources;
- they enable the introduction to new markets (if the supplier comes from a different country);
- they enable expanding the network of relations – they increase the structural capital of the organization.

From the viewpoint of knowledge management, an attempt might be made to describe outsourcing as [11, p. 394]:

- a manner of acquiring specialized knowledge and skills which the organization does not have;
- a form of stabilization of the knowledge related to the operation of selected areas in the organization (if the organization cannot handle e.g. the fluctuation of the IT staff, outsourcing could in a way secure the organization against a possible loss of critical employees);

- a guarantee of keeping pace with technological development (in this case, the outsourcing contract should provide for appropriate terms and conditions imposing the obligation of continuous development and improvement of services on the service supplier);
- a replacement of the internal know-how with the same type of knowledge from the outside.

Selecting the appropriate sourcing model is one of the critical aspects in the outsourcing planning.

It is worth emphasizing here that the selection of the optimum organization operation model is conditioned on numerous factors, such as: the held intellectual capital, the sector of activities, the phase of the organization's lifecycle, etc. Hence, every organization needs to make an independent decision on which of the operation options can have the most beneficial impact on the enterprise development and monitor on an ongoing basis if the selected model is still optimum in the context of strategic renewal.

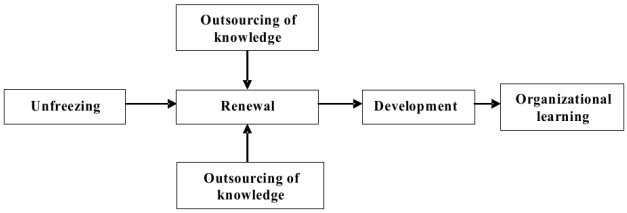


Fig. 6. The universal model of organizational learning through renewal and outsourcing of knowledge
Source: own work

The enterprise deciding on outsourcing ought to make an attempt to build such trust and create conditions facilitating knowledge sharing between employees and suppliers so that the organization's tacit knowledge could be co-created by external companies. It increases the probability that, along with broadening employee knowledge and experience, owing to continuous improvement, the outsourcing process will mature, which in turn will be reflected in the acceleration of organizational renewal processes.

VI. CONCLUSION

The execution of the strategic renewal process initiated inside the organization and pursued based on the exclusive use of internal resources of the enterprise is an appropriate action. However, it is indicated that the quality and thoroughness of renewal can be considerably increased by the application of external resources. In order to obtain the enhancement effect of the discussed process, it is suggested to use an outsourcing tool which has been effective in practice for years. Outsourcing, like many other modern management concepts, is a response of enterprises to the dynamically changing conditions of the environment as well as the new management trends that are being formed. Outsourcing is a complex enterprise management tool and its impact might involve numerous aspects of the enterprise operation.

Outsourcing creates a new type of strong correlations between partners, which do not arise from legal provisions. A successful collaboration cannot be ensured exclusively by contracts. Various unpredictable events might occur during the implementation of the project and therefore it is good when collaboration with the external supplier(s) is based on trust, respect and informal business coexistence principles.

REFERENCES:

[1] Agarwal. R, Helfat C., Strategic renewal of organizations, Organization Science Vol. 20, No. 2, March-April, 2009.
[2] Al.-Salti Z., Hackney R., Ozkan S., Factors impacting knowledge transfer success in information systems outsourcing, http://bura.brunel.ac.uk/bitstream/2438/4371/1/C74.pdf , (accessed: 26.11.2012).
[3] Banaszyk P., Cyfert S. Strategic renewal of a company (in Polish), Difin, Warszawa 2007.
[4] Bełz G., System zarządzania jako regulator odnowy i wzrostu przedsiębiorstw, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 2011.
[5] Ciesielska D., Offshoring of services. (in Polish), Wolters Kluwer Polska, Warszawa 2009.
[6] Cieśliński B.W., Mierzyński J., Nosek W., Model of renewal processes management - towards organizational learning (in Polish) [in:] Zmiana warunkiem sukcesu. Współczesne trendy i przeobrażenia metod i praktyk zarządzania w przedsiębiorstwach. ed. J. Skalik, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2012 – in the process of publishing.
[7] Kotlarsky J., Willcocks L., Special issue on global sourcing of business and IT services, Journal of Information Technology Teaching Cases (2012) 2, 57–60; published online 13 November 2012.
[8] Lacity M.C., Willcocks L.P., The Practice of Outsourcing. From Information Systems to BPO and Offshoring. Palgrave Macmillan. L, New York 2009.
[9] Oshri I., Kotlarski J., Willcocks L.P., The handbook of global outsourcing and offshoring. Second edition, Palgrave Macmillan Ltd. – Houndmills Basingstoke Hampshire (UK) 2011.
[10] Perechuda K., Zarządzanie wiedzą w przedsiębiorstwie, Wydawnictwo Naukowe PWN, Warszawa 2005.
[11] Perechuda K. Sobińska M., Information and knowledge management in IT outsourcing (in Polish), [in:] Systemy informacyjne w zarządzaniu przedsiębiorstwem, ed. J. Korczak, I. Chomiak-Orsa, H. Sroka, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2010 (s. 393-403).
[12] Prahalad C.K., Krishnan M.S., The New Age of Information (in Polish), Wydawnictwa Profesjonalne PWN, Warszawa 2010.
[13] SSC/BPO Sector in Poland, Report prepared on the request of the Association of Business Service Leaders in Poland, September 2010.
[14] Senge P.M., The Fifth Discipline (in Polish), Oficyna Ekonomiczna, 2006
[15] Skalik J., The key areas of strategic renewal of an organization (in Polish), [in:] Zmiana warunkiem sukcesu. Odnowa przedsiębiorstw – czego nauczył nas kryzys?, ed. J. Skalik, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2010.
[16] Sparrow E., Successful IT Outourcing. Springer, London 2003.
[17] Willcocks L.P., Lacity M.C., The new IT outsourcing landscape. From innovation to claud computing, Palgrave Macmillan Ltd. – Houndmills Basingstoke Hampshire (UK) 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1293–1299

Student Response to Educational Games – An Empirical Study

Urszula Świerczyńska-Kaczor
Jan Kochanowski University
in Kielce, Żeromskiego 5,
25-369 Kielce, Poland
Email: swierczynska@ujk.edu.pl

Jacek Wachowicz
Gdańsk University of Technology
ul. G. Narutowicza 11/12,
80-233 Gdańsk, Poland
Email: jacek.wachowicz@zie.pg.gda.pl

***Abstract*—In this article we explore students' experience with digital educational games. We analyze and discuss the factors which determine a student's perception of the educational benefits of game-based learning. The organization of the research is structured by the main question - how do variables such as player satisfaction, game features (e.g. the perceived quality of the educational content, the design) and enhancement in the player's educational process are interconnected. The study proves that games as an educational tool are assessed very favorably by undergraduate students of business and economics. Moreover game features are correlated with educational benefits and player satisfaction. A player satisfaction is also linked with enhanced learning.**

I. Introduction

SERIOUS games – intended to improve players' learning skills – are regarded to be a more effective educational tool than traditional lecturers when it comes to meeting the needs of the 'digital millennium generation'. The 'virtual generation' ('digital natives') learns differently compared to previous less technologically immersed generations (Pasin & Giroux 2011), although it is arguable that 'differently' means 'better'.

The number of publications referring to digital game-based learning has significantly increased in the first decade of the 21-st century (Hwang & Wu 2012). Nowadays the body of literature empirically or theoretically embracing the problems and issues connected with game-based learning is quite significant. Publications which clearly show the direct comparison between the different studies (the aims, participants, main findings), such as Pasin & Giroux (2011) or Wu et al. (2012), are particularly insightful and useful for understanding the scope of research that has been conducted in this field.

Therefore, due to the wide variety of forms (strategic games, solitary games, social games, board games, computer games etc.) the conclusions from different studies about the educational effectiveness of the game are not easily comparable. Most studies argue for the educational effectiveness of games, although different authors set different approaches for conducting their studies. For example Tao at al. (2009) proposed a model developed from the technology acceptance model, the expectation confirmation theory, and the agency theory. Other authors, Lin & Tu (2012) implemented the concept of means-end chain (MEC) to explore the value sought by players. Also in the literature researchers refer to different learning theories (behaviorism, cognitivism, humanism, constructivism) or they do not refer to any theory of learning at all (see – Wu et al 2012). Analysis of game-based learning is even more difficult due to the lack of a clear definition of game features. Games have game mechanics (which includes such elements as points or virtual gifts), game dynamics (e.g. status, reward, individual achievement and self-expression), an immersive environment which includes the rules, the story which outline the theme of the game, the embedded risks and competition (Derryberry 2007, Simões et al. 2013). A game's interactivity, which is the imminent aspect of a digital game (Rouse III 2005) is also difficult to assess and define. (Even as a website feature is difficult to measure due to the incongruence in the actual and the perceived interactivity of the website – Voorveld et al. 2011).

The study presented in this article is an exploratory study. So far business simulation games are not very popular in higher education in Poland and this presumption was confirmed by the empirical study presented below – only three participants of over 100 reported previous game-related experience. Do students perceive games as useful tools in formal education? Do educational games in higher education meet academic standards from the students' point of view? Although we would point out that we do not focus on answering the question whether or not a university's educational program can be substantially based on games, or if the games can replace the academic reading, class discussion and live lectures. We rather see that games can be useful tools for the purpose of introduction and invitation to more in-depth analysis.

II. Research Design

This article presents part of data and analysis conducted as part of a wider research project "e-Education within the social Internet". In the part of the project presented here two simulation games - Trade Ruler Game and Marketing Manager (connected with economic and managerial problems) - were tested for their educational benefits (see Table I).

This work was a part of the project 'Badanie statutowe 614564' Jan Kochanowski Univeristy in Kielce

We understand simulation games as to be games which have an embedded risk of losing or winning, are based on a backstory, and have game mechanics; on the other hand, they embed simulation (see Tao et al. 2012) One of the selected games – Marketing Manger - can be classified as a functional game around the specific topic of business. The other – Trade Ruler Game – can be classified as concept simulation in referring to a specific type of decision making (classification of the management simulation games - Pasin & Giroux 2011). We used games which are not time-consuming and can be completed within one-hour of play. This kind of game increases participants' full attention and motivation for the whole duration of the study. (Very complex games, such as multi-module business simulation games, need much more time and effort to play and therefore bring different challenges for educators).

In the present study we measured the constructs – the navigation of the game, the design, the educational content – as the player's perception-based construct. This means that the participants of the survey answered questions about their feelings and perceptions (e.g. they agree or disagree with the statement "I can easily find information which I need'). Most students played the games as an out of class task and then they filled out the questionnaires. We did not measure the progress which students made playing the games several times – instead we asked them to generally assess their feelings about using this new form of learning. We also did not include in the study data referring to the heterogeneity of the participants such as risk avoidance, general attitude to the university or interest in their studies.

We chose the constructs to measure arbitrarily, agreeing that they are under theorized and often have different meaning in the literature (with the concept of design as a prime example). In this study we wanted to 'capture' the students' general feelings and experience regarding game-based education.

In this empirical study we assumed that the game features – the educational content, the ease of navigation and the design of the game - influence player satisfaction and that satisfaction is linked to the player's enhanced learning (Fig. 1). Therefore the hypotheses are:

H1: The game features - the perceived value of the educational content, the perceived design of the game, and the ease of navigation and playing - enhance player satisfaction

H2: The features of a game and a player satisfaction are linked with the player's educational benefits.

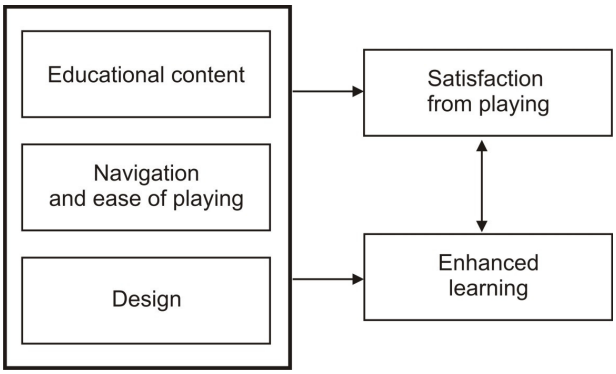


Fig. 1 The framework for empirical research

For this research, a questionnaire was developed to gain students' feedback (see Table II–VIII). The study was conducted in April 2013. A total of 208 filled questionnaires were received, 99% of the respondents were under the age of 25. All of the participants are business and economics students at one of three different Polish universities. In the study:

- for the Trade Ruler Game 106 participants took part - 72% of which were women,
- for the game Marketing Manager – 102 participants took part with the same proportion of women – 72%.

Very few – only three participants stated that they had previous experience with the game-based learning connected with economics or management.

III. Results and Discussion

Tables II-VIII present the main results and the structure of the survey questionnaire. Summarizing the results we can point out that:

1. The students' perception of the game Trade Ruler is affected by the language of the game which is not

TABLE I.
CHARACTERISTICS OF THE GAMES

Trade Ruler Game – available: http://www.nobelprize.org/educational/economics/trade/index.html	This is an educational game available at Nobel Prize official website. The theme of the game is based on the Heckscher-Ohlin trade theory. The player takes the role of the leader of one country and has to make decisions - depending on the country's resources– regarding production and trading with another country. After a few turns the player receives information about the score - calculated on the basis of the welfare of the country as a result of international trading. The language of the game - English
Game Marketing Manager http://www.nbportal.pl/pl/cw/gry/gry_decyzyjne/marketing	The game is available at the educational website of the National Bank of Poland. The task of the player is to choose marketing options for a car company: to decide to buy or not to buy a market research report, choose the client base, establish a delivery chain, etc. The player receives information on how the decisions affect the company's finances in areas such as sales, profits, and costs. A player's score is based on the total number of points given for the correctness of the player's business decisions. The language of the game - Polish

their native language. Almost 40% of the students stated that the game's language was a problem or a significant problem (table II).

2. Students assessed the educational content of the analyzed games very favorably (table III). The majority (over 80%) of the participants claimed that The Trade Ruler Game or Marketing Manager Game helped them to better understand the economic issue in focus (the trading between different countries or basic marketing strategic options).
3. The games analyzed were perceived by the players to be easy to navigate and easy to find the necessary information (table IV). In both games over 90% of the participants found the game easy to play, over 80% found information easily and the majority of the players found their score without problems
4. The design for both games was also well received (table IV). Interactivity was positively assessed by over 80% of informants in both games, graphic design over 55%, with the general quality of The Trade Ruler Game received an average score of 2.8 (on a scale of 1-4) and Marketing Manager Game-3.1.
5. Player satisfaction is high for both games (table VI). The majority – nearly 80% of players ranked their satisfaction at a 3 or 4 in Trade Ruler game (the average being 3.0). In the case of Marketing Manager the percentage was 90% (the average being 3.3).
6. Players perceived the games to be very useful and effective educational tools (table VII). They claimed that the games triggered their interest in economical or marketing issues (over 80% of participants), made learning more effective (over 85% of participants), helped to increased engagement in the subject (over 80%) and linked the learning process to a positive emotion (over 85% of participants).

In order to test the hypotheses we looked for correlation between variables (table VIII) and a weak or moderately positive correlation between some of the variables were found. Below we emphasizes the findings for the moderate positive correlation (R Spearman above 0.33, $p<0.05$).

TABLE II.
THE PLAYERS' ASSESSMENT OF HOW LANGUAGE AFFECTS PLAYING THE GAME

How much of a problem was the use of the English language in "The Trade Ruler" game in your playing of the game?	Not a problem at all	It was a problem	It was a significant problem
N=106	61%	30%	9%

TABLE III.
EDUCATIONAL CONTENT [TRADE/MARKETING MANAGER]

			Strongly disagree			Strongly agree	It's difficult to say	Negative opinion	Positive opinion	Average (1-4)
Code			1	2	3	4	-			
QE1	The game helps me to understand [international exchange with cost-and labor product/marketing management: segmentation]	Trade (N=106)	0.9%	9.4%	44.3%	38.7%	6.6%	10.3%	83.0%	3.3 (n=99)
		Marketing Manager (N=102)	1.0%	5.9%	33.3%	56.9%	2.9%	6.9%	90.2%	3.5 (n=99)
QE2	The game clearly explains [the Heckscher-Ohlin theory/the idea of segmentation and marketing mix]	Trade (N=106)	7.5%	14.2%	38.7%	24.5%	15.1%	21.7%*	63.2%*	2.9 (n=90)
		Trade (without language barrier, n=65)	7.7%	9.2%	33.8%	29.2%	20.0%	16.9%	63.0%	3.1 (n=52)
		Marketing Manager (N=102)	12.7%	42.2%	38.2%	6.9%	0.0%	54.9%*	45.1%*	3.3 (n=95)

Note: students fills two different questionnaires about two games, but to make the presentation of the findings more clear, they are presented in one table. The average is only calculated for participants who ranked the game, and excluded the 'It's difficult to say'. Mark * means that there is a significant difference ($p<0.05$) between the percentage of Trade Ruler users and the percentage of Marketing Manger users in assessing the game positively or negatively

TABLE IV.
NAVIGATION AND EASE OF PLAYING

Code			Strongly disagree			Strongly agree	It's difficult to say	Negative opinion (1-2)	Positive opinion (3-4)	Average (1-4)
			1	2	3	4	-			
QN1	I find this game simple to play	Trade (N=106)	0.9%	3.8%	22.6%	72.6%	0.0%	4.7%	95.2%	3.7 (n=106)
		Marketing Manager (N=102)	1.0%	1.0%	24.5%	71.6%	2.0%	2.0%	96.1%	3.7 (n=100)
QN2	I easily find information which I need	Trade (N=106)	1,9%	15.1%	46.2%	34.0%	2.8%	17.0%	80.2%	3.2 (n=103)
		Marketing Manager (n=102)	2.9%	12.7%	41.2%	41.2%	2.0%	15.6%	82.4%	3.2 (n=100)
QN3	Without any problems and I easily established my score in this game	Trade (N=106)	0.9%	9.4%	26.4%	63.2%	0.0%	10.3%*	89.6%	3.5 (n=106)
		Trade (without language barrier, n=65)	0.0%	9.2%	23.1%	67.7%	0.0%	9.2%	90.8%	
		Marketing Manager (n=102)	1.0%	2.0%	32.4%	57.8%	6.9%	3.0%*	90.2%	3.6 (n=95)

1. Positive perception of the educational content is interlinked with the belief that introducing the games into the university curriculum increases the effectiveness and the engagement of the students.
2. Ease of playing the game is correlated with positive emotions during learning process.
3. The interactivity of the game is linked with the perceived effectiveness of game-based learning, the student's engagement in learning, and positive emotions.
4. Graphic design is correlated with positive emotions during the learning process.
5. Player satisfaction is correlated with interactivity, the design, the perceived total quality of the game, the educational content of the game and the ease of navigation.
6. Player satisfaction is correlated with the perceived effectiveness of learning, the engagement in learning, and positive emotions connected with the learning process.

Therefore the general hypotheses were confirmed. Although there are variables within main constructs which were not interlinked as we presumed.

TABLE V.
DESIGN

Code	I asses …		Very low			Very high	Negative opinion	Positive opinion	Average (1-4)
			1	2	3	4			
QD1	Interactivity of the game	Trade (N=106)	2.8%	17.0%	59.4%	20.8%	19.8%	80.2%	3.0
		Marketing Manager (N=102)	2.0%	9.8%	55.9%	32.4%	11.8%	88.3%	3.2
QD2	Graphic design	Trade (N=106)	13.2%	31.1%	30.2%	25.5%	44.3%	55.7%	2.7
		Marketing Manager (n=102)	4.9%	31.4%	40.2%	23.5%	36.3%	63.7%	2.8
QD3	General quality of game design	Trade (N=106)	6.6%	27.4%	44.3%	21.7%	34.0%*	66.0%*	2.8
		Trade (without language barrier, n=65)	7.7%	29.2%	44.8%	18.5%	36.9%	66.3%	2,7 (n=65)
		Marketing Manager (n=102)	1.0%	14.7%	52.9%	31.4%	15.7%*	84.3%*	3.1

TABLE VI.
SATISFACTION FROM PLAYING THE EDUCATIONAL GAME

Code	I asses …		Very low			Very high	Negative opinion	Positive opinion	Average (1-4)
			1	2	3	4			
QS	Satisfaction from playing	Trade (N=106)	3.8%	17.9%	50.9%	27.4%	21.7%*	78.3%*	3.0
		Trade (without language barrier, n=65)	6.2%	13.8%	55.4%	24.6%	20.0%	80.0%	3.0
		Marketing Manager (n=102)	1.0%	7.8%	55.9%	35.3%	8.8%*	91.2%*	3.3

TABLE VII.
ENHANCED LEARNING PROCESS [TRADE/MARKETING MANAGER]

Code			Strongly disagree			Strongly agree	It's difficult to say	Negative opinion (1-2)	Positive opinion (3-4)	Average (1-4)
			1	2	3	4	-			
QL1	The game builds my interest in [economics/marketing management]	Trade (N=106)	2.8%	10.4%	33.0%	48.1%	5.7%	13.2%	81.1%	3.3 (n=100)
		Marketing Manager (N=102)	0.0%	6.9%	30.4%	54.9%	7.8%	6.9%	85.3%	3.5 (n=94)
QL2	Implementing the games similar to [Trade/Marketing Manager game] to the university curriculum would enhance the effectiveness of the learning process	Trade (N=106)	2.8%	5.7%	22.6%	65.1%	3.8%	8.5%	87.7%	3.6 (n=102)
		Marketing Manager (N=102)	1.0%	8.8%	16.7%	69.6%	3.9%	9.8%	86.3%	3.6 (n=98)
QL3	Implementing the games to the university curriculum would enhance my engagement in learning process	Trade (N=106)	2.8%	10.4%	28.3%	53.8%	4.7%	13.2%	82.1%	3.4 (n=101)
		Marketing Manager (N=102)	2.0%	7.8%	24.5%	59.8%	5.9%	9.8%	84.3%	3.5 (n=96)
QL4	Implementing the games to the university curriculum would link learning to positive emotions	Trade (N=106)	1.9%	5.7%	19.8%	67.0%	5.7%	7.6%	86.8%	3.6 (n=100)
		Marketing Manager (N=102)	1.0%	5.9%	31.4%	57.8%	3.9%	6.9%	89.2%	3.5 (n=98)

TABLE VIII.
THE CORRELATION BETWEEN VARIABLES – R SPEARMAN, P<0.05. CORRELATIONS OF MODERATE STRENGTH ARE HIGHLIGHTED.

Trade		Marketing Manager	
Correlation between educational content and enhanced learning			
QE1-QL2	**0.33**	**QE1-QL2**	**0.36**
QE1-QL3	0.30	**QE1-QL3**	**0.34**
		QE1-QL4	**0.35**
		QE2-QL1	0.28
Correlation between navigation and enhanced learning			
QN3-QL1	0.21	QN1-QL1	0.25
QN2-QL2	0.26	QN1-QL2	0.24
QD2-QL2	0.21	QN1-QL3	0.27
QD3-QL2	0.29	**QN1-QL4**	**0.33**
		QN2-QL4	0.25
		QN3-QL4	0.26
Correlation between design and enhanced learning			
QD2-QL3	0.28	QD2-QL1	0.29
QD3-QL3	0.24	QD3-QL1	0.27
QD2-QL4	**0.34**	QD1-QL2	0.30
QD3-QL4	0.28	**QD1-QL3**	**0.41**
QD1-QL1	**0.33**	QD1-QL3	0.25
		QD1-QL4	**0.43**
		QD2-QL4	0.29
		QD3-QL4	0.23
Correlation between educational content and player satisfaction			
QE1-QS	0.25	**QE1-QS**	**0.36**
		QE2-QS	**0.38**
Correlation between satisfaction and design			
QS-QD1	**0.44**	**QS-QD1**	**0.48**
QS-QD2	**0.38**	**QS-QD2**	**0.40**
QS-QD3	**0.37**	**QS-QD3**	**0.44**
Correlation between satisfaction and navigation			
		QS-QN1	0.31
		QS-QN2	0.27
Correlation between player's satisfaction and enhanced learning			
QS-QL2	0.25	QS-QL1	0.24
QS-QL1	0.31	**QS-QL2**	**0.35**
		QS-QL3	**0.38**
		QS-QL4	**0.40**

Note: while both variables have the option 'It's difficult to say' the correlation was calculated using the full scale. If only one variable had the option 'It's difficult to say' the correlation was calculated with the exclusion of answers "It's difficult to say"

IV. CONCLUSIONS

The main conclusions of the study:

- Introducing 'one-hour-play' simulation games into the university curriculum would be connected with positive educational outcomes such as increased student engagement in learning, positive emotions, and the perceived effectiveness of learning. This means that 'one-hour play' games (the games tested were assessed to be simple to play by over 95% of students) - which are more affordable and easier to prepare than very complex multi-modul simulations games - can be useful tools for introducing managerial or economic issues to students.
- Unfortunately – as our survey also showed – very few students have had previous experience in using the games as the part of their courses and games are a neglected educational tool in the Polish universities. Perhaps this is one of the reason why participants of our survey evaluated the games so favorably. Students were very enthusiastic about the games and they assessed them

very highly in different areas, especially their the educational content, simplicity and interactivity.

- Our survey shows there is a positive link between the player satisfaction and perceived enhancement of the learning process. It suggests that when there are two or more different games for evaluation it may be worth focusing more on measuring general satisfaction than on a particular feature of the game.
- The finding emphasis the importance of game design. Interactivity, graphic design and general quality of the game design can influence the player satisfaction.

There are a few limitations for study presented. First of all, the size and the process of choosing the participant sample limit the way in which results of the study can be interpreted and generalized. Also the study did not include some variables which could influence the results including a student's level of achievement at university and learning style. Future studies should extend the scope of analysis.

REFERENCES

[1] Derryberry A. (2007), Serious games: online games for learning, Adobe White Paper [online], http://www.adobe.com/resources/elearning/pdfs/serious_games_wp.pdf [01.04.2013]

[2] Hwang G.-J. & Wu P.-H., Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010, British Journal Of Educational Technology, January 2012;43(1):E6-E10

[3] Lin Y.-L. & Tu Y.-Z. (2012). The values of college students in business simulation game: A means-end chain approach, Computers & Education, 58(4), 1160-1170

[4] Pasin F. & Giroux H. (2011). The impact of a simulation game on operations management education, Computers & Education, August 2011;57(1):1240-1254.

[5] Simões J., Redondo R. & Vilas A. (2013). A social gamification framework for a K-6 learning platform, Computers In Human Behavior, 29(2), 345-353

[6] Tao Y.-H., Cheng C.-J. & Sun S.-Y. (2009), What influences college students to continue using business simulation games? The Taiwan experience, Computers & Education, November 2009;53(3):929-939

[7] Tao Y.-H., Yeh C. R. & Hung K. C. (2012) Effects of the heterogeneity of game complexity and user population in learning performance of business simulation games, Computers & Education, December 2012;59(4):1350-1360.

[8] Wu W.-H., Chiou W.-B., Kao H.-Y., Hu C.-H. A. & Huang S.-H. (2012), Re-exploring game-assisted learning research: The perspective of learning theoretical bases, Computers & Education, December 2012;59(4):1153-1161.

[9] Rouse III R. (2005), Game Design: Theory & Practice, Second Edition, Wordware Publishing, Inc., 2005, xx-xxi

[10] Voorveld H, Neijens P & Smit E. (2011) The Relation Between Actual And Perceived Interactivity, Journal of Advertising,. Summer2011 2011;40(2):77-92

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1301–1307

Social Network Framework for Deaf and Blind People based on Cloud Computing

Mahmoud El-Gayyar1[1], Hany F. ElYamany[1], Tarek Gaber[1*] and Aboul Ella Hassanien[2*]
[1]Computer Science Department, Faculty of Computes and Informatics
Suez Canal University, Ismailia, Egypt
{elgayyar, hany_elyamany, tarekgaber}@ci.suez.edu.eg
[2]Cairo University, Faculty of Computers and Information, Cairo, Egypt
[*]Scientific Research Group in Egypt (SRGE) www.egyptscience.net
aboitcairo@gmail.com

***Abstract*—Most of the governments and civil society organizations work hardly to promote the disabled people especially blind and deaf persons to join the normal community and practice the regular daily life activities. Indeed, Information Technology with its modern methodologies such as mobile and Cloud computing has an impressive role in enhancing the intercommunication among the people with different disabilities and normal pupils from one side and among the disabled people themselves who have the same or different impairments. However, a few numbers of suggested systems are quite limited for the Arabic Region. Additionally, according to our knowledge, there is no proposed system for connecting the blind and deaf people within direct Arabic language-based conversations. In this paper, we propose a comprehensive framework constructed upon three main modern technologies: mobile devices, Cloud resources and social networks to provide a seamless communication between the blind and deaf people especially for those living in the Arabic countries. Moreover, it is designed to facilitate the communication with normal people through various directions by using recent methodologies such as time-of-flight camera and social networks. The main modules and components of the suggested framework and its possible scenarios are fully analyzed and described.**

***Keywords*—Cloud, Mobile Devices, Social Networks and Blind/Dead people**

I. Introduction

LACK of awareness of the blind/deaf needs could lead to discrimination in a single community. It could also leave many talented blind/deaf people out of productive members of society. In addition, it could have a substantial effect on the educational performance of children. Furthermore, children with hearing loss and deafness, specifically in developing countries, hardly receive any education. Moreover, adults with hearing loss suffer from a much higher unemployment rate [1].

As recently reported by the World Health Organization, 285 million people around the world are visually impaired (39 million are blind and 246 have low vision) [2]. About 80% of all visual impairment around the world could be cured or avoided [2]. However, this is unachievable due to fact that 90% of the 285 million visually impaired people live in developing countries. It is also reported that, 360 million of the worldwide population have disabling hearing loss [1]. In addition, "current production of hearing aids meets less than 10% of global need" [1].

In Egypt, as an example of the developing countries, about 6% of the total Egyptian population is visually impaired [3] and nearly three million people in Egypt suffer from hearing impairment. Nonetheless, there is no a national project on how the government helps those people to be active members in the community. For example, according to the report in [4], the Egyptian government has provided sign language services for News Programs on Television (10 minutes to 7 hours a week). However, there are no current programs or plans to provide subtitles or captions. In addition, the government does not provide any access for Deaf people to get governmental documents in their sign language [4]. Indeed, the blind/deaf communities are still out of the government interests and future plans.

As a matter of fact, social interaction is a life-process and a crucial part of our success in life. It supports independent living, community experiences and relationships. However, social skills are learned by repeated visual observations (e.g. facial expressions, body language) that are translated to cues that help us to develop and understand concepts of social behavior. In other words, vision plays an important role in establishing and maintaining social interactions that is a great challenge for individuals who are visually impaired or blind.

To enrich the social interaction process, recent technologies such as social networking websites and mobile devices have been introduced and used in a wide scale as new means for social communication between people. These services are widely popular in Egypt – a February 2013 survey found that around 13 million out of 32.5 million online Egyptians use Facebook [5]. In addition, in January 2013, MCIT of Egypt [5] reported that the mobile subscribers are 96.11 million and 10.08 million of these subscribers are accessing the Internet through their mobile phones. Meanwhile, there are no real interests or efforts for helping disabled people to get benefit of these technologies to develop their social skills. The main goal of our work in this paper is to introduce a new framework that may help to shorten the distance between the blind/deaf people and the normal life. In other words, the proposed solution should support a seamless communication between blind, deaf and normal people from one side, and facilitate the interconnection between the blind/deaf and daily-based activities from the other side.

In this research paper, we propose an integrated framework to provide a Mobile-Cloud system to help blind and deaf people to gain better social skills for a more successful life. The main goals of the system are:

- Seamless communication between blind and deaf people. This includes ways to (a) detect the face to tell the blind about the identity of the conversation partner, (b) detect

978-83-60810-53-8/$25.00

emotion to tell the blind about his feelings, and (c) detect sign language and convert it to Arabic speech if it exists.

- Social network that can help blind and normal people to communicate with each other. Such network could help blind people to (a) detect certain objects or identify Arabic text, (b) use a social website to communicate with their friends while protecting their privacy, and (c) find and contact the nearest friend in case of critical situations.

The rest of this paper is organized as follows. The related work is discussed in Section 2. Section 3 presents the proposed framework and its involved modules. Section 4 describes the seamless communication between blind and deaf persons. Section 5 demonstrates the roles of social networks for connecting a disabled person to reality. Section 6 shows the possible challenges that may encounter the framework implementation. Finally, Section 6 concludes the presented solution and highlights the future work.

II. RELATED WORK

In this section, we will introduce all related research according to three main bases or technologies that may assist the blinds to have a normal life with regular activities which are: social network, mobile device and Cloud infrastructure. The social network is proposed as an electronic hand that would guide them to the right position or decision. The mobile device is the portable eye that might describe and recognize the objects or people who are naturally invisible for them. Finally, the Cloud infrastructure is the co-brain for processing the captured photos and associated data.

There are several innovative trials to build a particular social network for disabled people in general and for the blind community in particular such as [www.disabilinet.com, www.disabledcommunity.net, audioboo.fm/about/social_blind, and theblinduniverse.com] in which some special features have been added to facilitate the interaction and communication among the involved subscribers including the audio and media contents. These networks may give the disabled and blind people what they wish to connect them with their peers or friends; However, they may lead to (1) isolating the blind people from the normal world, (2) increasing the feeling by their hindrance, consequently fallen in more sociological and depression problems. On the other side, to the best of our knowledge, the popular social networks such as Facebook and Twitter do not provide any particular features for the blind/deaf people for attracting them to their community or helping them to easily participate in regular life with its different activities (e.g. searching jobs) as well as bringing them in a direct relationship with several parties including near friends and families members.

Some previous solutions have been accomplished for helping blinds to recognize the surroundings and figure out the objects they interact with such as the VizWiz [6] and VizWiz Social systems [7]. The VizWiz backbone is mainly based on two important concepts: crowdsourcing [8] and Mobile-Q&A [9] [10]. In the VizWiz system, the blind recruit a number of employees (i.e. crowdsourcing) to answer the audio questions related to a photo taken by their mobiles (i.e. Mobile Q&A). The suggested VizWiz system gives them the full confidence that they are basically depending on their selves rather than asking for help from friends or family members through the traditional social networks (i.e. friendsourcing) [8]. However, it is not a free service and does not support any automated process for recognizing the captured photos that clarity degree basically depends on the mobile model that the blind regularly use. At the same time, it is difficult for blinds to correctly target their mobiles' camera in order to capture clear and easily recognizable pictures.

The VizWiz Social is a free application installed on iPhone and has three separated distinguished features: (1) accessing the original VizWiz with no cost, (2) automated object or photo recognition, and (3) social-based application. The study run over this application in [7] demonstrated that most of the involved blinds, who have involved in a survey sample, preferred to access the original paid VizWiz instead of the new two features added to VizWiz social. This is due to the some technical challenges in the automated object recognition such as the picture obscurity and the inaccurate results received. Regarding the social-based feature, there is a delay (approaches to an average of a full day) in the answers the blind users obtain due to the unavailability of the trusted friends or family members. Additionally, the blinds have anxious feeling about their privacy and independence when sharing their data and photos online.

The work in [11] [12] presented a hybrid system of special camera glasses connected to a mobile device through Bluetooth technology in order to transfer the captured images to a Cloud platform to speed up the matching and recognition processing time and obtaining accurate results in which the blind society can depend. Also, the work in [13] introduced the possibility of combining social networks with Cloud environment through analyzing the taken photos from blinds phones by either Clouds, or friends subscribed in social networks who are sharing the same geographical locations in order to obtain quick and accurate answers.

Neither of the previous work is considered as a complete solution for the blind community. VizWiz [6] [7] does not support full automated answers and imposes a charge of the Q&A service. The VizWiz Social [7] [14] fails to address the accuracy and the users' privacy while the remaining stated work [11] [12] [13] are only focused on the navigation problem. Moreover, neither of the described work discusses how the taken photos and sent data are proceed and stored. Furthermore, none of the explained work in this section is suitable for the Arabian region as English is the main language within all those applications/solutions. Additionally, the blinds are the main focus of the study of the listed work and nothing is there about deaf people and their own needs and language. Last but not the least, the interactions emotions and feelings of both the blinds and their peers (e.g. friends) have been neglected and should be studied properly due its gorgeous impact in improving the interconnectivity among them. By including these two features, the blinds would be able to feel like normal ones.

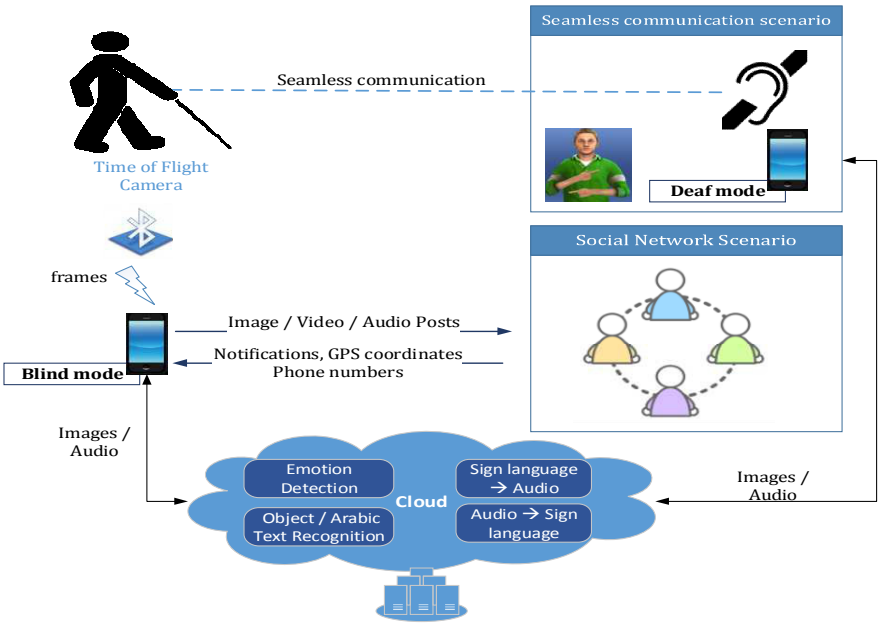


Figure 1. SoNetDBlue General Framework

III. THE PROPOSED FRAMEWORKE

The blind and visually impaired people, especially in Arabic countries, have limited opportunities for social interaction compared to those without visual impairments. At the same time, there is no much attention by assistive technology researchers for this topic. In this paper, we present SoNetDBlue (**So**cial **Net**work for **D**eaf and **Bl**ind **us**ers) framework that tries to help those people to improve their social life.

SoNetDBlue proposes a mobile-Cloud framework shown in Figure 1. In which a speech interface on the mobile is used to take commands in Arabic language from the user while visual data is captured by camera modules integrated into sunglasses and fed to the mobile device through the Bluetooth technology. Bluetooth has been chosen here due to its proven effectiveness and popularity. Machine instances in the Cloud are utilized for running complex and time-consuming tasks that cannot be achieved on the mobile device with its limited CPU power and memory capacity.

The proposed framework is based on a collaboration model between everyday mobile devices, and the computational power provided by the Cloud infrastructure. All required complex algorithms will run in the Cloud while the thin client on the mobile device is used to capture images or audio signals and send them to the Cloud for further processing. The proposed framework is based on three major subsystems:

- ***Integrated Camera***: We consider here the time-of-flight camera, a new technology used in real-time three-dimensional imaging. This technology has produced promising results in many fields including face recognition [15], gesture recognition [16], and real time motion capture [17]. These cameras provide real-time depth information about pixels of a captured image and the camera modules are made available by manufacturers at decreasing prices with the advances in the underlying technology. Currently available time-of-flight cameras provide ranges of about 10 meters and high frame rates of about 100 frames/second making them even more attractive for dealing with dynamic environments with fast moving objects.

  As opposed to using the camera of the mobile device, the time-of-flight camera module will be integrated into glasses to be worn by blind users (an eye-level placement) that will help them to easily capture context-relevant pictures

- ***Mobile Application***: We are planning to support iOS and Android mobile platforms due to their great popularity, support for multi-tasking and accessibility features. Android and iOS based devices come with integrated speech recognition and text-to-speech engines that will facilitate the design of an easy-to-use interface for disabled people. The application supports two basic functionality modes: blind and deaf. More details about these modes will be provided in the following sections.

- ***Cloud Infrastructure***: The suggested framework is built upon the computational power of Cloud to overcome the shortage of available resources on mobile devices. The Cloud is settled to accomplish computationally-intensive algorithms including emotion detection, object and Arabic text recognition, and the conversion from Arabic language to sign language and vice versa.

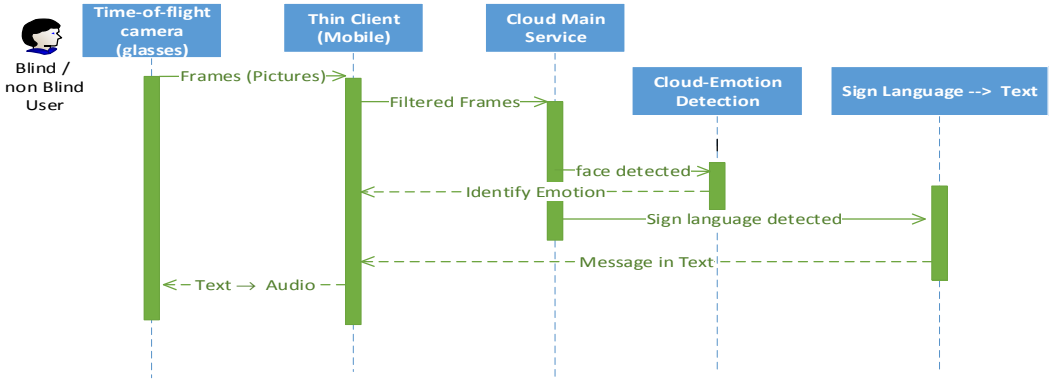


Figure 2. Sequence Diagram of Blind Mode – Seamless Communication Scenario

Our initial prototype is deployed on a private Cloud built using the OpenStack [1] Cloud operating system upon the infrastructure of the Suez Canal University, Egypt. OpenStack is open source software designed to provision and manage large networks of virtual machines, creating a redundant and scalable Cloud computing platform. It gives you the software, control panels, and APIs required orchestrating a Cloud, including running instances, managing networks, and controlling access through users and projects. OpenStack APIs are compatible with Amazon EC2 and Amazon S3 [2] and thus client applications written for OpenStack can be used with Amazon Web Services with minimal porting effort. [12]

This degree of compatibility will help us to port our final work to a commercial Cloud (e.g. Amazon EC2) to determine the overall cost expected by our system and to compare the system efficiency and reliability over different Cloud platforms.

SoNetDBlue focuses mainly on two basic scenarios: seamless communication with deaf people and social network for blind users. In the following sections, we are going to deliberate the two scenarios in detail and discuss the major challenges expected during the implementation of the system.

IV. SEAMLESS COMMUNICATION

In this section, we will concentrate on the first objective of the SoNetDBlue framework that is to provide an independent communication between blind and deaf people as depicted in the top portion of Figure 1. To achieve this objective, the thin client on the mobile phone should be configured to one of two available modes: blind and deaf modes. The thin client behavior is automatically adapted according to the selected mode. In the rest of this section, the functionalities of each mode are explained in more details.

The sequence diagram shown in Figure 2 concludes the flow of actions in case of the blind mode:

1) The time-of-flight camera integrated in the sunglass records a video for the conversation's partner (deaf user) and sends it to the thin client resident on the mobile device through a Bluetooth connection.
2) Depending on the CPU power of the mobile device, the thin client may decide to apply a preprocessing stage to filter the captured frames and to extract the key ones while the other frames will be regenerated on the Cloud. This step will reduce the amount of data transferred on the network and thus reduce the overall cost of the service.
3) The thin client submits the filtered frames to a machine instance on the Cloud responsible for running, controlling and synchronizing the required processing.
4) On the Cloud, a scientific workflow management system (e.g. SWIMS [18]) is used to process received frames in parallel over available computational nodes. Various algorithms, explained below, are applied to notify the blind user about the current feelings of the conversation's partner and a translation of his sign language into Arabic text that is sent back to the thin client.
5) The received text is converted into speech using the text-to-speech engine assimilated in the mobile device.

In our work, we focus on facial expression recognition as a key index of human emotion based on the six basic emotion-specified facial expression (i.e. happiness, sadness, fear, disgust, surprise and anger) defined by Ekman [19]. Based on our architecture requirements (i.e. accuracy and response time), we will consider and evaluate the work by Mase and Pentland on advanced face detection and recognition using relatively low computational power [20] and the algorithm that classifies face emotions through eye and lip features using particle swarm optimization [21].

[1] http://www.openstack.org/
[2] http://aws.amazon.com/ec2

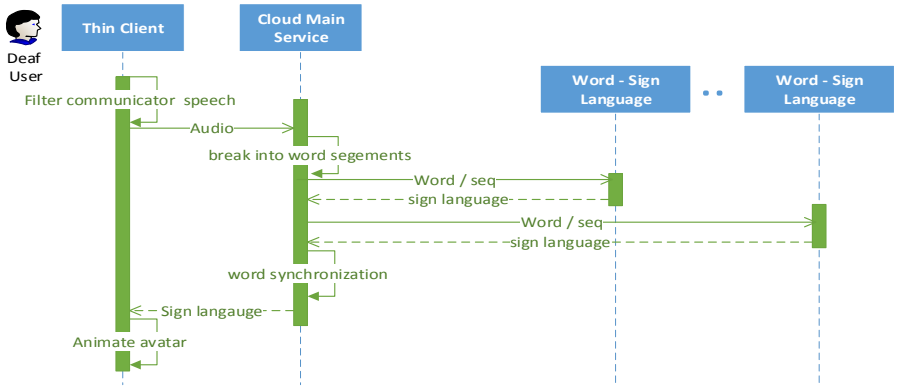


Figure 3. Sequence Diagram of Deaf Mode – Seamless Communication Scenario.

For the Arabic Sign Language (ArSL) recognition, we concentrate on testing two algorithms that achieve real time translation of dynamic gestures. The first algorithm involves two stages for automatic translation of dynamic gestures into the ArSL; in the first stage, it recognizes the group of the gesturer and the second stage interprets the gestures within the groups based on spatial domain analysis and hidden markov model [22]. The second interesting algorithm is a vision-based automatic sign language recognition system for Arabic letters with no need for any additional hardware such as gloves or sensors. The algorithm uses predefined Haar classifiers to track and detect the hand's position, then it detects the skin color, transforms the images into frequency domain, and finally uses a simple classification technique (K nearest neighbor); this algorithm achieved up to 90.55 % recognition accuracy at real time [23].

The deaf users should configure their mobile devices to the deaf mode in which the thin client records the Arabic speech and filters it to extract the signal of the conversation's partner (blind), and submits it to the Cloud for processing as illustrated in Figure 3:

1) The speech is translated into Arabic text and segmented into separated words.
2) Each word is translated into its equivalent sign language in parallel on different available computational nodes to speed up the total response time. The outcome of this step is an animation sequence that represents the sign language of the word under processing.
3) Accumulated results of step 2 are synchronized and sent back to be displayed on the thin client through a digital avatar as shown in Figure 1.

Several research projects have made efforts in translating English text into sign language [24]. However, research work focusing on Arabic language is quite limited. An early work in the field of ArSL translation shows a poor consideration of the deaf community in the Arabic world, for example, the system build by Mohandes wrongly assumes that ArSL depends on the Arabic language and shares the same structure and grammar [25]. A more powerful work has been presented in [26]; in this work the authors considered the ArSL's unique linguistic characteristics (e.g. its own grammar, structure, and idioms) and provided a full working prototype, ArSL translation system, for helping Arabic deaf community to access published Arabic text. Another interesting work that we have to consider is the translation tool introduced in [27] as a part of a full chat system for deaf people; it provides two different modes for translating from ArSL to Arabic language and vice versa; it is based on a word to word translation, and if a word does not exist in the system's database, a letter by letter translation is encountered.

An important question that may jump to the mind is how can blind and deaf people initiate communication while the blind cannot see the deaf and the latter cannot hear the blind one? For sure the one with vision capability (the deaf) should start his program that broadcasts an audio message. This message will notify the blind to initialize his application and to target his eyes to the deaf person to be able to capture his signs.

V. SOCIAL NETWORK

The second major objective of the SoNetDBlue architecture is to connect deaf and blind people with their cycle of friends (with or without disabilities) anywhere and at any time. This will help them to integrate in the community and to gain better social skills for a more successful life. To hit this goal, SoNetDBlue framework embraces a social site connecting blind and deaf users with their cycle of friends including people with or without disabilities. Figure 4 presents a use case to show the various activities that can be accomplished in this scenario:

- **Post text / audio messages:** users can post messages either in text or audio. Text messages can be entered in ArSL using a special keyboard as the one designed for the chat system presented in [27]. The received message can be also translated into one of the available format (i.e. audio, text or sign language). The camera module integrated in the blind users' sun glasses or the mobile's camera can be exploited to capture an image or a video to be attached with the posted message.

- **Find/call nearest friends:** SoNetDBlue users can utilize the thin client installed on their mobile devices to locate their nearest friends, using the GPS module installed on the mobile device, and to contact them either through calls or SMS messages. This can be very helpful for disabled people, especially in case of emergency situations.
- **Identify objects / Arabic text:** The goal here is to help blind users to identify the class of objects (e.g. car, building) in front of them. In addition, the algorithm should search for a textual note in the image to give a more descriptive explanation about the detected object (e.g. a building of Suez Canal University). According to the evaluation in [28], the Lehigh Omnidirectional Tracking System [29] is one of the best algorithms for object detection in video streams. For detecting textual notes, we have to evaluate existing Arabic optical character recognition and natural language processing algorithms to select the best of them while considering that the textual notes may contain bilingual sentences (e.g. Arabic / English) and different number system (i.e. Arabic and Indian numbers).

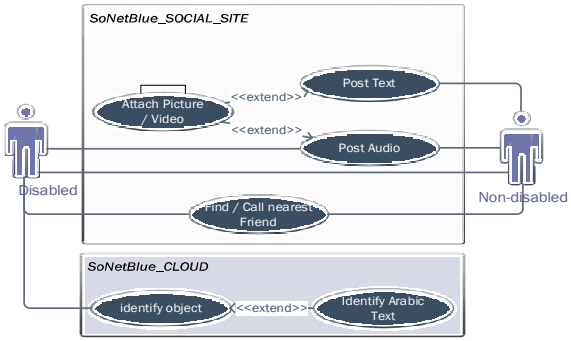


Figure 4. Use Case of the Social Network Scenario

VI. ARCHITECTURE CHALLENGES

The major challenges of the SoNetDBlue framework are fourfold: data stored on the Cloud, user's security (privacy and authentication) , Arabic language recognition, and evaluation of the proposed solution. Data stored on SoNetDBlue Cloud infrastructure is enriched with semantic annotations to determine (a) the owner of the data (user) and (b) the context in which the data has been collected [30]. This may lead to data duplications to some extent, but it will have a great impact on the response time of available Cloud services as the search scope will be limited to a certain user at a certain context. At the same time, the Cloud infrastructure will utilize advanced frameworks to manage the replication of its services [31] and expected privacy issues [32]. SoNetDBlue could be used to help a blind user to get a specific location using Global Positioning System. However, this could put the user in risk while submitting his location information to the Cloud. This information could be used by a malicious party to locate the blind user and then exploit this user for his/her own benefit. Thus, unlink-ability techniques of the user identity should be used to overcome this problem. Another aspect of privacy concern is as follows. SoNetDBlue, using Time-of-Flight camera in the blind mode, requires continuous capturing of the surroundings around the user to be sent to a dedicated website. It is crucial that the user anonymity is preserved, i.e. providing unlink-ability to the recordings. This is because these videos would disclose lots of information about the places visited by the blind user.

Another security issue is the mobile authentication. A survey in [33] reported that7jhli 40% of mobile users used to enter a password on a daily basis. In addition, 56% of these users mistype a password at least once out of ten. Users find that entering password on Mobile Internet Devices (MIDs) is more frustrating than lack of coverage, small screen size, or poor voice quality. Surely, for blind users, these limitations of using traditional password as authentication method are more frustrating. Therefore, especially for blind users, MIDs require another way for authentication such that the users' involvements are very limited or without any involvement. Furthermore, as reported in [34] after interviewing 13 blind users of smartphones, it has found that most them were not familiar with potential security threats of not using authentication methods such as a password-protected screen lock. Implicit authentication [33] [35] could be used to address these limitations. It could be used as a secondary factor for authentication to augment passwords, thus achieving a higher-assurance authentication. For blind users, this implicit authentication is very promising as they are not required to memorize how to enter a password.

The challenging bit of the ArSL is mainly concerning the accuracy of the obtained results. Comparing to the American Sign Language, the ArSL still requires more work to reach the same level of accuracy. As reported in [36], American Sign Language has achieved a word accuracy of 99.2% (users have to wear a special glove) whereas in [37] American Sign Language has accomplished an accuracy of 98%. On the other hand, the best result of the ArSL is 90.55 %, as reported in [9]. The reasons of this problem are summarized as follows [12]. ArSL, like other natural language, is an independent language which has its own structure, grammar, and idioms. In addition, it is not hand/finger spelling of the Arabic alphabet. The finger spelling is only used for places or names which that do not exist in ArSL. Furthermore, ArSL does not have a documentation system which could be utilized while building a translation corpus. Therefore, deep research should be done to address this difficulties and to fill the gap of the accuracy between the ArSL and American Sign Language. It is believed that the high accuracy of ArSL, the high adoption of the proposed SoNetDBlue solution.

VII. CONCLUSION AND FUTURE WORK

In this work, a Mobile-Cloud framework is presented to help blind and people with visual impairments to gain better social skills that are important for healthy and successful life. Moreover, the introduced framework is structured to bridge the communication gap between the blind and deaf persons, particularly in the Arabic section. Finally, we showed how several technologies and methodologies including social networks can be integrated to recognize all possible obstacles and persons to assist the blind people to feeling and visualizing the surroundings.

Our future work involves two main steps: the first one, as mentioned above, we have started to implement the suggested architecture in an OpenStack Cloud environment, we aim to complete the entire architecture implementation and its interrelated components and testing its behavior in order to meet the analyzed requirements and estimated objectives. The second step includes moving the implemented system to a commercial Cloud and practicing the expected behavior with real pupils. Also, this stage will be concluded by a real-time survey to measure the blind/deaf people satisfaction about the produced system including the performance and accuracy.

References

[1] "World Health Organization: Deafness and hearing loss.," [Online]: http://www.who.int/mediacentre/factsheets/fs300/en/index.html.

[2] "World Health Organization: Visual impairment and blindness.," [Online]: http://www.who.int/mediacentre/factsheets/fs282/en/.

[3] "Ministry of Information and Communication Technology (MICT).," Egypt's ICT Golden Book, December 2006. [Online]: http://mcit.gov.eg/Upcont/Documents/Golden%20Book%20Final20072 11155321.pdf.

[4] *Global Survey Report WFD Interim Regional Secretariat for the Arab Region (WFD RSAR) Global Education Pre-Planning Project on the Human Rights of Deaf People,* World Federation of The Deaf, 2008.

[5] "Ministry of Information and Communication Technology (MICT)," ICT Indicators in Brief, February 2013. [Online]: http://www.egypticindicators.gov.eg/en/Publications/PublicationsDoc/ English%20Flyer%20new%20Feb%202013.pdf.

[6] J. Bigham and et al., "Vizwiz: Nearly real-time answers to visual questions.," in *In Proceedings of UIST*, 2010.

[7] E. Brady and et al., "Visual Challenges in the Everyday Lives of Blind People," in *In Proceedings of CHI 2013, ACM.*, 2013.

[8] M. Bernstein and et al., "Personalization vs. Friendsourcing," *ACM Transactions on Computer-Human Interaction,* vol. 17, no. 2, 2010.

[9] U. Lee, E. Yi and M. Ko, "Mobile Q&A: Beyond Text-only Q&A and Privacy Concerns.," in *In Proceedings of CHI 2013, ACM*, 2013.

[10] J. Yang and et al., "Culture Matters: A Survey Study of Social Q&A Behavior.," in *In Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)* , 2011.

[11] P. Angin, B. Bhargava and S. Helal, "A Mobile-Cloud Collaborative Traffic Lights Detector for Blind Navigation," in *In Proceedings of 11th IEEE International Conference on Mobile Data Management*, 2010.

[12] P. Angin, "Real-time Mobile-Cloud Computing for Context- Aware Blind Navigation," *INTERNATIONAL JOURNAL OF NEXT-GENERATION COMPUTING,* vol. 2, no. 2, 2011.

[13] A. Roy, H. Ren and S. Contractor, *Integration of Human Resources and Cloud System for Blind People,* technical report, Purdue University, 2010.

[14] E. Brady, "Unique Considerations in Social Network Question Asking by Blind Users," in *In Proceedings of CSCW, ACM*, 2013.

[15] S. Meers and K. Ward, "Face Recognition Using a Time-of-Flight Camera," *Sixth International Conference on Computer Graphics, Imaging and Visualization,* pp. 377-382, 2009.

[16] E. Kollorz, J. Penne, J. Hornegger and A. Barke, "Gesture recognition with a Time of flight camera," *Int. J. Intell. Syst. Technol. Appl.,* pp. 334-343, 2008.

[17] V. Ganapathi, C. Plagemann, D. Koller and S. Thrun, "Real time motion capture using a single time-of-flight camera," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[18] M. El-Gayyar, Y. Leng and A. Cremers, "Distributed Management of Scientific Workflows in SWIMS," in *Ninth International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES)*, 2010.

[19] P. Ekman, "An argument for basic emotions," *Cognition and Emotion,* vol. 6, pp. 169-200, 1992.

[20] K. Mase and A. Pentland, "Recognition of facial expression from optical flow," *IEICE Trans.,* vol. E, no. 74, pp. 3473-3483, 1991.

[21] A. Navin and M. Mirnia, "A New Algorithm to Classify Face Emotions through Eye and Lip Features by Using Particle Swarm Optimization," *IPCSIT,* vol. 22, pp. 268-274, 2012.

[22] M. Al.Rousan, O. Al-Jarrah and M. Al-Hammouri, "Recognition of dynamic gestures in arabic sign language using two stages hierarchical scheme," *International Journal of Knowledge-Based and Intelligent Engineering Systems,* vol. 14, no. 3, pp. 139-152, 2010.

[23] N. Albelwi and Y. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition," *ICCIT,* pp. 497-501, 2012.

[24] M. Huenerfauth, *Generating American Sign Language Classifier Predicates For English-To Asl Machine Translation,* Philadelphia, PA, USA: Ph.D dissertation, University of Pennsylvania, 2006.

[25] M. Mohandes, "Automatic translation of Arbaic text to Arabic Sign Language," *ICGST International Journal on Artifical Intelligence and Machine Learning,* vol. 6, no. 4, pp. 15-19, 2006.

[26] A. Almohimeed, M. Wald and R. Damper, "Arabic Text to Arabic Sign Language Translation System for the Deaf and Hearing-Impaired Community," in *The Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, Edinburgh, UK, Scotland,, 2011.

[27] H. Al-Dosri, N. Alawfi and Y. Alginahi, "Arabic Sign Language Easy Communicate," in *The 2nd International Conference on Communications and Information Technology* , 2012.

[28] J. Nascimento and J. Marques, "Performance evaluation of object detection algorithms for video surveillance," *Multimedia, IEEE Transactions on,* vol. 8, no. 4, pp. 761 - 774, 2006.

[29] T. Boult, R. Micheals, X. Gao and M. Eckmann, "'Into the woods: Visual surveillance of non-cooperative camouflaged targets in complex outdoor settings," *Proceedings of the IEEE,* pp. 1382-1402, 2001.

[30] M. Roshdy, K. Fadel and H. ElYamany, "Developing a RDB-RDF Management Framework for Interoperable Web Environments," *to appear in the proceeding of the 4th IEEE Eurocon Conference,* 2013.

[31] M. Mohamed, H. ElYamany and H. Nassar, "An Adaptive Service Replication Framework for Managing Different Responsiveness Levels," *International Journal of Intelligent Computing and Information Science,* vol. 13, no. 2, 2013.

[32] D. Allison, M. CApretz, H. El Yamany and S. Wang, "Privacy Protection Framework with Defined Policies for Service-Oriented Architecture," *Journal of Software Engineering and Applications,* vol. 2, 2012.

[33] M. Jakobsson, E. Shi, P. Golle and R. Chow, "Implicit authnetication for mobile devices," in *HotSec'09*, Berkely, CA, USA, 2009.

[34] S. Azenkot, K. Rector, R. Ladner and J. Wobbrock, "Passchords: secure multi-touch authentication for blind people.," in *14th international ACM SIGACCESS conference on Computers and accessibility, ASSETS '12,*, 2012.

[35] A. Luca, A. Hang, F. Brudy, C. Lindner and H. Hussmann, "Touch me once and i know it's you!: Implicit authentication based on touch screen patterns.," in *CHI 2012, ACM*, 2012.

[36] T. Starner, J. Weaver and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transations on Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, p. 1371–1375, 1998.

[37] T. Starner, J. Weaver and A. Pentland, "A wearable com-puter-based American sign language recogniser," *personal and ubiquitous computing,* 1997.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1309–1313

Tracking the node path in wireless ad-hoc network

Artur Sierszeń
Lodz University of Technology, Institute od Applied Computer Science, ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
Email: artur.sierszen@p.lodz.pl

Łukasz Sturgulewski
Lodz University of Technology, Institute od Applied Computer Science, ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
Email: lukasz.sturgulewski@p.lodz.pl

Agnieszka Kotowicz
Lodz University of Technology, International Faculty of Engineering, ul. Żwirki 36, 90-924 Łódź, Poland
Email: agnieszkakotowicz88@gmail.com

***Abstract*—This article provides an insight into the topic of ad-hoc protocols used for routing, namely proactive and reactive protocols. It depicts the general concept how these protocols can find a path in a network between two nodes and it also presents the evaluation of the methods of tracking the node path in a wireless ad-hoc network through investigating the available mobile routing protocols.**

The main focus is on the throughput and the average end-to-end delay in a network, using for the simulation OMNeT++ environment. Three protocols were chosen for the final testing: Ad-hoc On-demand Distance Vector (AODV), Optimized Link State Routing (OLSR), and Dynamic Source Routing (DSR).

I. Introduction

AD-HOC networks originated in 1960s when the ALOHA project was emerging from the shadows. Even though the dynamically established network was not the first outcome of this project (it was based on fixed nodes with the single-hop option only), the idea of a shared medium for client transmissions remained. The earliest wireless ad-hoc networks were the "packet radio" networks (PRNETs) already proposed in 1970. Since then, project and ad-hoc networks have been developed continously.

In general, an ad-hoc network is a collection of wireless mobile nodes (e.g. smart phones, laptops, cameras etc.) that is formed only for a short period of time when wireless devices come within each other's communication ranges. Nodes are the users or devices forming the network [1]. This set-up is created dynamically without using a preconfigured network infrastructure (a simple example of an ad-hoc network is shown in Fig. 1). If a network is set up for a longer period of time, it is just a plain old local area network (LAN). Finally, it is said that an ad-hoc network does not have any centralized architecture, what means that any node is a peer. In a peer network, each node is a client, a receiver (server), or a mediator of a packet that routes packets to other nodes that are out of range of the sender [2]. Moreover, an ad-hoc network can operate as a stand-alone, closed group as well as a network with a connection to the Internet.

This definition indicates that the mobility of the nodes leads to fast and sometimes enormous changes in the wireless network topology. In addition, other obvious attributes such as a large size of the network, bandwidth, large diversity of available devices, and their power consumption may cause large problems for today's routing protocols. They may all be a huge challenge if ad-hoc network users want to receive a reliable and high quality service, not to mention other problems that can be enumerated: physical obstacles, indirect communication between two nodes, imperfections of network elements causing delays, battery constraints etc.

The first idea is based on fast routing protocols. User mobility influences the changing topology of an ad-hoc network, so it is possible that some old nodes are no longer available but new have just appeared. In theory, a routing protocol could still handle this change somehow in order to connect the required nodes, but there are some protocols that cannot do that. Therefore, it is necessary to use the protocols that are dedicated for ad-hoc networks and, providing they are fast enough, they can solve the mobility problem. Routing in ad-hoc networks is a combination of dealing with topology adjustments and minimizing the routing overhead. There are proactive and reactive protocols as well as the hybrid of those two solutions which tries to combine the best features of each protocol [1].

In networking, a hop represents one fragment of a path between the source and the destination. It is a well-known phenomenon that data passes through an unknown number of intermediate gateways until it reaches its destination. For example, on the Internet packages are routed between various sub-networks. Moreover, the definition of a hop distance should be useful. It is a unit of measurement used to express the number or routers that a packet must pass through on its way to its destination.

Therefore, in a wireless network, single-hop means that there is only one hop between the source station and the destined host. At the same time, multi-hop refers to a situation when the packet of data must travel through more than one hops. The hop count is important for the basic network operating principles. Fig. 1 clearly presents both expressions.

The perfect routing protocol has to combine the goal of dynamic adjustment to changing conditions in an ad-hoc network and of low overhead. Due to this combination, several different approaches were introduced in the field of routing protocols. Some of them will be presented and discussed in the following sections. Figure 2 shows a possible classification of routing protocols that is taken from Latiff et al. [3].

978-83-60810-53-8/$25.00

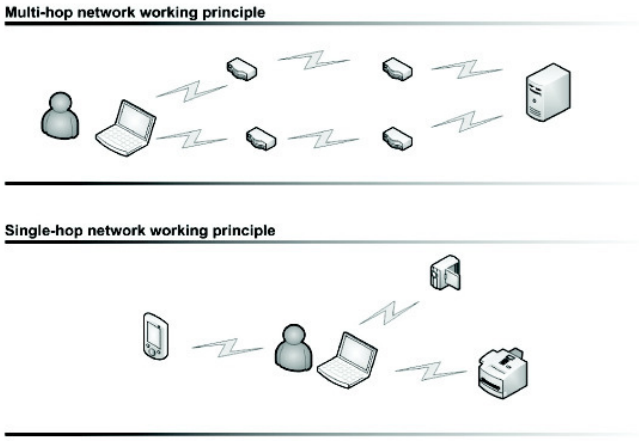


Fig. 1 Examples of single-hop and multi-hop ad-hoc networks.

II. Flooding

According to Mohapatra and Krishnamurthy [4], flooding (network-wide broadcasting or pure flooding) is a way to deliver data from the source node to the destination node through every outgoing link. It means that every attached node will receive source data packets via a MAC layer broadcast mechanism and, finally, every node in the connected component of the network will deliver the data.

There is the basic rule that is followed in order to avoid looping in the network: "every node transmits only once". If a node collects data for the first time, it re-broadcasts it. This algorithm guarantees the end of the procedure eventually and is easy to implement. Additionally, no prior knowledge about the network topology is required and, in some cases, when mobility of the nodes in the network is so high that even unicast protocols cannot handle it, the flooding may become the only reasonable alternative for routing data rationally [4].

This protocol technique can have a big contribution to an overall throughput in the network – the higher number of packets in a network means that there is a higher chance for a collision, what influences the success rate of the packet delivery directly.

III. Proactive Protocols

The main operating principle of proactive routing protocols is that they maintain unicast paths between all pairs of nodes, even when routes are currently idle. They are also called "table-driven" routing protocols. A node can decide to update its routing table after either receiving an update message from a neighbor or detecting a change in the status of a link to a neighbor. Hence, when the source wants to start a connection with a remote destination node, the process can immediately begin because the path is ready and available at any time. No other request or path discovery is required and, therefore, the delay of such nature can be eliminated. It is assumed that the protocols are capable of finding the shortest and the most optimal route for a given model of link costs.

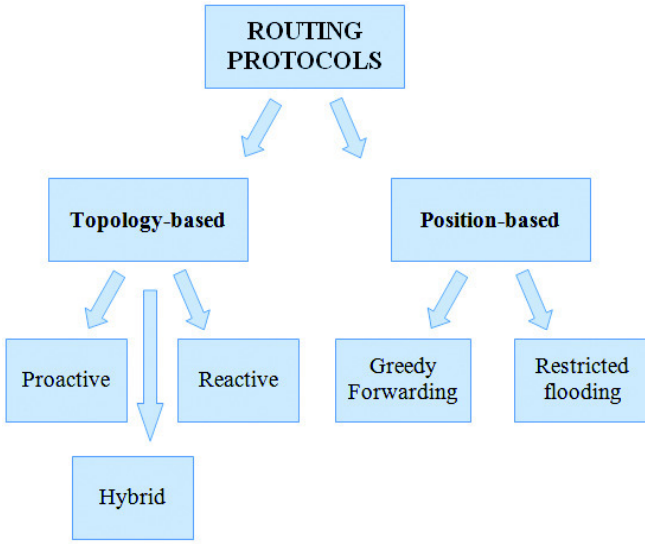


Fig. 2 Categorization of MANET routing protocols.

Optimized Link State Routing (OLSR) is a member of the proactive protocols group and, as such, it is also a table-driven protocol, which assumes that nodes in an ad-hoc network will update each other regularly and will cooperate in order to send data from the source to the destination using the most optimal path. This protocol uses the Mohapatra and Krishnamurthy [4] concept of Multipoint Relays (MPRs), mentioned in this paper before.

In a general operating mechanism, only the nodes that were selected as responsible for their area are allowed to generate link state updates. Additionally, these updates must include information on the links between MPR nodes only [2]. No other node has been granted the privilege to do so in order to keep the update size as small as possible. This way, even though there may be other routes available, only a part of the network topology is revealed to other nodes. This may seem dangerous for network routing; however, this partial information is fully sufficient in order to locally calculate the hop count to every node because it is certain that a path that consists only of MPRs exists.

One of basic principles of OLSR is that it uses only periodic updates in order to keep all nodes up-to-date with the link state. Whenever traffic in a network is dense, the protocol reduces the overhead as compared to the time when traffic is lower or the network is sparse. Additionally, the interval between subsequent updates is critical for reacting to topology changes and should be accurately considered.

What is really distinguished for OLSR is that it can minimize the overhead from flooding of control traffic effectively only by using carefully selected MPRs to retransmit control messages. This way, not all nodes have to be occupied with retransmission but messages still reach all nodes in an ad-hoc network [4]. Moreover, in order to find the most optimal route, the OLSR protocol needs only a partial link state to be sent through the whole network. This minimal information about link states includes the links to all nodes in the region under MPR responsibility (however, the redundancy is also possible).

The performance of the OLSR protocol was also tested in comparison to other various types of protocols. It was discovered that OLSR shows a good resilience to a suboptimal link state situation in a network where the routes are constantly changing (as nodes move in counter-rotating circles) and the network picture never converges permanently [5].

IV. Reactive Protocols

Reactive routing protocols, also called "on-demand" protocols, are quite different from the traditional proactive manner. The main difference lies in a route preservation mechanism – while proactive protocols keep all routes available for use at any time, reactive protocols maintain only the paths that are currently needed. The advantage of this technique is that a huge amount of routing data does not have to be stored and updated all the time. However, good algorithms are needed for instant path discovery that would not create too big delays and queues. Still, this kind of protocols should be perfect for networks where the traffic is small and sporadic.

Ad-hoc On-demand Distance Vector [6] belongs to the reactive protocols family and discovers a route from the source to the destination when it is needed. All possible routes are not maintained the whole time. Basically, AODV relies on the distance vector technique. This term refers to the method which uses the arrays of distances to other nodes in a network. Instead of saving knowledge about all routes in a network, it is enough to know the direction of forwarding the message (or the interface that should be used) or the distance from its destination (in reasonable units).

So, keeping those two basic rules in mind, AODV depends on dynamically established route table entries at nodes between the source and the destination. This means that AODV protocol requires a much larger overhead in order to piggyback source routes in each packet, what is unthinkable for proactive protocols. Each entry consists of the destination address, the next hop address, the destination sequence number, and the hop count.

Another characteristic of AODV protocol is the sequence number, which is incremented monotonically at each node of the network separately. The combination of above features results in an algorithm that can use an available bandwidth efficiently and can adapt to changes spotted in an ad-hoc network.

Each router based on the AODV protocol is more or less a state machine that works using a simple algorithm. If a route exists, then the message is forwarded. Otherwise, the message enters a queue and the router sends a route request in order to search for a possible path. According to received information, the router can update the table and even transmit the message if the path to the destination has built up.

The AODV protocol uses four types of messages that the nodes can distinguish [4]. The route discovery is handled by Route Request (RREQ) and Route Reply (RREP). The needed paths are maintained by Route Error (RERR) and HELLO messages.

Dynamic Source Routing is the reactive protocol that uses a source routing mechanism. The sender of the packet generates a header that can contain all addresses of the nodes in a network which a packet must be forwarded through in order to reach the destination node [7]. It means that the source needs to know the whole hop-by-hop path that can be stored in a route cache. This memory should be maintained by each node of an ad-hoc network which wants to participate in the traffic share. If this cache does not enclose the required path, the node simply needs to use the standard discovery process for the wanted route in order to dynamically determine the path to the destination node. It is accomplished by flooding the network with RREQ messages, also called queries [4].

The route discovery technique is based on route requests which are re-broadcasted by each intermediate node if it is not a destination node or if it does not know the path to the destination based on a route cache. Otherwise, the node answers with the PREP message and the packet with the entire route is sent back to the origination node. And finally, this path is, of course, saved for later in a route cache by each node which does not know it [4]. Like in the case of the AODV protocol, the RERR packet is generated if any node detects a broken link that cannot be longer used for the traffic. This kind of message triggers the removal of the given route from the route cache as well as all entries that are affected [4].

V. Running Simulations

We have decided to use OMNeT++ [8] for the simulations library, which is rather a good provider of infrastructure and tools than a simple simulator of a network. The main advantages of this tool are as follows: free for academic use, the engine runs event-driven simulations of communicating nodes on a wide variety of platforms, support of graphical network creation, the framework is fully extensible and modular (based on C++ language), the documentation and the tutorials are properly maintained and developed by an increasing number of new users, and there exists a great diversity of available libraries and featured projects compatible with the OMNeT++ platform.

We performed the tests only in a random grid, but we were aware of the disadvantages of the linear or grid topology that may cause problems in ad-hoc network routing. Nevertheless, we wanted to have a more realistic topology, so the random one was the most reasonable (Fig. 3). All the nodes in our simulations were moving all the time with variable speed and direction of movement without a pre-determined path.

All tests had the common goal of adjusting the final simulation parameters because the default ones are not always the most suitable for the wanted results. Table (Table I) presents the output of all testing and verification processes.

The preparation for simulation was not only focused on investigating the best parameter set but also on adjusting the behavior of a singular node. The inner construction of the mobile device was based on the TCP/IP model. It was decided that it would employ 802.11g technology for MAC layer, UDP was used in transport layer, and UDP APP for application layer.

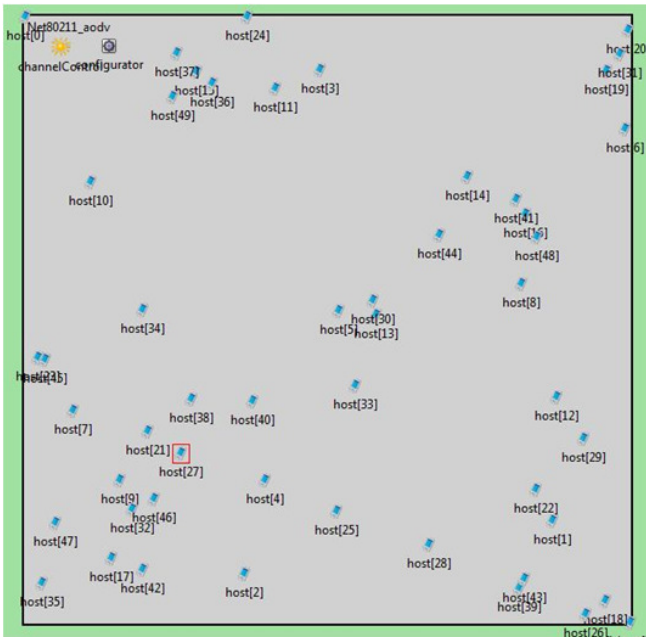


Fig. 3 Snapshot of the random topology.

VI. RESULTS

With all predefined parameters of the simulations, we were able to obtain the relevant output – the average end-to-end delay and the average throughput of the network. The average delay time involves all possible reasons, such as queuing time, packet transmission, and propagation time or retransmission time.

We think that the delay is important for a dynamic ad-hoc network and should be as small as possible but, at the same time, the successful rate must be tolerable. Please have a look at the results that we got separately for the AODV, OLSR and DSR protocols (the plots are presented in Fig. 4, 5 and 6, respectively).

With the increased number of nodes in a network, packet collisions may occur more often and this will lead to a higher number of retransmissions. This kind of situation would definitely influence the overall delay and it can be observed in all of these plots.

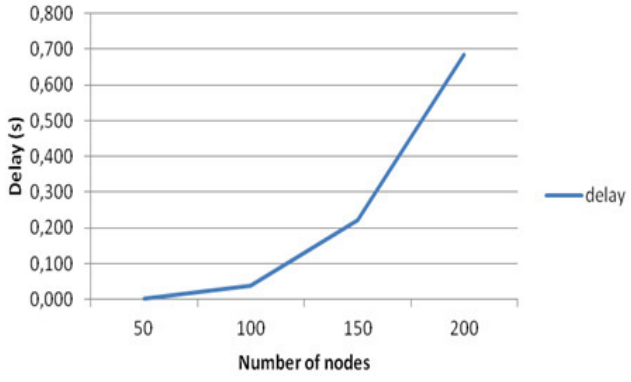


Fig. 4 Average end-to-end delay for the OLSR protocol.

TABLE I.
SIMULATION PARAMETERS CHOSEN FOR THE FINAL TESTS.

Parameter	Value
Simulation time	600 s
Topology	Random location of nodes (network of mobile devices)
Number of nodes	50, 100, 150, 200
Ad-hoc protocols	OLSR, AODV, DSR
Transmission range	100 m
Mobility model	Random way-point

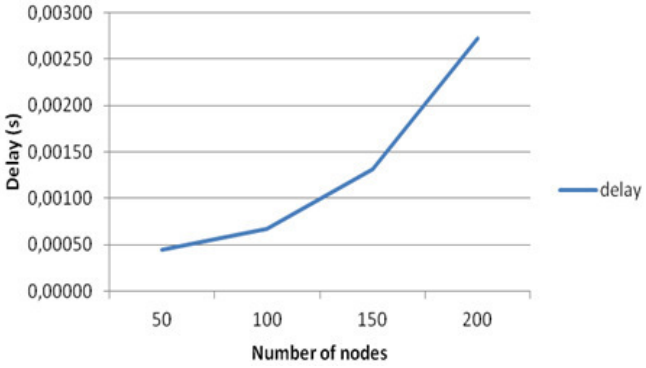


Fig. 5 Average end-to-end delay for the AODV protocol.

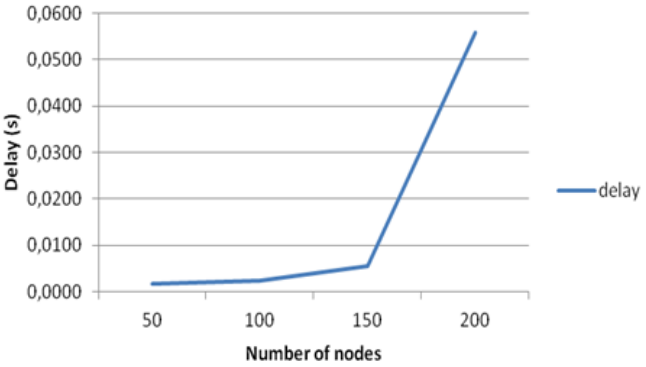


Fig. 6 Average end-to-end delay for the DSR protocol.

However, the smallest delay can be noticed in case of the AODV protocol application.

The average throughput of all three protocols is compared by measuring the average rate of successful message delivery over a communication channel. This is calculated in bits per second, what emphasises the vitality for ad-hoc network operation. The higher this number is, the higher the throughput is.

When the number of nodes increases, more packets of data come to the network; it can be observed that the highest throughput of all three investigated protocols was reported in the AODV protocol (please compare the plots presented in Fig. 7, 8 and 9).

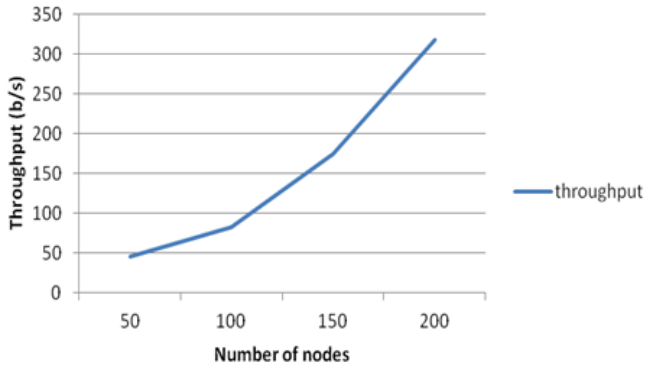


Fig. 7 Average throughput for the OLSR protocol.

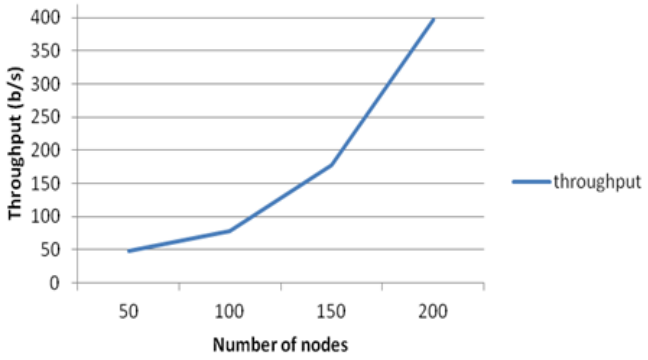


Fig. 8 Average throughput for the AODV protocol.

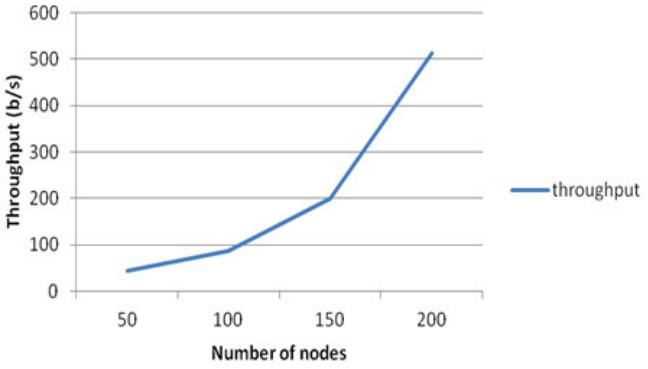


Fig. 9 Average throughput for DSR protocol.

VII. Discussion

We have examined and analyzed three routing protocols from both proactive and reactive groups, namely Ad-hoc On-demand Distance Vector (AODV), Optimized Link State Routing (OLSR), and Dynamic Source Routing (DSR). The simulation tests involved measuring metrics such as throughput and end-to-end delay.

The presented results indicate that the performance of the AODV protocol is superior as compared to the two other protocols that have been taken into account. It can be easily noticed that the AODV protocol handles the network traffic better when more and more nodes are added to an ad-hoc network. Still, according to the theoretical background, smaller networks (up to 10 nodes) might have been handled better by the DSR protocol.

Nevertheless, the DSR protocol was inferior in both end-to-end delay and throughput, even when there were only 50 nodes in the network. The poor performance of DSR with respect to time for packet delivery is mainly due to caching the routes and the lack of mechanism for deleting the stale paths. We think that it is the reason why DSR did not perform so well, even though it belongs to the same group of reactive protocols as the AODV protocol.

It was also observed that, for a relatively small number of nodes, all routing protocols are similar when throughput was under test. However, the average end-to-end delay could be easily compared for even the smallest number of nodes and the AODV protocol was incredibly fast in delivering packets. This is why we assume that the AODV protocol would be preferred for real time traffic over DSR or OLSR.

Whenever throughput is considered, results show that DSR does not handle it well because it consumes a considerable amount of power. If it was a real environment involving mobile devices, the batteries would run out of power pretty quickly.

During the testing process, some problems occurred and those influenced directly the developing and testing time. The parameters were difficult to adjust because the default ones did not give the wanted results and looking for better values consumed a lot of time. Additionally, testing the performance of the DSR protocol was difficult in terms of the processing power of computer (this protocol uses the caching routes technique, so the bigger the number of nodes was, the more resources it was consuming for the test). Finally, the OMNeT++ testing environment gives different debugging messages in each operating system, so we had difficulties solving, for example, Cygwin problems during the testing process.

References

[1] R. Hekmat, Ad-hoc Networks: Fundamental Properties and Network Topologies, Springer, 2006
[2] S. K. Sakar, T.G. Basavaraju, C. Puttamadappa, Ad-hoc Mobile Wireless Networks. Principles Protocols and Applications, Auerbach Publications, 2008
[3] L.A.Latiff, N. Fisal, S.A. Arifin and A. Ali Ahmed, Directional Routing Protocol in Wireless Mobile Ad Hoc Network, article from "Trends in Telecommunications Technologies", book edited by Christos J Bouras, 2010
[4] P. Mohapatra, S. V. Krishnamurthy, AD-HOC NETWORKS: Technologies and Protocols, Springer Science, 2005
[5] J. Hsu, S. Bhatia, M. Takai, R. Bagrodia, M. J. Acriche, Performance of mobile ad hoc networking routing protocols in realistic scenarios, IEEE Military Communications Ceonferenc, MILCOM 2003, 2003
[6] C. E. Perkins, E. M. Royer, Ad Hoc On-Demand Distance Vector Routing, In Proceedings of IEEE Workshop on Mobile Computing Systems and Applications (WMCSA), 1999
[7] D. Johnson, D. Maltz, Dynamic Source Routing in Ad Hoc Wireless Networks, Mobile computing, 1996
[8] OMNeT++ Network Simulation Framework, www.omnetpp.org

Proceedings of Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1315–1318

User Positioning System for Mobile Devices

Artur Sierszeń
Lodz University of Technology, Institute od Applied Computer Science, ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
Email: artur.sierszen@p.lodz.pl

Łukasz Sturgulewski
Lodz University of Technology, Institute od Applied Computer Science, ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
Email: lukasz.sturgulewski@p.lodz.pl

Karol Ciążyński
Lodz University of Technology, International Faculty of Engineering, ul. Żwirki 36, 90-924 Łódź, Poland
Email: karol.ciazynski@gmail.com

Abstract. **In the recent years, the Global Positioning System (GPS) has become a standard for the location and navigation for a huge number of people all over the world. This system is unquestionably one of the most significant developments of the twentieth century. GPS employs a great variety of applications from car navigation and cellular phone emergency positioning even to aeronautic positioning. Despite the fact that it plays an essential role in today's world, GPS has some limitations. The main disadvantage is the inability to operate inside the buildings because of the loss of signal from the satellites. During the last decade, the interest in location based services has significantly increased. It is related to the existence of ubiquitous computers and context awareness of mobile devices. Information about the position plays the great role in the field of security, logistics and convenience nowadays. Thus, it is necessary to fill the gap at the point where Global Positioning System does not perform satisfactorily.**

I. Introduction

THE idea is to design the system complementary to GPS which would be able to determine the location in the places where GPS is not. It is possible to use a wireless LAN system and its existing infrastructure to find the users location indoors. The wireless communication system was designed to provide users with a possibility to move around and still be connected to the local network or be able to use Internet access without cables. Nevertheless, it is possible to use some of the properties of wireless communication to determine the location. Analyzing the transmitted signal, a mobile device can estimate the distance between the access point and the terminal itself. Then, by combining the measurements from more than one access points, the mobile device can determine the exact location of the terminal.

II. WLAN Positioning Principle

In recent years, the WLAN IEEE 802.11g standard has gained high popularity; the number of devices using wireless networks is still constantly increasing. Nowadays, WLAN infrastructure is maintained nearly everywhere where people appear frequently. This enables customers to connect to the Internet in public places such as airports, hospitals, universities, or shopping malls. The majority of modern mobile devices is equipped with a WLAN interface and that enables them to connect to wireless access points. In recent years, the position information as well as the WLAN standard IEEE 802.11g have become very common. This motivates developers to produce systems based on WLAN networks which are able to determine a user's location.

To estimate the real-time location of a user, location systems have to perform a number of steps and various calculations [1],[3]. The estimation of the distance from the access point is the first phase needed to determine the exact location. It is the method of calculating the radius of a circle in two dimensional spaces or a sphere in three dimensional systems. The calculation of more than one distance from several Access Points (APs) could be used to estimate the exact location [2]. A location could be described as a set of coordinates pointing at a particular position in a space or on a map. The majority of the positioning applications require that the position of the user's device be estimated with a good accuracy, but sometimes another criterion, such as no complexity or low costs, is more important. Positioning systems based on wireless networks use the properties of the access to a medium. They use various physical attributes to measure the distance between two terminals [4-5].

The main principle states that signal strength at the receiver is inversely proportional to the square of the distance that the signal travels. Based on that rule and the characteristics of a wireless signal in a researched environment, the distance between two terminals can be determined. In comparison to methods that are based on the time of flight technique, the Received Signal Strength (RSS) approach has a number of advantages. To apply the RSS method, no hardware changes are usually required. This method can be implemented in a customary wireless communication system such as the IEEE 802.11g standard with the facility to read a received signal strength indicator. The special synchronization and timing techniques are not relevant. Owing to the low costs of implementation and the simplicity, the RSS approach has a great chance to become the most popular technique used for indoor positioning systems. Nevertheless, there are some restrictions of the RSS location technique. The electromagnetic signal is very prone to interferences and the effect of multipath propagation. A position awareness system based on RSS method must use a specific database or increase a number of static base stations to achieve higher accuracy [6][8].

978-83-60810-53-8/$25.00

The RSS fingerprinting approach is based on sampling and recording of characteristic patterns of a radio signal in a specific environment and is called pattern recognition or fingerprinting. The location patterning technique is implemented in the software entirely. This reduces the complexity as well as the cost of performance and, at the same time, guarantees high positioning precision. However, a special database must be created for every single area and every change which effects radio propagation requires the re-creation of the whole database[7],[9].

The deployment of a location system based on the position patterning has two measurement phases. The first one, which is called the offline or the calibration phase, results in the creation of the database. The second phase, called the online or operational phase, takes place when the real time signal strength values are matched with the previously constructed database components associated with the reference points.

III. Project Implementation

The created WLAN Positioning applies the position determining approach based on the received signal strength and access point information from the 802.11g wireless network. Tests were conducted in a three-storey building with the overall surface area of about 250 square meters using three access points which were distributed all over the building, each on a different floor.

The client-based approach is applied in the system that includes the offline phase as well as the online phase. The developed application could operate in both modes. The first one is a calibration mode and includes the construction of a radio map of an indoor area containing measurements of the received signal strength from access points in each reference point.

The localization system uses three access points that provide overlapping coverage area, with the strongest signal power in the calibrated arena. Another wireless networks detected by the system are not used by the application because they do not guarantee the sufficient coverage and the high enough signal strength value. The second mode of application is the operational mode and this mode includes the determination of position in real time. In this phase, the application receives a signal from fixed access points and performs the calculations on signal strength to obtain the current position of the device. The approach used to determine the device position is the nearest neighbor method based on the Euclidean distance.

The designed positioning system operates on mobile devices in conjunction with the 802.11g standard wireless access points. The calibration and the operational phase were conducted using already installed access points in the three storey building. The coverage region includes three floors with the surface area of about 250m2. This place consists of 15 different locations where 55 reference points were defined. The plan of each floor with the position of the reference points is presented in Fig 1.

The User Positioning System is a software-based project developed to work on the Sony Ericsson Xperia mobile device. The mobile phone operates using Android 2.1 and is equipped with WLAN card working with 802.11g technology.

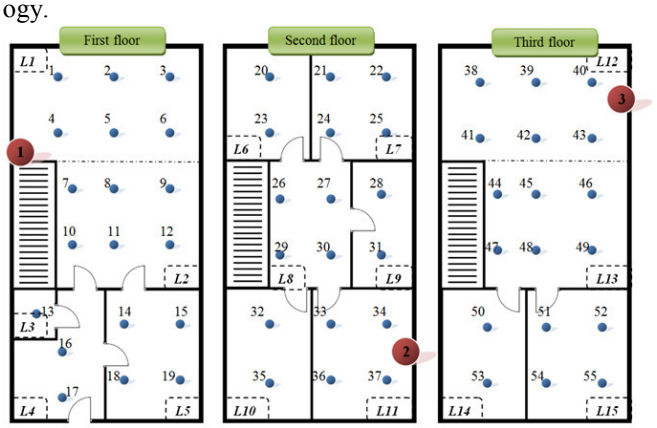


Fig. 1 Plan of the building with tagged reference points.

Due to the fact that the project was designed to work on the mobile phone Sony Xperia, there were some specific hardware restrictions. The system uses the embedded wireless network card and its ability to deliver the received signal strength values. The positioning approach could be chosen from the set of methods that use the RSS approach to determine the location. The approach based on the angle or time of arrival may not operate on available mobile device. (In order to measure the angle, it is necessary to use a directional antenna or an antenna array, where each antenna is tuned to a different optimum frequency. Unfortunately, it is not possible to use this method on a mobile phone, which uses an antenna with properties similar to these of omnidirectional antennas. No possibility of tuning the frequency of the antenna makes it also impossible to construct a suitable antenna array).

Due to the fact that the system was created to operate in a very complex environment where the propagated signal was prone to various distortions caused by multipath phenomena, the propagation model method could not be applied in the described system. The approach which could give the satisfactory results in that complex environment and which was applied in the project was the received signal strength fingerprinting approach that is based on the nearest neighbor algorithm.

The User Positioning System could operate in two modes. The first mode uses the offline phase which contains a collection of reference points, what results in the creation of the database. This phase is performed before the real-time operational phase and provides references for the localization algorithms used in the actual position localization. The constructed database consists of reference points at a specific location and three median values of RSS that correspond to the fixed access points. The main database is associated with the described environment – the three-storey building. This database is used in the real-time operational mode as a reference databank for the application to find the most accurate position of the device.

The second mode of the program is called the positioning phase. In that mode the application matches real time measurements with the database that was already created in the

previous phase. The mobile application in the positioning mode measures the real time RSS values from three access points. Due to hardware limitations, the program could obtain only one sample per two seconds. During this phase, the analysis including a number of sample values resulting from the calibration process was compared with the systems accuracy. The application depending on user preferences could operate in three different modes that vary in the number of samples. The first one uses instantaneous measurements of the RSS value to determine the best reference point, while the other two use three and nine samples to localize the device. The corresponding time of the calibration is two, six, and eighteen seconds. The received RSS values from more than one sample are translated into their analogue median values in order to create a single vector of signal strength values.

The constructed real time vector of measurements is compared with the already created database, using the nearest neighbor algorithm which was already described in detail in the previous section. Each value of received signal strength in the vector is compared with the corresponding values with each reference point from the database, using Euclidean distance. The reference point with the smallest distance is considered the best position by the program. The application displays the determined position at the screen.

IV. Test Measurements

The tests of the project were carried out with various parameters. The actual accuracy of the system varies in the number of samples obtained in real time measurements. The results presented below are divided into three parts as three different approaches have been considered.

The first part includes only real time RSS values from all access points. This leads to a very quick determination of the position; however, this approach cannot provide the efficient accuracy. The next approach compromises on the time of calculations and the accuracy and uses the measurements of three samples which require six seconds for locating the position. The last approach uses measurements of nine samples which give the best accuracy; however, this method takes eighteen seconds to determine the current position.

The first approach, which measures one instantaneous RSS value, does not provide the satisfactory accuracy (Fig 3.). The analyses have shown that only in 17% of the tested positions the location was determined correctly. Only in 36% of the cases, the position was estimated correctly as being located within a room.

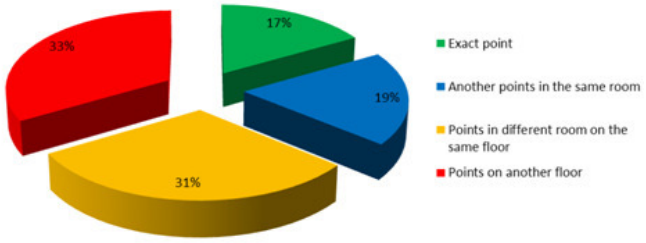


Fig. 3 Accuracy for the single sample approach.

The second approach, which measures three instantaneous RSS values, provides better accuracy (Fig 4). The analyses have shown that in 32% of the tested positions the location was determined correctly and in 53% of cases the position was estimated correctly as being located within a room.

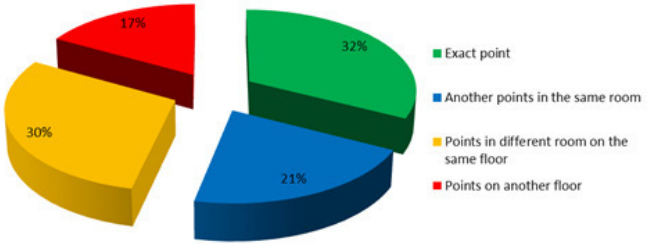


Fig. 4 Accuracy for the 3 samples approach.

The last approach, which measures nine instantaneous RSS values, provides the best accuracy. The analyses have shown that in 48% of the tested positions the location was determined correctly and in 81% of cases the position was estimated correctly as being located within a room.

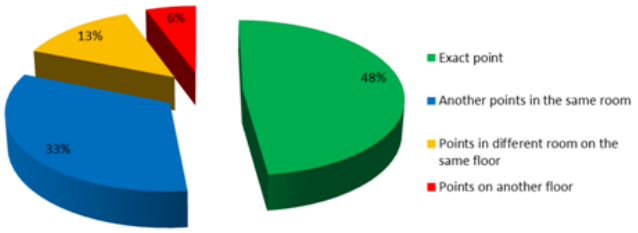


Fig. 5 Accuracy for the 9 samples approach.

V. Conclusions and Future Work

The system tests were performed with three different approaches. They vary in the amount of samples in real time measurements, what results in the difference of position estimation time as well.

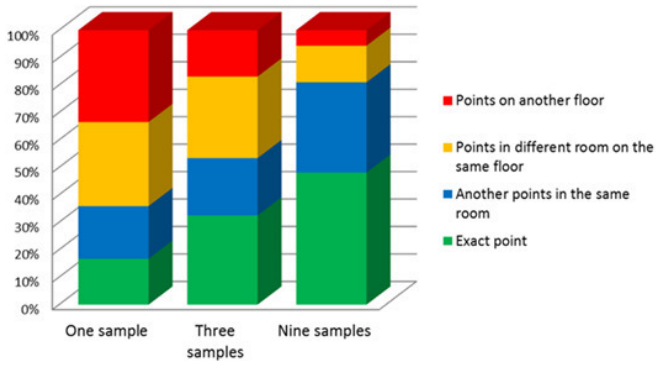


Fig. 4 The comparison of the system performance for the three approaches.

Undoubtedly, the approach that measures 9 samples provides the best accuracy. However, 18 seconds to obtain the position is too long (during this time the real positioning can change). The compromise might be the three samples approach that has the satisfactory accuracy and does not take too long to calculate the location.

The goal of the project was to design the positioning system for mobile devices that would perform satisfactorily in an environment where the Global Positioning System could not operate effectively. The project was successfully implemented using the already existing infrastructure of WLAN networks in a three-storey building. The most suitable approach for WLAN positioning was chosen after a careful literature review. The method implemented in the localization system is the RSS fingerprinting technique based on the nearest neighbor algorithm.

The User Positioning System for Mobile Devices that works on mobile phones with the Android operating system has been successfully implemented. The designed application could work in two modes. The program working in the first mode constructs the database with access points and RSS values corresponding to them. The result of the calibration mode is the database that is a reference map for the operational mode. The second mode operates in the localization mode and compares real time measurements with the database in order to point the current position. The system has been tested in a three-storey building and has achieved the accuracy of about 80% for the localization within 20 seconds with the accuracy to one area of one room (about 15 square meters). The User Positioning System belongs to a minor group of applications which were developed to operate on mobile devices. The implemented system is not expensive and it can become very popular for handheld indoor positioning because nowadays a lot of people have Android mobile phones and the wireless local area networks are becoming more and more common.

Nevertheless, the system implementation could be improved in some steps in order to increase accuracy and effectiveness. Firstly, the accuracy could be improved by implementing more access points in the building. The number of access points will not prolong the time of calibration dramatically but the higher number of access points is, the better quality of signal strength is and, as a result, the better accuracy of the system is. Moreover, the implemented system is not resistant to environmental changes. Each little change in an environment such people moving, doors closing or opening or even furniture moving could dramatically increase the position inaccuracy. The solution could be the creation of a flexible database that would be able to adapt to environmental changes.

REFERENCES

[1] Salter J., Li B., Woo D., 802.11 Positioning In Home, University of New South Wales, Sydney 2008
[2] Nuno-Barrau G., Paez-Borrallo J., A New Location Estimation System for Wireless Networks Based on Linear Distriminant Functions and Hidden Markov Models, Hindawi Publishing Corporation., Madrid 2006
[3] Zekavat S., Tong H., Tan J., A Novel Wireless Local Positioning System for Airport (Indoor) Security, Dept. of Electrical and computer Engineering, Michigan Tech University, Houghton 2007
[4] Hur H., Hyo-Sung A., Hybrid-style Wireless Localization Network for Indoor Mobile Robot Applications, Gwangju Institute of Science and Technology, Gwangju 2009
[5] Hyo-Sung A. Wonpil Y., Indoor Localization Techniques based on Wireless Sensor Networks, Mobile Robots – State of the Art in Land, Sea, Air and Collaborative Missions, Shanghai 2009
[6] Lorincz K., Welsh M., MoteTrack: A Robust, Decentralized Approach to RF-Based Location Tracking, Harward University Division of Engineering and Applied Sciences, Cambridge 2009
[7] Prasithsangaree P., Krishnamurthy P., Chrysanthis P., On Indoor Positioning Location with Wireless LANs, Pittsburgh 2006
[8] Li B., Quader J., Dempster A., On outdoor positioning with Wi-Fi, Journal of Global Positioning Systems, Sydney 2008
[9] Mustafa Y., Agrawala A., On the Optimality of WLAN Location Determination Systems, Maryland 2006

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1319–1325

Development of a Mobile Application for People with Panic Disorder as augmentation for an Internet-based Intervention

Stefan Kleine Stegemann*, Lara Ebenfeld*, Dirk Lehr*, Matthias Berking†, Burkhardt Funk*
*Leuphana University Lueneburg, Germany
†Philipps University of Marburg, Germany

***Abstract*—Smartphone technology has recently gained attention in the field of E-Mental Health research and mobile applications for measuring health-related aspects as well as mobile mental health interventions have emerged. However, little work has been done on leveraging mobile technology in combination with internet-based interventions. We argue, that mobile applications can not only enrich mental health treatments but also foster the commercial success of E-Mental Health applications. To this end, we have developed GET.ON PAPP, a mobile application for panic disorder that integrates an internet-based treatment into daily life. In this work, we present the development and structure of GET.ON PAPP and a perspective for its evaluation.**

I. Introduction

E-MENTAL Health, the use of information and communication technology to "support and improve mental health conditions and mental health care" [1], has been and still is a growing field in research and practice [2]. One of the most prominent forms of an E-Mental Health application is the internet-based intervention (IBI), the use of web-sites to deliver self- or low-guided psychological treatments over the internet. With the advent of smartphones in the past years, however, researchers recently started to explore the potential of this technology for E-Mental Health [3, 4].

The ubiquitous nature of smartphones makes them ideal for measuring data on health conditions in daily life. To this end, a number of applications have been developed and evaluated. Examples are the MONARCA self-assessment system [5, 6] and the work of Morris et al. on emotional self-awareness [7]. Furthermore, smartphone sensors have been used in order to amend or replace self-reported data with more objective measures [8, 9].

Self-contained interventions are another area where mobile technology is used in E-Mental Health [3]. Examples include mobile interventions for bipolar disorder, schizophrenia and depression [10, 8]. However, to our knowledge, little work has been done on using smartphones in combination with IBIs.

We argue that by combining mobile applications with IBIs, smartphones can be leveraged to overcome limitations of the internet-only approach, thereby integrating the treatment more closely into daily life. This is important not only from a psychological point of view but also from a commercial perspective. In Europe and in particular, in the Netherlands, UK and scandinavian countries, IBIs are currently being integrated into the routine healthcare system. In recent years, vendors have emerged that specialize in development and distribution of E-Mental Health systems. The use of new technologies together with established approaches can make systems more attractive for both, customers who offer mental health services and end-users.

To explore the potential of a mobile application in combination with an IBI, we developed the GET.ON Panik program, a treatment for people with panic disorder with and without agoraphobia. The treatment consists of two parts, an online training that is delivered via the internet-browser and the mobile application GET.ON PAPP that amends the training. In this paper, we present GET.ON PAPP and discuss its foundation, structure, development process and perspective.

The GET.ON[1] project is an international and interdisciplinary cooperation of researchers and practitioners from the Leuphana University Lueneburg, Phillips University of Marburg, VU University Amsterdam and Minddistrict, a Dutch company specialized in the development and distribution of IBIs. The goal of the EU-founded project is to develop, evaluate and disseminate interventions for common mental disorders, including depression, insomnia and panic disorder.

This paper is structured as follows. First, a brief introduction is given into the background of panic disorder and its treatment as well as IBIs and their limitations. After that, the development process and structure of GET.ON PAPP are presented. Following this, we discuss the evaluation perspective of the mobile application and finally, a conclusion is drawn.

II. Background

Panic disorder (PD) is a common and severe mental disorder that belongs to the category of anxiety disorders. People suffering from PD experience recurrent panic attacks, which happen in an unpredictable manner. A panic attack involves heavy body-symptoms such as palpitations, breathing difficulties and dizziness[2]. Because of this, people often fear that they will die which further exacerbates the symptoms. Panic disorder often occurs in combination with agoraphobia, the anxiety about being in places or situations in which help might not be available or from which escape might be difficult. Examples include driving a car, shopping in a supermarket and going

[1]GesundheitsTraining.Online, http://geton-training.de
[2]cp. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) for diagnose criteria

978-83-60810-53-8/$25.00

to the cinema or theater. As a consequence, people tend to avoid such situations and over time, the fear of panic attacks often becomes a central problem because it severely limits their lives. This leads to a high level of distress as well as social and job-related disabilities [11].

Panic disorders are one of the most prevalent anxiety disorders. Goodwin et al. reported (12-month) prevalences between 0.3% and 3.1% for PD with and without agoraphobia in Europe [12]. For 2010, the total economic cost for Europe was estimated to 11,894 million EUR PPP[3] for PD and to 9,634 million EUR PPP for agoraphobia [13].

A. Treatment of panic disorder

Cognitive behavior therapy (CBT) is a commonly used and effective method for treatment of panic disorder [14] that is adapted by the majority of contemporary computer-based treatment programs, not only for PD but also for other mental disorders [15]. In CBT, "patients are trained to collect information in a systematic fashion to offset the influence of maladaptive information-processing strategies and to conduct behavioral experiments to test the accuracy of their negative beliefs" [16].

CBT for panic disorder typically incorporates two central components: cognitive restructuring and exposure exercises. Cognitive restructuring teaches strategies on how to identify and change negative thoughts. Exposure is a technique where people with PD and agoraphobia expose themselves to situations where they fear to have a panic attack. For people with PD without agoraphobia, a method is used where specific exercises such as hyperventilation are carried out in order to trigger body symptoms similar to those that occur during a panic attack. In both cases, it is vital to the success of the treatment that exposures are done frequently and properly.

In addition to these core components, clients[4] are usually encouraged to keep a self-monitoring diary. This not only helps the therapist to give feedback but also increases the client's awareness for panic symptoms and disorder patterns. Furthermore, self-monitoring accounts for the natural human drive for self-understanding [17] and can thereby reinforce a client's involvement with the treatment.

B. Internet-based interventions

An internet-based intervention for a mental disorder is a variant of computer-based psychotherapy [15] that is provided over the internet. In a nutshell, an IBI is similar to a self-help textbook with the content being delivered over the web-browser, usually in an interactive form and enriched with multimedia elements. Participants work through the intervention either completely independent or guided by a coach who gives feedback and motivation [18].

Compared to traditional face-to-face therapy, IBIs are a low-threshold form of intervention that offer high availability at any time as well as anonymity and thus less stigmata. Protection of privacy is another commonly noted strength of IBIs [2, 15].

Psychological research has shown effectiveness of IBIs for several mental disorders, including PD with and without agoraphobia [18, 19]. A couple of online-programs for PD were developed and evaluated in the past years, such as, for example FearFighter[5] in the UK and Panikprojektet[6] in Sweden. Furthermore, IBIs are now gradually implemented in practice as part of the routine healthcare system in Europe. The leading countries in this development are the Netherlands and Sweden but others start to catch up.

C. Limitations of IBIs

IBIs commonly make use of an internet-browser that is running on a stationary computer or laptop to deliver the treatment. This approach is advantageous to serve larger blocks of information as well as for exercises or quizzes where users fill in (textual) answers. The timeframe of a single interaction between an user and an IBI is typically rather long, ranging up to several hours.

While mobile devices can be used to access an IBI on the go, screen size and unstable internet-connections often impose problems. In addition, entering text on a mobile device can be cumbersome as most of them do not have a dedicated keyboard [20]. As a result, IBIs are limited when it comes to supporting clients in their daily life.

While the above can be said for IBIs in general, we identified the following two key problems with respect to panic disorder.

1) In-situ support: Exposure is a notoriously difficult task to perform. Clients not only need to overcome their inner resistance but also carry out the exercises in a specific and prescribed way while at the same time experiencing high levels of anxiety. Although principles and procedures can be described in an IBI, exposure usually happens far away from the computer. As a consequence clients need to memorize and recall what they have to do. In addition, they can not easily verify that the exposure was performed correctly.

2) Self-monitoring: Diaries are already an inherent part of many contemporary IBIs. However, the limited accessibility of these programs requires clients to recall and sum up their panic attacks and anxiety feelings, typically on a day-to-day basis. As a result, a retrospective bias is introduced, which reduces the ecological validity of the diary data [3]. This is not only unfavorable for clients and therapists but also for the researcher who wants to analyze the data.

III. A MOBILE APPLICATION FOR PANIC DISORDER

Smartphones have become increasingly powerful and almost ubiquitously available in the past years. Consequently, they "play a vital role in the practice of medicine today" [21]. In a review on mobile technology in psychological treatments, Heron and Smyth argued that mobile technology is in particular suitable to bring interventions closer to people's daily

[3]Purchasing power parity

[4]In a psychotherapeutic context, people who seek the help of a psychotherapist are commonly referred to as "clients".

[5]http://www.fearfighter.com/

[6]http://www.kbt.info/behandling/

life [3]. To this end, we have build the mobile application "GET.ON PAPP"[7] for people with PD (with and without agoraphobia) to overcome the limitations we identified above by providing a tool that integrates the treatment in their daily life, where panic actually happens. GET.ON PAPP is not designed to be used stand-alone but as an extension to an IBI that structures the treatment and offers information as well as instructions. In doing so, we strive to use both technologies to their potential. For the remainder of this paper, however, we will concentrate on the mobile part only.

GET.ON PAPP is structured into a diary that covers the self-monitoring aspect and an exposure-guide that supports people in performing exposure-exercises. In the following, we describe and discuss the development process, the different parts of the application and the technologies we used for its implementation.

A. Development

GET.ON PAPP was developed in an iterative process, incorporating experts from various disciplines. Each iteration started by creating a paper mockup for a specific function together with psychologists. The mockup was then refined until it represented the functionality to our satisfaction. In the next step, we created a functional prototype that could be executed on a smartphone. After that, the prototype was tested and gradually improved.

Researchers have mentioned the importance of credibility for the adoption of IBIs [22, 23]. While it is unclear if these findings can be generalized to mobile interventions, Fogg argued that credibility is crucial to facilitate behavioral change in general [24]. He defines credibility has the combination of perceived trustworthiness and perceived expertise and highlights the importance of a product's visual appearance for the perceived first-hand and long-term experience [25, 24]. For that reason, we included visual and interaction designers in the team from the early project stages on.

While it was originally planned to also incorporate potential end-users early in the process, it turned out to be difficult to find people who were willing to talk about their experiences with the disorder. Instead, we decided to employ psychotherapists and researchers with experience in the treatment of PD. Furthermore, we conducted regular but informal tests with team members who were not directly involved with the development.

B. Documentation of panic-related events

As mentioned before, self-monitoring is an important aspect in CBT based treatment of PD. To this end, a diary has been created that enables clients to document and view their panic events. This is not only helpful to increase awareness for how, when and where panic attacks develop but also as an introduction to the treatment that encourages people to reflect on their disorder.

The frequency of panic attacks varies greatly between individuals, ranging from few attacks a week to several attacks a day. Some clients do not even have panic attacks but live in a more or less constant fear of an attack. Therefore, we opted for an event-based design for the documentation of panic-related events [26].

We define a panic-related event as either a full-blown panic attack or the feeling of fear that a panic attack may happen in an ongoing situation. To reduce the retrospective bias, clients are encouraged to document these events immediately after their occurrence. Smartphones are ideal for this task because they are accessible most of the time.

Because panic-related events are documented "on the go", an efficient and unobtrusive user interface is required. For this reason, we tried to reduce the information that a user has to enter to a minimum. We discussed the attributes that are required to accurately describe a panic-event with scientists from the field and psychotherapists. As a result, a set of four mandatory items has been developed, which are answered on a Likert-like scale from 1 to 10.

A goal was to make the rather boring task of entering a number more interesting and engaging yet efficient. Based on the single-dimension mood-scale presented by Morris et al. [7], we created a novel input component for entering data on a numeric scale (fig. 1). Users select a value by moving a finger up and down on the display. The number moves with the finger and the intensity of the background color changes in order to provide an additional visual feedback (high values have a high intensity and vice-versa).

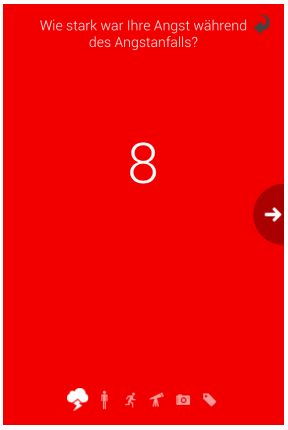


Fig. 1. Entering data on a scale from 1 to 10

To facilitate learning and recognition throughout the interface, we assigned a dedicated color to each question, or more precisely to the concept that underlies a question. For example, the question "How strong was your anxiety during the panic attack?" relates to the concept of anxiety, which is represented in red color. Moreover, the use of colors is not restricted to the diary. Wherever a particular concept appears in the application, the same color is used for it's representation.

We did not do any formal usability testing of the input component so far. However, we conducted a couple of informal tests with PhD-students, asking them to document a panic-event. There was no further information provided on how to

[7]A supercool acronym for GET.ON Panik App

use the component. We observed that when the interface was first presented to the students, it was not immediately clear how to operate it. However, most of them immediately realized the principle when touching the screen and reported that they liked the use of colors. To overcome the initial barrier, we added instructions to the screen which disappear when the user moves the finger. That said, formal usability testing is needed to validate if the design actually works as intended and how it compares to alternative approaches.

Finally, we wanted users to be able to give individual meaning to a panic-related event in a non-prescribed way. We anticipated that some users are more visually oriented while others tend to prefer text. Therefore, we included two additional options in the documentation of events. First, we encourage users to take a photo of either the situation where the event occurred or something that is related to the situation. Second, they may enter textual notes in order to remember feelings or specific aspects of the situation.

C. Visualization of events

To make sense of the past, the diary allows users to browse through panic-related events. For the visualization, we developed a non-technical but more casual approach that is shown in figure 2. Pousman, Stasko and Mateas have discussed the need of casual information visualization for non-work related tasks and populations that are not experts in a domain [27]. To this end, we build on the work of Ljungblad, Skog and Holmquist on ambient information displays [28, 29] and developed a Mondrian-style approach to visualize the four questions that are used to document a panic-related event.

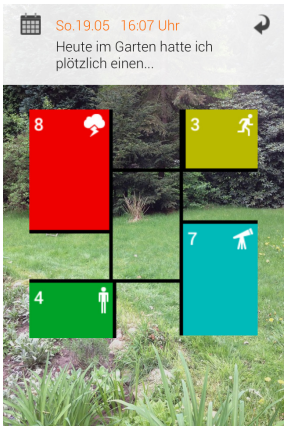


Fig. 2. Mondrian-style visualization of a panic-related event

Four colored rectangles are used to display the information. Each rectangle corresponds to a specific question which is represented by the color and an additional icon. The size of the square as well as the intensity of the color is used to visualize the value between 1 and 10. Apart from aesthetic aspects, this type of display has the additional advantage that users can see the values in relation to each other and thereby identify patterns such as the relationship between anxiety and avoidance. Furthermore, if a photo has been taken, it is displayed as background image to facility fast recognition of the documented event.

In their work on informative art, Ljungblad, Skog and Holmquist reported that their Mondrian-style display created an aesthetic experience for viewers [28]. However, they also noted that, without further information, no one was able to understand how the display actually worked. Some did not even recognize the artifact as an information display at all. That said, with a brief introduction, most people quickly grasped the underlying concepts. We got similar results when presenting our display to students, although at least some candidates interpreted it correctly without any help. This might be tributed to the fact that the context of the display was clear and that the amount of visualized information is rather small in comparison to the work of Ljungblad, Skog and Holmquist. However, in order to ensure that people understand the display properly, we give them a brief introduction.

D. Daily summaries

Fogg notes that giving users ongoing information about their state and progress on a task can help to stimulate their intrinsic motivation [24]. Therefore, we wanted to give clients the ability to document their progress and to be able to track their personal development over the course of the treatment. Furthermore, the development of a client is important not only from a client's perspective but also from a researcher's as well as from a therapist's point of view.

Panic-related events are not suitable to document progress because frequency as well as severity of attacks varies over time, even for a single person. Apart from this, it may be the case that a client has more frequent and more heavy panic attacks at the beginning of the treatment. From a therapeutic point of view, however, it is important how much the PD limits a client's life. To this end, the general level of anxiety as well as the degree of avoidance [8] are relevant variables. Building on this idea, daily summaries have been added to GET.ON PAPP in order to measure these variables not at the event but at a daily level.

A fixed-schedule diary was used to implement daily summaries [26]. Clients are asked to choose a fixed time each day, preferably in the evening, where they reflect on the past day and fill in a daily summary. In order to visualize a client's development and progress, the application can plot the daily summaries over time. Furthermore, the plot illustrates how exercising can reduce panic symptoms by depicting for each day how many exposure exercises have been performed.

E. Guiding exposures

When technology is used as a tool that leads people through a process which would otherwise be difficult or impossible to perform, it can support the change or adoption of a desired behavior [24]. As mentioned above, exposure exercises are the most difficult part of a treatment. Motivating people to confront themselves with a threatening situation or body

[8]Avoidance refers to the fact that people with PD and agoraphobia often avoid situations in which they fear to have a panic attack.

symptoms is difficult. We argue that by creating a tool that guides users and offers feedback, we can lower the barrier, thereby making it easier to persuade people to actually face exposure tasks. Therefore, we integrated an exposure guide into GET.ON PAPP.

The exposure guide supports a) in-vivo exposures, where people approach a situation in which they fear to have a panic attack and b) interoceptive exposures, where body symptoms are triggered by carrying out body-exercises. The guide is constructed as a wizard which leads the user through a sequence of steps while at the same time offering instructions and orientation. In addition, at certain points, the user is asked to answer questions in order to measure if the exposure has been performed correctly (for example the level of anxiety, the degree of avoidance and the severity of body-symptoms). For the sake of consistency, the same input component as in the diary is used to answer those questions.

Therapists noted that when they perform exposures together with clients in a face-to-face setting, people are sometimes so proud of what they achieved that they ask for a photo of the situation. Consequently, we added the ability to take a photo with the smartphone after an in-vivo exposure is completed. Users can view these photos in a gallery together with the feedback and a description of the situation. We hypothesize that taking photos and looking at them retrospectively can act as a self-rewarding mechanism and hence increase motivation for the treatment.

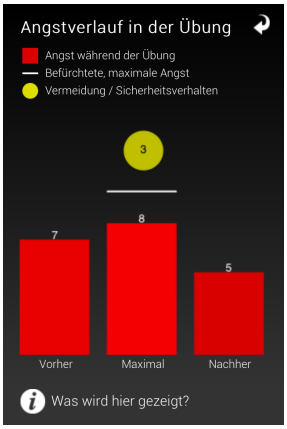


Fig. 3. Feedback for an in-vivo exposure

The feedback after an (in-vivo) exposure allows people to see if the exposure worked and to reflect on the situation (fig. 3). For example, it is important to stay in the situation until anxiety begins to drop without using any avoidance strategies. Furthermore, users can validate if their belief about how much anxiety they will feel was correct.

F. Cross platform development

A common setting for the evaluation of mobile applications in the field of e(Mental)Health is to conduct a study and hand out mobile devices with the pre-installed application to the participants. Examples for this approach are the work of Morris et al. [7] and Cafazzo et al. [30]. On the contrary, we want to reach people who are experienced with smartphones and thus most likely already own such a device. We argue that giving them a second phone is likely to harm the adoption of GET.ON PAPP because carrying around a second smartphone for a sole application is not very convenient. As a result, the ecological validity of the evaluation could be reduced. Consequently, we decided that people will use their own smartphones to execute GET.ON PAPP. To reach a larger population, the application has been developed for the two currently most widespread mobile operating systems, Google Android and Apple iOS.

Developing a mobile application for multiple platforms is challenging because each vendor has its own, very specific, development kit and environment [31]. As a consequence, development expertise for each platform is needed and implementation time is effectively doubled. Recently, practitioners as well as researchers have advocated the use of web technology as an alternative for the development of cross-platform mobile applications [31, 32]. Frameworks such as PhoneGap[9] carry this method even further by providing a runtime environment in which a web application is hosted inside a native app and thus pave the way for an almost native user experience.

Despite known performance issues [33], we employed the PhoneGap framework for the development of GET.ON PAPP. The use of web technologies not only enabled almost trouble free deployment on both platforms but also supported the rapid prototyping approach we used for development. That said, we experienced indeed a number of performance-related problems, for example when displaying photos that have been taken with the smartphone camera or transferring data to the server. However, we were able to solve these problems by implementing native plugins for each platform. The PhoneGap framework offers a dedicated API for this purpose and we consider the combination of a shared, web technology-based codebase with a few, specialized native plugins as ideal.

A major problem was, however, not related to performance but to the user experience. Creating a true mobile feeling with web technologies can be cumbersome, especially when it comes to gesture detection. For example, the swipe detection mechanism implemented in frameworks such as jQuery Mobile [10] turned out not be very robust. As a result, swipes where not properly detected on some devices or versions of the operating system. Another example is a delay between the user tapping on the screen and the "click"-event being fired, which resulted in a rather sluggish interface feeling. To work around these problems, we had to implement our own event handling system. Nevertheless, we would still recommend the web-based approach if the performance issues reported by Corall, Sillitti and Succi [33] are taken into account.

[9]http://phonegap.com
[10]http://jquerymobile.com

G. System structure

Figure 4 gives an informal overview over the system structure of the GET.ON Panik program. As mentioned above, the program is structured into an IBI and GET.ON PAPP, the mobile application which is presented in this paper. During the treatment, a client is guided by a coach who gives feedback and offers help with problems. For the feedback, the coach uses data from both, the IBI and the mobile application.

The IBI is developed and hosted inside a commercial content-management system that is provided by Minddistrict. Apart from serving the intervention content, it also offers facilities for secure conversations between coach and client.

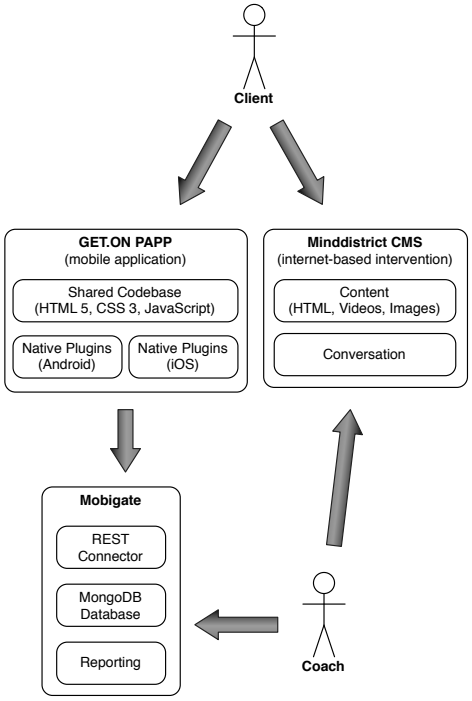


Fig. 4. Structure of the GET.ON Panik program

The mobile application periodically uploads the data entered by clients to the Mobigate backend through a REST interface which is implemented in Python using the flask micro framework[11]. The Data is stored using MongoDB, a document-oriented database from which it is accessed by a component that generates reports for the coach.

We intentionally kept the system structure simple, especially with respect to the integration of the mobile system with the CMS. In the context of this project, it was not feasible to implement an interface between the two systems for both, organizational and security reasons. However, for the system to be used in practice, closer integration would be necessary.

IV. Perspective evaluation

We currently undertake a feasibility and acceptance study (n=10) of the GET.ON Panik program that covers the IBI as well as GET.ON PAPP. Apart from getting first insights into the clinical effectiveness, at the end of the study, we will carry out a semi-structured interview with participants about their usage of GET.ON PAPP. Our goal is to investigate how the mobile application is accepted as well as potential technical problems and opportunities for improving the application. While the study did start only recently, first comments from participants on GET.ON PAPP have been positive. However, at the time of this writing, we cannot make any substantial statement about how the application is received.

After incorporating feedback from the feasibility study, a randomized controlled trial (RCT) will be conducted (n=90). The primary outcome of this RCT is the clinical effectiveness of the GET.ON Panik program. However, in the course of the trial, we plan to perform an in-depth evaluation of the usability of GET.ON PAPP. To this end, the System Usability Scale [34] will be used as well as usability data collected from the mobile application [35]. In addition, we strive to carry out a systematic investigation of the acceptance of GET.ON PAPP using the Technology Acceptance Model in a longitudinal study [36].

Finally, we plan to conduct a formal usability test for the Likert-like scale input component, in particular, to evaluate its efficiency and engagement potential. Therefore, we will carry out an experiment to compare our input component with more traditional approaches. In addition, we want to know if and how the use of colors influences a user's input.

V. Conclusion

The ubiquitous nature and extended capabilities of contemporary smartphones have made them attractive for researches from the field of E-Mental Health, not only to measure health behavior but also to integrate interventions into the daily life. Furthermore, the use of mobile technology can make E-Mental Health more attractive for end-users, thereby fostering its dissemination and implementation in practice.

To combine the potential of mobile applications with the advantages of traditional internet-based interventions, we developed the GET.ON Panik program which integrates an IBI with a mobile application. The program can be used by people with panic disorder with and without agoraphobia.

In this paper, we presented the mobile application of the GET.ON Panik program, GET.ON PAPP. The application was developed in an iterative process, incorporating experts from psychology, computer scientists and visual designers. In order to target Android and iOS systems, a cross-platform approach based on web-technology was used. We outlined how the application was designed to integrate the treatment into daily life while at the same time striving to engage and motivate users.

Future research will show if GET.ON PAPP is accepted by its users and if the approach is as beneficial for the treatment as we believe. To this end, a feasibility study has been started and a randomized controlled trial will follow. We hope that systematic studies on usability and technology acceptance will contribute to our understanding on the potential of mobile technology in E-Mental Health.

[11] http://flask.pocoo.org

REFERENCES

[1] H. Riper, G. Andersson, H. Christensen, P. Cuijpers, A. Lange, and G. Eysenbach, "Theme issue on e-mental health: a growing field in internet research." *Journal of medical Internet research*, vol. 12, no. 5, p. e74, Jan. 2010.

[2] J. Proudfoot, B. Klein, A. Barak, P. Carlbring, P. Cuijpers, A. Lange, L. Ritterband, and G. Andersson, "Establishing Guidelines for Executing and Reporting Internet Intervention Research," *Cognitive Behaviour Therapy*, vol. 40, no. 2, pp. 82–97, Jun. 2011.

[3] K. E. Heron and J. M. Smyth, "Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments." *British journal of health psychology*, vol. 15, no. Pt 1, pp. 1–39, Feb. 2010.

[4] V. Harrison, J. Proudfoot, P. P. Wee, G. Parker, D. H. Pavlovic, and V. Manicavasagar, "Mobile mental health: review of the emerging field and proof of concept study." *Journal of mental health (Abingdon, England)*, vol. 20, no. 6, pp. 509–24, Dec. 2011.

[5] J. E. Bardram, M. Frost, K. Szántó, and G. Marcu, "The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients," in *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*. New York, New York, USA: ACM Press, Jan. 2012, p. 21.

[6] J. E. Bardram, M. Frost, K. Szántó, M. Faurholt-Jepsen, M. Vinberg, and L. V. Kessing, "Designing mobile health technology for bipolar disorder: a field trial of the monarca system," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 2627–2636.

[7] M. E. Morris, Q. Kathawala, T. K. Leen, E. E. Gorenstein, F. Guilak, M. Labhard, and W. Deleeuw, "Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness," *J Med Internet Res*, vol. 2, no. 12, 2010.

[8] M. N. Burns, M. Begale, J. Duffecy, D. Gergle, C. J. Karr, E. Giangrande, and D. C. Mohr, "Harnessing context sensing to develop a mobile intervention for depression." *Journal of medical Internet research*, vol. 13, no. 3, p. e55, Jan. 2011.

[9] A. Grünerbl, P. Oleksy, G. Bahle, C. Haring, J. Weppner, and P. Lukowicz, "Towards smart phone based monitoring of bipolar disorder," in *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare - mHealthSys '12*. New York, New York, USA: ACM Press, Nov. 2012, p. 1.

[10] C. A. Depp, B. Mausbach, E. Granholm, V. Cardenas, D. Ben-Zeev, T. L. Patterson, B. D. Lebowitz, and D. V. Jeste, "Mobile interventions for severe mental illness: design and preliminary data from three approaches." *The Journal of nervous and mental disease*, vol. 198, no. 10, pp. 715–21, Oct. 2010.

[11] G. L. Klerman, M. M. Weissman, R. Ouellette, J. Johnson, and S. Greenwald, "Panic attacks in the community. Social morbidity and health care utilization." *JAMA : the journal of the American Medical Association*, vol. 265, no. 6, pp. 742–6, Feb. 1991.

[12] R. D. Goodwin, C. Faravelli, S. Rosi, F. Cosci, E. Truglia, R. de Graaf, and H. U. Wittchen, "The epidemiology of panic disorder and agoraphobia in Europe." *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, vol. 15, no. 4, pp. 435–43, Aug. 2005.

[13] J. Olesen, A. Gustavsson, M. Svensson, H.-U. Wittchen, and B. Jönsson, "The economic cost of brain disorders in Europe." *European journal of neurology : the official journal of the European Federation of Neurological Societies*, vol. 19, no. 1, pp. 155–62, Jan. 2012.

[14] S. D. Hollon, M. O. Stewart, and D. Strunk, "Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety." *Annual review of psychology*, vol. 57, pp. 285–315, Jan. 2006.

[15] J. A. Cartreine, D. K. Ahern, and S. E. Locke, "A Roadmap to Computer-Based Psychotherapy in the United States," *Harvard Review of Psychiatry*, vol. 18, no. 2, pp. 80–95, Mar. 2010.

[16] A. T. Beck, A. J. Rush, B. F. Shaw, and G. Emery, *Cognitive Therapy of Depression*. New York: Guildford, 1979.

[17] L. Festinger, "A theory of social comparison process," *Human Relations*, vol. 7, pp. 117–140, 1954.

[18] V. Spek, P. Cuijpers, I. Nyklícek, H. Riper, J. Keyzer, and V. Pop, "Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis." *Psychological medicine*, vol. 37, no. 3, pp. 319–28, Mar. 2007.

[19] P. Cuijpers, I. M. Marks, A. van Straten, K. Cavanagh, L. Gega, and G. Andersson, "Computer-aided psychotherapy for anxiety disorders: a meta-analytic review." *Cognitive behaviour therapy*, vol. 38, no. 2, pp. 66–82, Jun. 2009.

[20] P. Bao, J. Pierce, S. Whittaker, and S. Zhai, "Smart phone use by non-mobile business users," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '11. New York, NY, USA: ACM, 2011, pp. 445–454.

[21] E. Ozdalga, A. Ozdalga, and N. Ahuja, "The smartphone in medicine: a review of current and potential use among physicians and students." *Journal of medical Internet research*, vol. 14, no. 5, p. e128, Jan. 2012.

[22] L. M. Ritterband, F. P. Thorndike, D. J. Cox, B. P. Kovatchev, and L. A. Gonder-Frederick, "A behavior change model for internet interventions." *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, vol. 38, no. 1, pp. 18–27, Aug. 2009.

[23] W. Brouwer, A. Oenema, R. Crutzen, J. de Nooijer, N. de Vries, and J. Brug, "What makes people decide to visit and use an internet-delivered behavior-change intervention?: A qualitative study among adults," *Health Education*, vol. 109, no. 6, pp. 460–473, Oct. 2009.

[24] B. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco: Morgan Kaufmann Publishers, 2003.

[25] B. J. Fogg, G. Cuellar, and D. Danielson, "Motivating, Influencing, and Persuading Users," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, A. Sears and J. A. Jacko, Eds. New York: Taylor & Francis, 2009, pp. 133–46.

[26] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: capturing life as it is lived." *Annual review of psychology*, vol. 54, pp. 579–616, Jan. 2003.

[27] Z. Pousman, J. Stasko, and M. Mateas, "Casual information visualization: depictions of data in everyday life." *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, pp. 1145–52, 2007.

[28] S. Ljungblad, T. Skog, and L. Holmquist, "From Usable to Enjoyable Information Displays," in *Funology*, ser. Human-Computer Interaction Series, M. Blythe, K. Overbeeke, A. Monk, and P. Wright, Eds. Kluwer Academic Publishers, 2004, vol. 3, pp. 213–221.

[29] T. Skog, S. Ljungblad, and L. Holmquist, "Between aesthetics and utility: designing ambient information visualizations," in *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*. IEEE, 2003, pp. 233–240.

[30] J. A. Cafazzo, M. Casselman, N. Hamming, D. K. Katzman, and M. R. Palmert, "Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study." *Journal of medical Internet research*, vol. 14, no. 3, p. e70, Jan. 2012.

[31] A. Charland and B. Leroux, "Mobile application development: web vs. native," *Commun. ACM*, vol. 54, no. 5, pp. 49–53, 2011.

[32] H. Heitkötter, S. Hanschke, and T. A. Majchrzak, "Evaluating Cross-Platform Development Approaches for Mobile Applications," in *Web Information Systems and Technologies, 8th International Conference, WEBIST 2012*. Porto, Portugal: Springer, 2012, pp. 120–138.

[33] L. Corral, A. Sillitti, and G. Succi, "Mobile Multiplatform Development: An Experiment for Performance Analysis," *Procedia Computer Science*, vol. 10, no. 0, pp. 736–743, 2012.

[34] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. London, UK: Taylor & Francis, 1996, pp. 189–194.

[35] X. Ma, B. Yan, G. Chen, C. Zhang, K. Huang, J. Drury, and L. Wang, "Design and Implementation of a Toolkit for Usability Testing of Mobile Apps," *Mobile Networks and Applications*, vol. 18, no. 1, pp. 81–97, Nov. 2012.

[36] V. Venkatesh and F. D. Davis, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science*, vol. 46, no. 2, pp. 186–204, Feb. 2000.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1327–1330

Vertoid: Exploring the Persuasive Potential of Location-aware Mobile Cues

Paweł Woźniak
t2i Interaction Lab
Chalmers University of Technology
Gothenburg, Sweden
Email: pawelw@chalmers.se

Andrzej Romanowski
Institute of Applied Computer Science
Lodz University of Technology
Lodz, Poland
Email: androm@kis.p.lodz.pl

***Abstract*—This paper presents the design, implementation and user study of *Vertoid* — a mobile system for providing context-aware cues that help users limit domestic greenhouse-gas emissions. We have designed an Android-based mobile application that provides user with tips on simple eco-friendly actions in relevant locations. We then conducted a medium-term field study to evaluate the system. Our study shows that while context-aware cues have the potential to be a useful way to deliver customised content, they may as well provide unnecessary distractions. Based on the results of our study, we discuss how location awareness can be used to support persuasive systems and outline several design considerations for providing context-aware cues.**

I. Introduction

MANY researchers share the belief that technology can be effectively used as means to persuade users to change their habits and alter current behaviours [1]. Emerging technologies will try to persuade us to enjoy a healthier lifestyle and buy products from preferred suppliers. In our work, we aimed to investigate if this kind of potential can be used to support a more social cause where direct impact is harder to observe. Inspired by many social campaigns by those battling for reducing greenhouse gas emissions, we investigated that problem in detail. A unique feature of this particular case is the fact that only a large collective of users over a long period of time may make an impact. We have decided to explore possible answers to that challenge by designing a mobile system that provides information on how to limit greenhouse gas emissions only in locations where users can affect the emission levels.

In the remainder of this paper, we discuss related research and describe *Vertoid*'s design process. We then provide details on how the system was implemented and evaluated in a user study. The main contributions of our work include: the design and implementation of a persuasive mobile application for limiting greenhouse gas emissions, an exploratory user study of the application and a set of design notes that can be used with future systems.

II. Related work

Several past projects have explored the use of mobile devices in persuasive systems. Consolvo *et al.* [2] pointed to the unexplored potential of mobile phone displays for persuasive use. They investigated a system called *UbiFit* that facilitated physical activity self-monitoring. Their results show that a personal mobile display can increase awareness and support positive lifestyle changes. Contrary to *UbiFit*, in *Vertoid* we aim to avoid introducing additional devices to be carried, but try to cater to users already carrying a smart phone on a daily basis. In a similar vein, Gasser *et al.* [3] compared mobile and web-based activity and nutrition monitoring systems. While the study found no significant differences in the effectiveness of the systems, a significant advantage of an enhanced user experience was observed in favour of the mobile application. We aim to build on that potential and explore the mobile setting as a change enabler.

A critical design perspective on designing for self-reflection on the go was presented in *Fit4Life* by Purpura *et al.* [4]. Similarly to *Vertoid*, this project aims at promoting social change through small, individual steps. *Fit4Life* aims at supporting individual healthy behaviours to stop the spreading of obesity in the population. Similarly, *Vertoid* employs a system of persuasive messages to stimulate the user and provide guidelines on proper behaviours. As *Fit4Life* is mostly a conceptual prototype it uses a wide range of sensing technologies to provide context-aware cues. As *Vertoid* uses everyday smart phones, its sensing capabilities are limited, but it follows a similar design pattern. However, it is significantly different from *Fit4Life* as it addresses a problem where positive actions are harder to observe. An analysis of the literature available shows that most persuasive mobile applications have focused on improving health and fitness. Several other examples worth noting include: Exergaming [5] where a mobile app encouraged youngsters to exercise; MONARCA [6] which was used for helping patients with bipolar disorder; and CAMMInA [7] which promoted physcial activity among elders.

Gabrielli *et al.* [8] conducted design studies for a mobile application aimed at supporting sustainable transportation solutions. Their investigation included users that are not motivated by environmantal factors. *Vertoid* is aimed at users that are aware of the global warming problem and attempts to improve overall attitudes and promote principles rather than focusing on specific point-to-point transport tasks.

Creating user engagement is also an emerging theme in recent research. As Rogers [9] suggests, sometimes designers should sacrifice seamless usage to create excitement and cause

978-83-60810-53-8/$25.00

change. in *Vertoid*, we try to build on that notion and determine if messages provided in between everyday actions can be an effective persuading factor and, potentially, produce a long-term effect. This is also in line with an emerging trend for promoting engagement through mass-scale easy prototyping with high participation of the user base [10].

III. Design

We have endeavoured to design a system that would help the users make environmentally-conscious decisions when performing everyday tasks. We have decided to utilise the smart phone as the tool to achieve our goal as many users already carry these devices at all times. Recent analyses have proven that smart phones have the potential to create new habits consisting of brief usage sessions [11]. In principle, our goal was to capitalise on that potential by establishing a set of triggering contexts (in our case driven mainly by location) that would promote the desired behaviours. We have conducted semi-structured interviews with potential users and constructed severla user journeys and personas to aid the design.

To engage the user in environmental activities, Vertoid uses a set of environmental challenges that suggest possible actions that might help reduce greenhouse-gas emissions. Location tagging assures the challenges are posted at appropriate locations. The tasks may range from simple ones, like replacing a light bulb, to more complicated, like replacing the furnace filter. Whenever possible, a specific amount of carbon dioxide that can be saved by completing the challenge is indicated. Some challenges concern changing habits and others are designed to incline the user to make a single contribution. Each task is associated with a specific location type. Our aim was to increase the chances for an immediate completion of the tasks by asking the user to manipulate objects in their direct physical proximity.

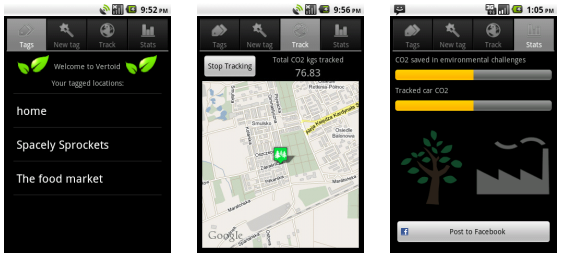
The application keeps track of the declared contributions and can always provide an overview of the current progress of the user. The users can also use the phone to track the carbon dioxide produced during car travel by activating *Vertoid* while driving. The personal statistics are also easy to share on social networks.

IV. Implementation

In order to evaluate the *Vertoid* concept, a high-fidelity prototype was implemented on the Android mobile platform. Here, we present the key features of *Vertoid*.

The tag list (Figure 1(a)) — Technical limitations require the users to declare their home and work locations as well as public places they frequent. This is needed only once. The main purpose of the list is to enable the user to verify which locations have already been tagged. There is a possibility to delete a given tag by using the context menu.

The map view (Figure 1(b)) — this screen enables recording carbon dioxide emissions while driving. The current position of the user is displayed on a map and tracking can be toggled on and off.



(a) The tag list screen (b) The map view during car tracking (c) The statistics screen

Fig. 1. Screen shots showing the most used screens in *Vertoid*

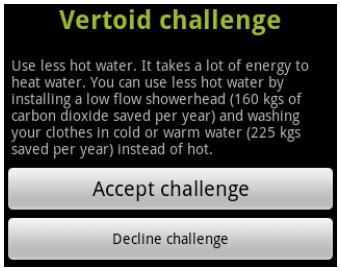


Fig. 2. The challenge prompt. *Vertoid* triggers environmental challenges based on user location

The statistics screen (Figure 1(c)) contains a visualisation of all the data accumulated during the usage of *Vertoid*. The user's status as to saving and generating carbon dioxide is presented as progress bars and waning pictures of a factory and a tree. A Facebook share button is provided.

Environmental challenges are *Vertoid*'s core feature. Whenever a user enters a location defined previously, a *Vertoid* notification appears in the Android notification panel. When reviewed, the notification scales up to a challenge prompt (an example is provided in Figure 2). The user can accept or reject a given challenge. The appearance of a new challenge is signalled by a vibration and challenges appear only once per visit to a location.

V. Evaluation

A. Methodology

We have conducted a limited-scope field study with 16 (9 male and 7 female) participants who were asked to try using the application in their everyday lives. Study subjects were recruited on a voluntary basis and consisted mainly of students. We looked for regular smart phone users already familiar with the Android operating system who expressed interest in actions against global warming. A detailed demographic profile of the participants is presented in Table I. The first user evaluation activity was conducting semi-structured interviews with all the participants to ensure their prior familiarity with smart phones and the intent to pursue environmental activities to some extent. The inteviews included a section where we discussed everyday actions and their impact on global warming. While the participants were quite enthusiastic about helping the

TABLE I
BASIC INFORMATION ON THE PARTICIPANTS AND THE STUDY. WE ASKED THE PARTICIPANTS TO SUBJECTIVELY RATE THEIR SMART PHONE USAGE FROM 1 — VERY SELDOM TO 5 — MORE THAN 10 TASKS PER DAY

Property	max	min	$mean$
Age [years]	33	19	24.73
Usage duration [weeks]	8	4	5.27
Smart phone usage [see caption]	5	3	4.10

common cause of reducing greenhouse gas emissions, their awareness to what one can do to help was rather limited. This created an opportunity for exploring *Vertoid*. We also recorded basic demographic data on the participants. We have then installed the application on the users' personal Android phones and provided a brief overview of the functionality with a practical walk through. Our initial concern was that the usability of the program might have affected its persuasive potential. We used expert evaluation from the start of the development, but a simple user test was required. Consequently, each user was asked to complete a simple questionnaire based on Nielsen's usability heuristics [12] after two weeks of use. We asked the users to rank the qualities of the application on a 1 to 5 scale. Short descriptions of the qualities were provided. We concluded that the usability of the application did not affect its possible impact as the users rated Vertoid (on average) 4.10 for clarity, 3.82 for appearance and 4.28 for error handling. We believe that these ratings are satisfactory for a prototype application. The participants had the opportunity to use *Vertoid* on their phones for periods from 4 to 8 weeks (the varying period lengths are a result of limiting the intrusiveness of the study by avoiding scheduling conflicts).

After the unsupervised usage period, the participants were invited for an interview once again. As our investigation was largely exploratory, we looked mainly for qualitative feedback. However, we gathered input on the frequency of use. In our preliminary talks, we have determined that some of the participants were reluctant to share their logged data as it contained their frequented location. This is the motivation behind relying on reported usage accounts rather than logging.

In order to evaluate the environmental awareness change caused by Vertoid, we included questions about green living in the interviews both before and after using the application. In the initial talk, we included both straightforward questions (*e.g.* "Do you try to maintain a sustainable lifestyle?") and indirect measures of green living ("How many incandescent light bulbs are there in your house?"). We asked similar indirect questions in the post-usage interview.

B. Results

Overall usage frequency varied significantly among participants. On average participants reported using *Vertoid* $\mu = 4.18$ times in a week with $SD = 2.36$. Such a discrepancy prompted us to investigate why some of the users refused to use *Vertoid* regularly. In Figure 3 we show usage frequency distribution among the participants.

Those of the participants who reported using the application

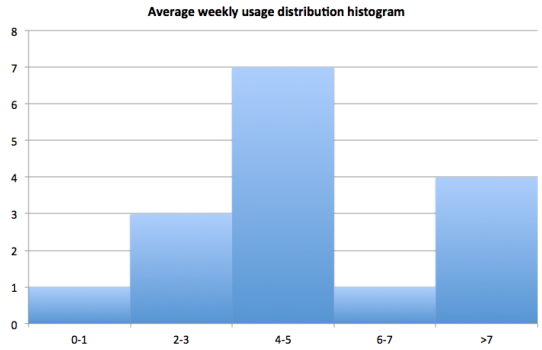


Fig. 3. The histogram shows participant counts and usage frequency intervals in times used per week for the field study.

only sporadically described the experience as interesting, but lacking the impact potential. *"A few lines of text cannot convince me to do additional housework."* remarked one participant. Two participants did not tag their frequented locations citing privacy issues, so they only received notifications in a predefined set of locations shipped with the installation. Even though 7 (33%) of the users explicitly stated that they did not use the application more than once a week, they still claimed that they wanted to improve their daily routines to make their lives more sustainable. On the other hand, those who reported using *Vertoid* regularly noted that the number of challenges was quite high (*"I never realised you could have so many tips on green living"*). All of the users who read at least one of the challenges (95%) reported learning new facts through the challenges. Those who used location tagging reported that it worked properly and the triggered challenges were relevant to the space. Accounts of the shopping experience with *Vertoid* provided us with additional insights. It surfaced that the coarse positioning was inadequate in the shopping areas while it worked at home and at work. For example, users were annoyed to receive suggestions about buying local produce once they have passed the vegetable section, but they did not mind being reminded to change to fluorescent light bulbs while not within reach of a lamp. Generally, we can conclude that the cues were quite successful when they concerned short, immediate actions performed with objects close to the user, *e.g.* choosing nationally locally meat.

We have observed a difference between declared environmental engagement and real actions taken everyday. Users often overestimated their efforts in living a green lifestyle despite being unaware of the simple actions that help. In the post-usage interview, we noted a slight increase in the knowledge of easy ways to reduce greenhouse gas consumptions. Figure 4 illustrates the subjective and objective pre-study awareness level as measured by our questionnaires and the results of a similar objective test after using *Vertoid*. We believe that the observed increase in awareness may suggest that mobile contextualised cues have cause a long-term effect.

VI. DISCUSSION

Overall, we believe that our exploratory field study confirmed the persuasive potential of location-aware mobile cues

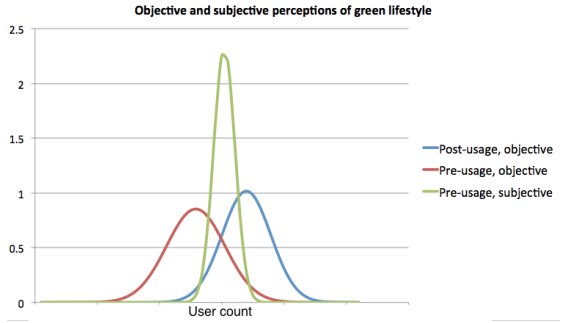


Fig. 4. User environmental awareness levels as measured by our questionnaire. The plot contains values declared by the participants and pre- and post-study results obtained with our questionnaire.

for stimulating environmental awareness. Those of the participants who decided to use *Vertoid* regularly reported positive experiences and were inclined to put some of the suggestions into practice. This was not the case for those using the application only sporadically. We can attribute this difference to the fact that *Vertoid* may constitute a significant intrusion to everyday life that must be consciously accepted by the user. As a consequence, we may speculate that context-aware mobile cues may affect everyday acrtion to a degree that requires conscious acceptance of the intrusion prior to delivering the cues, *i.e.* they are effective for users who want to change, but seek the means to achieve the goal. In a way, the field study differentiated between those who spoke of environmental matters as this was "a good thing to say" and those who were ready to take action.

We have consciously employed a user-driven design approach in creating *Vertoid* in order to embrace the practicality of everyday lives and focus on user accounts. This is supported by the fact that most known theory-driven methods revolve around health and well-being (discussed to a large extent in [13]). Using a design-oriented approach affects our insights as to the long-term effect of the application. The long-term experience of using *Vertoid* is a complex research question due to the mobile field setting and we cannot use theoretical means for modelling the answer. We have completed an exploratory study that clearly confirms the potential of mobile cues to become an effective behavioural change technology. Qualitative feedback from the participants seems to confirm that some of them learnt new ways to lead a more environmentally-friendly life and integrate some of the suggestions into everyday habits. Thus, we see an emerging need for studying this research context further to clearly determine if providing contextualised suggestion can lead to real changes in patterns of daily life. Long-term field studies are required, but these will only be possible if sensing techniques that will enable more accurate context recognition are made available. Recent advances in embedding information in physical objects like near field communication can be explored for opportunities to provide more relevant suggestions.

VII. Conclusions

In this paper we have presented the design, implementation and an exploratory field study of a mobile application that investigates the potential of contextualised mobile cues for persuading users to lead a more environmentally-friendly lifestyle. We have implemented a high-fidelity prototype and ran a field study that resulted in a better understanding of the persuasive effect of the application. *Vertoid* received positive feedback from most of the 16 study participants. Usage intensity varied significantly, but those participants who reported using the application on a everyday basis concluded that it was a useful addition to their everyday routines and created new opportunities for a greener lifestyle. We have observed that the location-based challenges functioned satisfactorily at home and at work. We see new potential emerging from increased accuracy at shopping places for an enhanced experience. We hope to explore the potential of context-awareness for persuasion further by employing more advanced sensing methods and long-term studies.

References

[1] B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*. Science & Technology Books, 1 ed., 2002.

[2] S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, and J. A. Landay, "Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays," UbiComp '08, (New York, NY, USA), pp. 54–63, ACM, 2008.

[3] R. Gasser, D. Brodbeck, M. Degen, J. Luthiger, R. Wyss, and S. Reichlin, "Persuasiveness of a mobile lifestyle coaching application using social facilitation," in *Persuasive Technology* (W. IJsselsteijn, Y. Kort, C. Midden, B. Eggen, and E. Hoven, eds.), vol. 3962 of *Lecture Notes in Computer Science*, pp. 27–38, Springer Berlin Heidelberg, 2006.

[4] S. Purpura, V. Schwanda, K. Williams, W. Stubler, and P. Sengers, "Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight," CHI '11, (New York, NY, USA), pp. 423–432, ACM, 2011.

[5] C. Wylie and P. Coulton, "Mobile persuasive exergaming," in *Games Innovations Conference, 2009. ICE-GIC 2009. International IEEE Consumer Electronics Society's*, pp. 126–130, 2009.

[6] G. Marcu, J. Bardram, and S. Gabrielli, "A framework for overcoming challenges in designing persuasive monitoring and feedback systems for mental illness," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pp. 1–8, 2011.

[7] M. Rodriguez, J. Roa, A. Moran, and S. Nava-Munoz, "Persuasive strategies for motivating elders to exercise," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, pp. 219–223, 2012.

[8] S. Gabrielli, R. Maimone, P. Forbes, J. Masthoff, S. Wells, L. Primerano, L. Haverinen, G. Bo, and M. Pompa, "Designing motivational features for sustainable urban mobility," CHI EA '13, (New York, NY, USA), pp. 1461–1466, ACM, 2013.

[9] Y. Rogers, "Moving on from Weiser's vision of calm computing: engaging ubicomp experiences," UbiComp'06, (Berlin, Heidelberg), pp. 404–421, Springer-Verlag, 2006.

[10] P. Wozniak and A. Romanowski, "Everyday problems vs. ubicomp: a case study," WIMS '12, (New York, NY, USA), pp. 57:1–57:4, ACM, 2012.

[11] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, "Habits make smartphone use more pervasive," *Personal Ubiquitous Comput.*, vol. 16, pp. 105–114, Jan. 2012.

[12] J. Nielsen, "Usability Heuristics," in *Usability Engineering*, ch. 5, London: Academic Press, 1993.

[13] R. I. Arriaga, A. D. Miller, E. D. Mynatt, C. Pagliari, and E. Shehan Poole, "Theory vs. design-driven approaches for behavior change research," CHI EA '13, (New York, NY, USA), pp. 2455–2458, ACM, 2013.

Software Systems Development & Applications

SSD&A is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the discipline of software engineering. The SSD&A area emphasizes the issues relevant to developing and maintaining software systems that behave reliably, efficiently and effectively. This area investigates both established traditional approaches and modern emerging approaches to large software production and evolution. Events that constitute SSD&A are:

- ATSE'13 – 4th International Workshop Automating Test Case Design, Selection and Evaluation
- IWCPS'13 – International Workshop on Cyber-Physical Systems
- PBDA'13 – Performance of Business Database Applications
- WAPL'13 – 4th Workshop on Advances in Programming Language

4th International Workshop Automating Test Case Design, Selection and Evaluation

TRENDS such as globalisation, standardisation and shorter lifecycles place great demands on the flexibility of the software industry. In order to compete and cooperate on an international scale, a constantly decreasing time to market and an increasing level of quality are essential. Software and systems testing is at the moment the most important and mostly used quality assurance technique applied in industry. However, the complexity of software systems and hence of their development is increasing. Systems get bigger, connect large amounts of components that interact in many different ways on the Future Internet, and have constantly changing and different types of requirements (functionality, dependability, real-time, etc.). Consequently, the development of cost-effective and high-quality systems opens new challenges that cannot be faced only with traditional testing approaches. New techniques for systematization and automation of testing are required.

Even though many test automation tools are currently available to aid test planning and control as well as test case execution and monitoring, all these tools share a similar passive philosophy towards test case design, selection of test data and test evaluation. They leave these crucial, time-consuming and demanding activities to the human tester. This is not without reason; test case design and test evaluation are difficult to automate with the techniques available in current industrial practice. The domain of possible inputs (potential test cases), even for a trivial program, is typically too large to be exhaustively explored. Consequently, one of the major challenges associated with test case design is the selection of test cases that are effective at finding flaws without requiring an excessive number of tests to be carried out. This is the problem which this workshop wants to attack.

This workshop will provide researchers and practitioners a forum for exchanging ideas, experiences, understanding of the problems, visions for the future, and promising solutions to the problems in automated test case generation, selection and evaluation. The workshop will also provide a platform for researchers and developers of testing tools to work together to identify the problems in the theory and practice of software test automation and to set an agenda and lay the foundation for future development.

Topics

Topics include (but are not limited to):

- techniques and tools for automating test case design:
  - model-based,
  - combinatorial.based,
  - optimization-based,
  - etc.
- Evaluation of testing techniques and tools on real systems, not only toy problems.
- Benchmarks for evaluating software testing techniques

Event Chairs

Eldh, Sigrid, Ericsson & Karlstad University, Sweden
Prasetya, Wishnu, University of Utrecht, Netherlands
Vos, Tanja, Universidad Politecnica de Valencia, Spain

Program Committee

Bagnato, Alessandra, Softeam, France
Bauersfeld, Sebastian, Universidad Politecnica de Valencia, Spain
Condori, Nelly, Universidad Politecnica de Valencia, Spain
Escalona, Maria Jose, Universidad de Sevilla, Spain
Lakhotia, Kiran, University College London, United Kingdom
Marchetto Alessandro, Centro Ricerche Fiat - CRF, Italy
Marin, Beatriz, Universidad Diego Portales, Chile
Memon, Atif, University of Maryland, United States
Polo, Macario, Uiversidad de Castilla la Macha, Spain
Shehory, Onn, IBM, Israel
Tonella, Paolo, Fondazione Bruno Kessler,, Italy
Tuya, Javier, Universidad de Oviedo, Spain

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1335–1342

Requirements on automatically generated random test cases

Thomas Arts
Quviq AB

Alex Gerdes
Quviq AB

Magnus Kronqvist
Ericsson AB

***Abstract*—Developing, for example, a simple booking web service with modern tools can be a matter of a few weeks work. Testing such a system should not need to take more time than that. Automatically generating tests from specified properties of the system using the tool QuickCheck provides professional developers with the required test efficiency. But how good is the quality of these automatically generated tests? Do they cover the cases that one would have written in manual tests? The quality depends on the specified properties and data generators and so far there has not been an objective way to evaluate the quality of these QuickCheck generators. In this paper we present a method to assess the quality of QuickCheck test data generators by formulating requirements on them. Using this method we can give feedback to developers of such data generators in an early stage. The method supports developers in improving data generators, which may lead to an increase of the effectiveness in testing while maintaining the same efficiency.**

I. Introduction

THIS paper provides a solution to a problem originating from the use of property-based testing of a simple, but realistic web service developed and used by a telecommunication company. Property-based testing [1] is a technique with which one describes properties of a software system using QuickCheck. QuickCheck has many implementations, for example [2], [3]. The general methodology is that one writes properties of the software under test, from which QuickCheck automatically generates test cases to validate these specified properties.

It has been shown that property-based testing increases efficiency and effectiveness of software testing [4]. Prowess [5], a recent EU STREP project, addresses the challenge to reduce time spent on testing, whilst increasing software quality, in order to quickly launch new, or enhancements of existing, web services and internet applications. In this paper we do not evaluate property-based testing, but focus on one particular challenge in using this technique. The use of property-based testing requires the definition of data generators that control QuickCheck's random data generation. There are many ways to define such data generators, and it requires some skills and experience to define a data generator with good data distribution. We explore how we can help developers to measure the quality of their data generators.

A danger in using QuickCheck is that we no longer see the generated test data. In fact, we would not want to see it, because QuickCheck can generate many test cases. As a result, we may be tricked into a false sense of security by a large number of passing tests, but fail to notice that the distribution is badly skewed. Even if we observe it by using QuickCheck's possibility to collect statistics on the test data, we would need an expert to judge whether the provided data is good test data.

We address the manual interaction of judging test data by capturing the expert knowledge in formal requirements. These requirements are used to automatically assess the quality of the test data generators. In this way, the generators can be developed with limited involvement of experts.

A clear example of the problem of judging the quality of data generators came to our attention when testing a web service created by a telecommunication company. Telecommunication systems often use special purpose hardware, which is rather expensive to build. This raises a cost issue for testing such systems; unlike commodity PCs, one cannot simply put as many machines in a test lab as one would like to. Hardware becomes a resource and more efficient use of this resource lowers the total production cost. When sharing resources, one needs a booking system. In our case, the interoperability requirements were to fit the already in-house built continuous integration and other test and deployment tools. Based on these specific requirements and experience with purchasing this kind of heavily integrated software in the past, this booking system was decided to be build in-house by spending a few weeks of effort. This resulted in a simple web service used by several sites in the world described in more detail in Sect. III.

Unit level testing of this system was performed following the existing literature [6], [7] and revealed that it is hard to judge the quality of the generated test cases. One can be mislead in believing that the system is well tested, although the randomly generated test cases do not cover the interesting test cases. We identified the need for assessing the quality of the generated test cases. In Sect. V we describe how we can express requirements on QuickCheck generators that we use to assess the quality of the generated test data.

After successfully using our method in this proprietary first application, we have evaluated the method in a different context. We have used the method to asses the data generators that we use to test our second application: the open source scheduling web application Dudle[8]. In Sect. VI we describe the requirements with which we validated the data generators for testing Dudle.

Although this paper describes the application of our method for two particular applications, the techniques we describe for

Partially supported by the EU STREP project Prowess, grant 317820

978-83-60810-53-8/$25.00

formulating requirements on generated test cases are generally applicable to many applications of this kind.

II. QUICKCHECK

QuickCheck [2] is a tool that tests universally quantified *properties*, instead of single test cases. QuickCheck generates random test cases from each property, tests whether the property is true in that case, and reports test cases for which the property fails. QuickCheck also "shrinks" failing test cases automatically, by searching for similar, but smaller test cases that fail as well. The result of shrinking is a "minimal"[1] failing test case, which often makes the root cause of the problem easy to find.

The original Haskell QuickCheck has inspired a number of different versions for a wide range of programming languages such as C++ [9], Java [10], or ML [11]. The work in this paper is based on the use of QuviQ QuickCheck. QuviQ QuickCheck[2] [3] is a commercial application that includes many advanced features, such as model-based testing using a state machine model [12]. State machine models are tested using a QuickCheck library, which invokes call-backs supplied by the user to generate and test random, well-formed sequences of API calls to the software under test.

An example of a QuickCheck property is shown below. This property is used to test a timeline datatype. A timeline is considered an ordered list of intervals and an interval is an ordered pair of dates. The property uses data generators for an interval (`interval()`) and for a timeline (`timeline()`). The software under test provides a function `add` that, given an interval and a timeline, should add this interval to the timeline, provided the interval does not overlap with an already existing interval. After successfully adding the interval, it should be a member of the newly created timeline.

```
prop_add() ->
  ?FORALL({I, T},
          {interval(), timeline()},
     begin
       case catch add(I, T) of
           {'EXIT', {overlap,_}} ->
             is_overlap(I, T);
           NewT ->
             member(I, NewT)
       end
     end).
```

The functions `is_overlap` and `member` are provided by the software under test as well.

At a unit testing level, one could express a number of such general properties for API functions. This would already be effective in finding a number of defects, but the problem is that it is hard to know when one has provided enough properties to cover the implementation. This problem has been addressed in literature [6] for datatypes. Based on this approach, the solution for the timeline example would be: create a model implementation, a generator for a timeline in which all possible constructors are used in the generation, and one property per operation. This solution, however, does not address the problem of generating data with a good distribution.

Let us consider how to write data generators for an interval and a timeline, which are used in the property above. A simple way to construct a timeline would be to generate a random number of intervals and put those in a timeline. If an interval consists of two completely random dates, the first less or equal to the second, then the size of the interval may be huge and the possibility to get overlapping dates in a timeline increases quickly. In this case, generating a timeline becomes problematic, since it must not contain overlapping intervals. We therefore should put the generation of intervals and dates under control and steer it in the right direction. For example, if we only randomly pick the first date and then add a random (small) number of days to this date to generate an interval, then we may end up with timelines that contain far more intervals. But the possibility of negative testing, i.e., testing that overlapping intervals are rejected, decreases.

Using QuickCheck, one needs to control the randomness in the generators. The problem we address in this paper is *not* to come up with different generators that are better or worse for certain kind of testing, the problem is to know whether the generated test cases provide a good coverage of the things you want to test.

III. APPLICATION 1: BOOKING WEB SERVICE

The booking web service is a tool used internally within a testing organisation at a telecommunication company to manage and enable efficient sharing of hardware equipment. Specialised hardware in the telecommunication industry is usually associated with profound costs, and efficient sharing of those resources is fundamental to achieve a cost-effective environment. It is also becoming more commonplace to configure larger networks of nodes. Those are setups that take longer time to install and configure, and re-using them between teams saves a considerable amount of time for testers.

The web service serves two main use cases with different needs: manual use of the test equipment and automatic use of the test equipment. In the first use case, engineers want to search the labs for hardware that they need for their tests. When they find what they need, they reserve the hardware for some days or a couple of weeks. Engineers can return to the service for additional information or to extend their bookings. The second use case is a fully automatic use of the test equipment during the continuous integration process. Every time a new software package is delivered, a few different sets of suites are run, organised as short, medium, and long term, each suite defined to run on a specific network setup. Every time this activity starts, the web service will be queried for available hardware of the proposed topology. Any network that contains this topology is accepted, and that network will be configured to disconnect the unwanted hardware. All networks not containing the proposed topology will not be considered. If there are no networks available that satisfy the request, then

[1] In the sense that it cannot shrink to a failing test with the shrinking algorithm used.

[2] We use QuickCheck from here on to denote this version

continuous integration will pause for some time and retry again some time later.

The web service is built in Erlang [13], based on a Mnesia database, and a YAWS front-end [14]. At the core of the application it uses a timeline data structure to manage bookings. This data structure represents a calendar in which certain intervals are blocked (the days that equipment is booked).

Keeping an efficient regression suite is important. Implementing a small booking system is a relatively small task compared to implementing other telecommunication software. It took less than five weeks to implement the first version of the functionality. However, as the system evolved over time, as is common in industry, it had to be adapted to new and changing requirements many times over. The first implementation, for example, was used with a single lab only, having users all at the same location. The system today has evolved and users can now be found in a handful of locations around the world. The software has had to retain its integrity over several adaptations like this.

Even though these small scale tools are not business critical, the consequences for an organisation can still be severe when they malfunction. In the case of the booking web service, it would have only modest effects on the engineers since their sole use of it is to find and book new equipment. The already booked nodes will not be affected, and thus tests can be carried on as normal. However, the global usage puts stress on the availability, and small errors at any time will most likely create annoyances or delays for someone, somewhere.

The effects are more severe for the automatic testing since it is continuously dependent on up to date information of hardware availability. It was decided that it should not occupy any resources when not running because the total number of hardware that would idle between runs would have an unacceptable impact on the lab size.

The web service shows a small and limited, but practical example of a software entity ubiquitously found in industry. These kind of systems are not business critical in themselves, but their cumulative effect on the business process as a whole is. Testing them sufficiently to ensure their quality will therefore be important to keep the bigger machinery running smoothly.

IV. Testing datatypes

The booking web application consists of a number of components, of which the implementation of the timeline datatype is central. This datatype is used to store, compare and remove time intervals for bookings. Since testing datatypes with QuickCheck is well documented [6], [7], we just need to follow the methodology described: create a generator using the constructors of the data type, create a model implementation of the datatype, and write one property for each operation on the datatype.

The timeline datatype represents a calendar in which certain intervals are blocked (the days that equipment is booked). A timeline can be constructed and manipulated using the following functions:

`new`	create an empty timeline,
`add`	add an interval to a timeline,
`delete`	delete a specific interval from a timeline,
`after_`	remove all intervals before a certain date from the timeline; used to prune old bookings,
`tail`	remove the first interval of a timeline.

According to [6] we should use all these operations in the timeline generator.

In addition to these operations, we also define functions to compare timelines, extract elements from a timeline, such as an particular interval, and to check whether two intervals overlap:

`equals`	check whether two timelines are equal,
`empty`	check whether a timeline is empty, i.e., does not contain any interval,
`member`	check whether a particular interval is already in the timeline,
`overlap`	check whether a given interval overlaps with any of the intervals in the timeline,
`valid`	check whether the intervals in a timeline are chronologically ordered and do not overlap,
`get_overlap`	return the first interval in the timeline that overlaps with a given interval,
`head`	return the first interval in a timeline,
`nth`	return the n^{th} interval in a timeline,
`overlap`	check whether two given intervals overlap.

Following the method mentioned earlier, we now need to create a generator for timeline data structures, and a model of a timeline that can be constructed with corresponding operations. We can then define a QuickCheck property for each operation, which applies the operation to a timeline and the corresponding operation to a model of that timeline, and validates if the resulting timeline conforms to the resulting model. For example, we define the following property for the `add` operation:

```
prop_add() ->
  ?FORALL(
    {I, SymT}, {interval(), timeline()},
    begin
      T = eval(SymT),
      case catch add(I, T) of
        {'EXIT', {overlap, _}} ->
          is_overlap(I, model(T));
        NewT ->
          equals(model(NewT),
                 model_add(I, model(T)))
      end
    end).
```

A random interval `I` and a timeline `SymT` are generated by the generators `interval()` and `timeline()` respectively. The timeline generator generates a symbolic timeline, that is, a value generated by this generator is a list of symbolic calls to constructor operations. Symbolic values allow us to inspect how an actual value is constructed. Whenever we need the actual value, we evaluate the symbolic value using the `eval` function. We also want to perform negative tests and check if a proper error message is produced. When an exception is

raised, we validate if it is raised for the right for the right reason, in this case if the generated interval overlaps with the generated timeline. If no exception is raised, the model of the newly obtained timeline should be equal to the model of the generated timeline to which we add the interval via the corresponding model operation.

The model and properties are easy to come up with following the aforementioned method, but the tricky parts are the generators for intervals and timelines. Decimal numbers [6] and ordered sets [7] can be generated from a simple recursive generator or grammar description, since there is little dependency between values generated in different recursive calls. However, in our case we have an invariant on the generated timeline, namely that intervals should be non-overlapping. This makes the data generation severely more difficult. The first contribution of this paper is *a method to evaluate the data distribution* for datatype generators that need to meet some requirements. The second contribution is *a timeline generator that performs well* with this evaluation.

A. Interval generator

We want to test the timeline functions on random input and need data generators for the arguments of those functions. Many functions take an interval as argument. We represent an interval as a pair of two triples: year, month and day. A naive approach would be to construct an interval using two (ordered) random generated dates. Using QuickCheck one would generate such a triple with the `choose` generator and use the `?SUCHTHAT` macro to filter dates that the Erlang `calendar` module rejects as valid date.

```
ymd() ->
 {choose(2012, 2013),
  choose(1, 12),
  choose(1, 31)}.

date() ->
  ?SUCHTHAT(Date, ymd(),
            calendar:valid_date(Date)).
```

A tuple of two such dates, however, does not provide good test data. We want the test data to typically be a few days, preferably around week, and containing month and year transitions. For example, 2012-12-28 to 2013-1-1 would make for a nice test case. We should create a generator that chooses such intervals with reasonable likelihood. As noted before, an interval generator that picks the date purely randomly would create intervals that are very large. Limiting the year to be either 2012 or 2013 reduces the number of extremely large intervals, but at the same time, choosing more than 4 non-overlapping random intervals in that domain is unlikely to happen with the uniform distribution of `choose`. We therefore steer the generation to make it more likely to select intervals that we are interested in by adding a few days to the date and discard dates that therewith become invalid.

```
interval() ->
  ?LET(D1, date(),
       ?LET(D2, larger_date(D1),
            {D1, D2})).

larger_date({Y, M, D}) ->
  ?SUCHTHAT(
    frequency(
      [{9, ?LET(Days, nat(),
                shift({Y, M, D}, Days))},
       {1, Date, date()}])
    Date > {Y, M, D}).
```

After picking the first random date, the second date is constructed by adding an arbitrary number of days to the date. Alternatively, in 10 percent of the cases we also allow a completely random date as second alternative, provided it is larger than the first date.

This is one attempt to get a good distribution of intervals in a timeline. The question is, how good? And are there any obvious cases that we do not test with such a distribution or cases that are unlikely to be generated in a run of hundred tests?

V. Testing Generator Requirements

We would like to be able to assess the quality of a test data, in order to convince ourselves that a generator is good enough. To assess the quality of a test data we propose to define *requirements* on values produced by such a generator. A requirement for a generator is a property that should hold for a certain percentage of the generated tests. So, we can specify that a minimum (or maximum) number of generated test values should adhere to a given property. We have extended QuickCheck with the possibility to define such requirements on generators in a convenient way. For example, a requirement on a generator for natural numbers between 1 and 10, may be that it should generate a 1 within say 12 tests. We can express such a requirement as follows:

```
req_has_one() ->
  Gen = eqc_gen:choose(1, 10),
  ?REQ_EXISTS(1, Gen, 12).
```

The `req_has_one` function returns a QuickCheck property that we can test, just as any other 'normal' property, with the `quickcheck` function:

```
1> eqc:quickcheck(req_has_one()).
OK, passed
true
```

Not surprisingly the generator meets this requirement. In case a generator meets a requirement, QuickCheck prints an acknowledgement and returns the value `true`.

A slightly larger example is the following requirement:

```
req_half_is_larger_than_five() ->
  Gen = eqc_gen:choose(1, 10),
  ?REQ_MIN(X, Gen, X >= 5, 50.0, 100).
```

This requirement demands from the generator that at least 50% of the generated values are equal or larger than 5. Running `quickcheck` on this requirement results in the following output:

```
2> eqc:quickcheck(req_half_is_larger_than_five()).
Failed! Only 46 percent meets the condition.
```

```
[6,4,1,9,2,8,9,1,6,2,7,2,10,9,3,9,8,5,6,1,2,
 ...
 2,4,3,2,5,3,10,2,8,2,5,5,9,4,1,9,2,10,8,5,8]
false
```

QuickCheck reports that the generator `choose(1, 10)` did not meet the `req_half_is_larger_than_five()` requirement. It shows the generated test data, which can be regarded as a counterexample, and the percentage of the data that did meet the requirement. Since the `choose` generator has a linear distribution, it is possible that we generate 50 numbers that are smaller than 5. The counterexample allows us to inspect the generated data. Using this information we can improve the generator, or, if we are satisfied with the data distribution, we could weaken the requirement.

We offer the following macros to construct requirements on QuickCheck test data generators:

`?REQ_EXISTS(X, Gen, N)`,
: check if a generator `Gen` will at least generate a value equal to `X` within `N` number of tests,

`?REQ_EXISTS_FOR(X, Gen, P, N)`,
: check if a value for which predicate `P` (that takes a value as argument and returns a Boolean value) holds, is generated within `N` tests,

`?REQ_BETWEEN(X, Gen, P, Min, Max, N)`,
: check if the percentage of the values for which `P` holds lies between `Min| and \erlang|Max`,

`?REQ_MIN(X, Gen, P, Min, N)`,
: same as `?REQ_BETWEEN` but only with a lower bound,

`?REQ_MAX(X, Gen, P, Max, N)`,
: same as `?REQ_BETWEEN` but only with an upper bound,

`?REQ(X, Gen, P, C, N)`,
: the above requirement macros are expressed in terms of this is general macro, which generalises the condition check (which takes an percentage as argument and returns a Boolean value).

The last argument, which specifies the number of tests, of all macros can be left out. If the number of test is not specified we use the default of a hundred tests. We can check individual requirements with the `quickcheck` function. In addition, we provide a function, named `req_module`, which checks all requirements defined in a module. The name of a requirement needs to be prefixed with `req_`.

QuickCheck already offers the possibility to *measure* the probabilities of different kinds of test data. This can be done by instrumenting a QuickCheck property to collect statistics during testing. For example, we might instrument a property as follows, to measure how often a one is generated by the `choose(1,10)` generator:

```
prop_has_one() ->
  ?FORALL(N, eqc_gen:choose(1,10),
          collect(N == 1, N < 11)).
```

The effect of the line `collect(N == 1, ...)` is to collect the value of `N` in each test, and after testing is complete, to display the distribution of the values collected. In this case, testing the instrumented property yields:

```
3> eqc:quickcheck(prop_has_one()).
.........................................
OK, passed 100 tests

89% false
11% true
true
```

The collected statistics show that `N` was 1 in 11% of the generated tests. This is already valuable information. However, we cannot use the collected data to give a judgement, nor can we let the property succeed or fail based on these statistics. As a consequence, an expert must (re)examine the result in order to check if a generator meets its requirements. Using the requirement macros defined above, we can. Note that the requirement functionality is not meant to replace the statistics collection functionality. Both are useful in their own right.

Interval generator: Let us now return to our running example. Using the above macros we can introduce some requirements on the interval generator, which we introduced in the previous section. For example, we can state that an arbitrarily generated list of intervals should consist of non-overlapping intervals in 75% of the cases:

```
req_non_overlap() ->
  ?REQ_MIN(Is, eqc_gen:list(interval()),
           non_overlapping_pair(Is),
           75.0).
```

The `non_overlapping_pair` function checks whether or not there is an overlap between one of the elements of `Is` with any of the other elements. When we check the requirement for the generator of intervals with two arbitrary dates (first smaller than the second), we get a requirement success rate of around 30%, thus in 70% of the generated lists of intervals, the lists contains overlapping intervals. Moreover, when we only generate lists containing five intervals, we seem to be unable to create any of these without an overlapping interval. However, for the smarter generator for intervals described above, we come close to a success rate of 80%.

A. Timeline generator

The problem of bad data distribution gets even more obvious if we follow the generator construction explained in literature, where we build a data structure by using the constructors defined by the datatype.

```
timeline() ->
  ?SIZED(Size, well_defined(timeline(Size))).

timeline(0) ->
  {call, ?API, new, []};
timeline(N) ->
  ?LAZY(oneof(
    [timeline(0),
     {call, ?API, add,    [interval(),
                           timeline(N-1)]},
     {call, ?API, tail,   [timeline(N-1)]},
     {call, ?API, delete, [interval(),
                           timeline(N-1)]}
    ])).
```

This generator creates an arbitrary timeline by recursively adding and deleting intervals from a previously defined timeline. We do this symbolically, which means that we build a data structure containing the calls to the API instead of calling the API directly. But this timeline generator does a very poor job. By deleting an arbitrary interval from the timeline it is most often the case that this interval is not present in this timeline. The software under test will in such case raise an exception. Exceptions are handled in the function `well_defined`, which takes a generator as input, and recomputes it if exceptions are raised under evaluation. When sampling this data generator, about 70% of all values created is the empty timeline, followed by timelines with one or at most two intervals in it. That does not make for good test data.

Lets improve the `timeline` generator. However, before doing so, we want to formulate requirements on the timeline generator we are trying to construct. A requirement that 2% of the test cases should have a symbolic timeline that after evaluation contains more than 10 intervals would be specified as follows:

```
req_length() ->
  ?REQ_MIN(SymT, timeline(),
           length(eval(SymT)) > 10, 2.0).
```

A requirement that any set of generated values should have at least one timeline with an interval that spans over a year border is specified as follows:

```
req_year_span() ->
  ?REQ_EXISTS_FOR(SymT, timeline(),
    lists:any(fun({{Y1, _, _}, {Y2, _, _}}) ->
                Y1 < Y2
              end, eval(SymT))).
```

Similarly, a requirement that checks if an interval is present that spans over a month, is defined as follows:

```
req_month_span() ->
  ?REQ_EXISTS_FOR(SymT, timeline(),
    [1 || {{_, M1, _}, {_, M2, _}} <- eval(SymT),
          M1 < M2] =/= []).
```

We have specified additional requirements on the timeline generator, but we omit the definition.

The symbolic representation of calls helps us to define requirements on the construction of timelines. Since the data generator has a structure in which we save which calls we apply instead of the final result, we can express a requirement that 10% of the generates timelines should have been build with both a `delete` and a `tail` in its construction. With those requirements and the above generator for timelines, we get the following result:

```
4> eqc_requirements:req_module(booking_eqc).
Failed! After 1 tests.
Requirement req_length failed:
  only 0.00% meets the condition.

Failed! After 1 tests.
Requirement req_consecutive failed:
  only 0.00% meets the condition.

Failed! After 1 tests.
Requirement req_mix1 failed:
  only 0.00% meets the condition.

Failed! After 1 tests.
Requirement req_year_span failed:
  only 0.00% meets the condition.

OK, passed 1 tests

false
```

The failure rate is 100% for all but the requirement that we should have an interval over the month border. This means that none of the generated values fulfils any of the other requirements.

By selecting existing intervals from the earlier generated timeline and only taking the tail from a timeline that contains at least one interval we can do much better. The improved generator is defined as follows:

```
timeline() ->
  ?SIZED(Size, well_defined(timeline(Size))).

timeline(0) ->
  {call, ?API, new, []};
timeline(N) ->
  ?LAZY(
  ?LETSHRINK(
    [SymT],
    [well_defined(timeline(N-1))],
    begin
      T = eval(SymT),
      frequency(
        [{50,{call, ?API, add, [interval(),SymT]}},
         {1, {call, ?API, after_, [date(),SymT]}}]
        ++
        [{5, {call, ?API, tail, [SymT]}}
          || T =/= []]
        ++
        [{5, {call, ?API, delete,
          [elements(T), SymT]}} || T =/= []])
    end)).
```

This generator performs much better and passes all requirements with good margins.

Specifying requirements provides the developers the tools needed to ensure that data generators meet the expectations on test cases that they would use in manually written unit tests. In a similar way as deciding which unit tests one should write, we now decide which particular data distributions provide valuable test data. After that, we specify one property per operation and check the result against a model. In this way, we do get the complete testing as described in literature plus an additional quality assurance on the generated test data. In practice, this has helped us to motivate the designers of the generators to realise the short-comings of early versions of the generators and to improve them iteratively.

VI. Application 2: Dudle

Dudle is an open source web service, which can be used to schedule a meeting or poll people for an opinion. It is a relatively small web service with a simple and well defined interface. In case of a schedule, users can vote for one or more time slots, and in case of a poll, users can choose several

options. Dudle has functionality for creating, deleting, editing a schedule or a poll. Participants can be invited via Dudle to take part in a schedule or a poll. And finally, the administrator of a schedule or poll can review the status in order to see which alternatives are preferred by the participants. Dudle is written in the programming language Ruby and can be deployed using a web server, such as Apache, via a common gateway interface (CGI).

We have tested the Dudle web service with QuickCheck, using the abstract state machine functionality. We maintain a model of the Dudle system while executing test commands, which are mapped to CGI-calls, and checking pre- and postconditions. We have developed a number of test data generators for this test, such as generators for a poll name, or a time slot. In this section we focus on a test data generator for a user name. We started out with a textbook case of a generator for random user names:

```
name() ->
  ?SUCHTHAT(
    Name,
    eqc_gen:non_empty(eqc_gen:list(eqc_gen:char())),
    not lists:member($\r, Name)).
```

We have constructed this generator in terms off standard QuickCheck generators. The `name()` generator produces a non-empty list of characters. We use the `?SUCHTHAT` macro to exclude user names containing carriage returns. To ensure that we pick equal names now and again we do not use this generator directly, but we use it to create a pool of names from which we choose.

We had to improve the user name generator, such that it generates more realistic user names. Dudle was not always able to handle peculiar user names, for example names containing newlines and spaces. We do not blame Dudle for this, instead, we blame our slightly naive generator. The improved version of the user name generator is defined as follows (where Erlang's notation for a character is preceded by a dollar sign):

```
name2() ->
  Gen = frequency([{100, choose($a, $z)},
                   {25,  choose($A, $Z)},
                   {25,  choose($0, $9)},
                   {5, $ },
                   {1, $-}, {1, $_}]),
  ?LET(Name,
       eqc_gen:non_empty(eqc_gen:list(Gen)),
       string:strip(Name)).
```

This generator also produces a non-empty list of characters, but the characters are selected more carefully. Instead of choosing random characters we now choose alpha-numeric characters and occasionally a slightly unusual character, such as a space or a dash.

We have used the above generator in testing Dudle and are quite satisfied with it. But does it actually produce the user names that we expect? That is, does it meet our implicit requirements? Lets find out and make these requirements explicit, and specify them using the macros from Sect. V.

We had the following implicit requirements in mind when we defined the user name generator:

1) at least 10% of the generated user names should contain an unusual characters, such as a dash,
2) we should not generate names with more than four spaces,
3) a quarter of the generated user names should be longer than 8 characters,
4) we want to generate user names containing both upper and lower case characters.

These implicit requirements can be translated to formal requirement using the requirement macros as follows:

```
req_unusual() ->
  Intersect =
    fun(Xs, Ys) ->
      [X || X <- Xs, lists:member(X, Ys)]
    end,
  ?REQ_MIN(Name, name2(),
           length(Intersect(Name, "-_ ")) > 0,
           10.0).

req_spaces() ->
  Spaces =
    fun(Xs) ->
      lists:filter(fun(X) -> X == 32 end, Xs)
    end,
  ?REQ_EXISTS_FOR(Name, name2(),
                  length(Spaces(Name)) < 4).

req_name_length() ->
  ?REQ_MIN(Name, name2(),
           length(Name) > 8, 25.0).

req_upper_lower_case() ->
  IsUpper =
    fun(X) ->
      X >= $A andalso X =< $Z
    end,
  IsLower =
    fun(X) ->
      X >= $a andalso X =< $z
    end,
  HasUpperAndLower =
    fun(Xs) ->
      length([X || X <- Xs, IsUpper(X)]) > 0
      andalso
      length([X || X <- Xs, IsLower(X)]) > 0
    end,
  ?REQ_EXISTS_FOR(Name, name2(),
                  HasUpperAndLower(Name)).
```

We have defined these requirements in the Dudle test module named `dudle_eqc`. We use the `req_module` function to test all requirements defined in the Dudle test module, which generates the following output:

```
5> eqc_requirements:req_module(dudle_eqc).
OK, passed 1 tests

Failed! After 1 tests.
Requirement req_unusual failed:
  only 7.00% meets the condition.

OK, passed 1 tests

Failed! After 1 tests.
Requirement req_name_length failed:
  only 8.00% meets the condition.

false
```

These results show that the `name2` generator does not meet two of the four requirements, namely `req_unusual` and `req_name_length`. The latter suggests that the length of the generated user names are too short. This may explain why the requirement `req_unusual` fails as well, since the unusual characters have a low probability of being generated. We adapt the user name generator such that it generates longer names:

```
name3() ->
  Gen = frequency([{100, choose($a, $z)},
                   {25,  choose($A, $Z)},
                   {25,  choose($0, $9)},
                   {5, $ },
                   {1, $-}, {1, $_}]),
  ?LET(Name,
       eqc_gen:non_empty(eqc_gen:longlist(Gen)),
       string:strip(Name)).

longlist(Gen) ->
  ?SIZED(Size,
         resize(Size*2, list(resize(Size, Gen)))).
```

Most of the generator is left as is, but we have replaced the standard `list` generator with our own `longlist` generator. The `longlist` generator produces lists that are double the size of lists generated by the `list` generator. Lets check the requirements again:

```
6> eqc_requirements:req_module(dudle_eqc).
OK, passed 1 tests

OK, passed 1 tests

OK, passed 1 tests

OK, passed 1 tests
true
```

The `name3` generator meets all the requirements. This example shows that testing requirements supports the development of good test data generators. Not only does testing requirements have added value for validating large complex generators, but also for simple straightforward generators, such as the user name generator. It is all too easy to overlook something, such as generating list of the proper length.

VII. Conclusions

From experience, strengthened by a scientific experiment [15], we know that it is difficult to write test cases that cover a good set of input data, both positive and negative data. Random generation of data makes testing immune to specific choices, but also introduces the possibility to generate data that does not cover border cases or specific inputs.

When QuickCheck data generators get more complicated to write and their distributions harder to grasp, one can get a false sense of trust by seeing many test cases pass. The actual generated data can be collected by built-in QuickCheck functions and printed as side-effect of testing. However, only presenting the data requires either domain experts to asses the statistics or forces engineers to subjectively judge whether the collected values are satisfactory.

In this article we contribute by showing how one can express and verify requirements on generators to convince oneself that the performed testing is sufficient. Interaction with domain experts is needed at the beginning of the test design, when the requirements on test data are stated. With data from testing two different web services, we have shown that with a naive approach to random data generation, we can easily produce test cases without the required quality. We have shown how we can make the quality requirements explicit and automatically verifiable. And finally, we have shown how to control the randomness so that the test cases we produce are of the required quality.

Mutation testing is a different way of judging the quality of a test suite. This is based upon introducing errors in the software under test and trying to find them by running the test suite. Mutation testing is a fundamentally different technique and requires code instrumentation with good mutants. It is further research how these techniques complement each other.

With the techniques presented in this paper, domain and test experts are able to write requirements to ensure that the tests they perform are of high quality. It allows for a high degree of automation by minimal intervention of domain experts and automatic feedback on the quality of the generated data.

References

[1] J. Derrick, N. Walkinshaw, T. Arts, C. B. Earle, F. Cesarini, L.-Å. Fredlund, V. M. Gulías, J. Hughes, and S. J. Thompson, "Property-based testing - the protest project," in *FMCO*, ser. Lecture Notes in Computer Science, F. S. de Boer, M. M. Bonsangue, S. Hallerstede, and M. Leuschel, Eds., vol. 6286. Springer, 2009, pp. 250–271.

[2] K. Claessen and J. Hughes, "QuickCheck: a lightweight tool for random testing of Haskell programs." in *Proceedings of ACM SIGPLAN International Conference on Functional Programming*, 2000, pp. 268–279.

[3] T. Arts, J. Hughes, J. Johansson, and U. T. Wiger, "Testing telecoms software with Quviq QuickCheck," in *Erlang Workshop*, M. Feeley and P. W. Trinder, Eds. ACM, 2006, pp. 2–10.

[4] A. Nilsson, L. M. Castro, S. Rivas, and T. Arts, "Assessing the effects of introducing a new software development process: a methodological description," *Int. J. on Software Tools for Technology Transfer*, pp. 1–16, 2013.

[5] "Property-based testing of web services," http://www.prowess-project.eu, 2012-2015.

[6] T. Arts, L. M. Castro, and J. Hughes, "Testing erlang data types with Quviq QuickCheck," in *Proceedings of the ACM SIGPLAN Workshop on Erlang*. ACM Press, 2008, pp. 1–8.

[7] T. Arts and L. M. Castro, "Model-based testing of data types with side effects," in *Proceedings of the 10th ACM SIGPLAN workshop on Erlang*, ser. Erlang '11. New York, NY, USA: ACM, 2011, pp. 30–38.

[8] "Dudle," https://dudle.inf.tu-dresden.de, 2013.

[9] C. Soldani, "QuickCheck++," http://software.legiasoft.com/quickcheck/, 2010.

[10] T. Jung, "Java implementation of QuickCheck," http://quickcheck.dev.java.net/, 2010.

[11] C. League, "Qcheck/sml," http://contrapunctus.net/league/haques/qcheck/, 2010.

[12] J. Hughes, "Quickcheck testing for fun and profit," in *Practical Aspects of Declarative Languages*, ser. Lecture Notes in Computer Science, M. Hanus, Ed. Springer Berlin Heidelberg, 2007, vol. 4354, pp. 1–32.

[13] J. Armstrong, *Programming Erlang: Software for a Concurrent World*. Pragmatic Bookshelf, 2007.

[14] Z. Kessin, *Building Web Applications with Erlang: Working with REST and Web Sockets on Yaws*. O'Reilly, 2012.

[15] S. Eldh, H. Hansson, and S. Punnekkat, "Analysis of mistakes as a method to improve test case design," in *Proceedings of the 2011 Fourth IEEE International Conference on Software Testing, Verification and Validation*, ser. ICST '11, 2011, pp. 70–79.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1343–1348

A method for selecting environments for software compatibility testing

Łukasz Pobereżnik
AGH University of Science and Technology, Cracow, Poland

***Abstract*—Modern software is developed to work with multiple software and hardware architectures, to cooperate with various peer components and can be installed in many different configurations. In order to test it, all possible working environments needs to be created. This requires software and hardware resources like servers, networks and software licenses and most important: man-hours of qualified engineers that will have to configure and maintain them. Because resources are usually limited we have to choose a set of configurations with highest impact on quality of software under test. In this paper we present a method of measuring effectiveness of given software environment for discovering defects in software by introducing environment sensitivity measure. We also show how it can be used in simple algorithm used to select best configurations by using only a selected subset of them and progressively modifying it thougout software development process.**

I. Introduction and problem description

SOFTWARE usually does not work alone. It must have an environment that it works in. This environment can be composed from many components like: servers, operating systems, databases, remote services etc. Those components can also have other components that they rely on. Eg. database might need an operating system to work on. Those dependencies create a Component Dependency Graph (CDG) that describes an environment for Software under Test (SUT). Example of such graph is given on Figure 1. This graph shows only general structure of environment. Each component may also have a set of properties like type, version number, architecture type, permissions, locales etc.

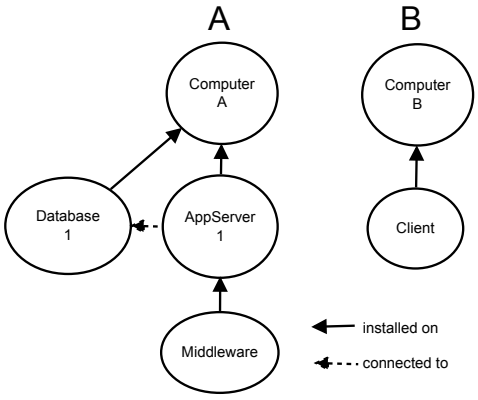


Fig. 1: Example of Component Dependency Graph (CDG) that shows dependencies between resources. Tree A represents environment for server application. Tree B is environment for client application.

Lets take a simple use case: an application working on two operating systems, with three database servers and two application servers. It will give about $2 \times 3 \times 2$=12 different environments to test. If we add another variable: 32-bit or 64-bit architecture, it will double possible environment configurations to at most 24. Adding new configurable element to environment tends to increase the number of possible setups exponentially. Not all configurations may be possible to create (for example some middle-ware may not be available for all operating systems), but it still is significant number of variants to test. Problems of generating test environments and possible solutions were mentioned in our other article [1].

There has been efforts to automate the process of creating those environments based on semantic description of CDG in [2] and [3]. Authors of those articles proposed to use virtualization to construct environments and then use snapshots to clone and then modify them to build other environments. This technique and additional simplifications allowed to reduce number of separate configurations from about 1200 to 160. However this is still to many environments to be build and maintained for everyday regressions tests or continuous builds.

In dissertation [4] same author came also to this conclusion and proposed a manual way to select subset of configurations based on testers' preferences. Decision on which configurations test software is in that case solely based on testers expert knowledge, without support of any analytical tools. In our research we tried to establish a method to measure how good is given configuration for testing and an algorithm to choose the best of them.

II. Selecting best environments for testing

As shown above number of possible environments can be quite high. This means that with limited resources we can only choose subset of them. One of the most popular methods is to use configurations that are most widely used by customers. However when number of software users is high, diversity of configurations may also be too high on and must also be limited.

We have to define what means that one configuration is better for testing purposes than the other and then create an algorithm for choosing the best of them. In our research we followed a common phenomena observed by testers: some of the software environments are causing more problems than the others - basically they fail more tests (or fail them more often). If configuration A is more problematic than configuration B that usually means that if we run tests on configuration A and

978-83-60810-53-8/$25.00

they will pass, so they will pass also on configuration B (with high probability). This means that we do not need to conducts tests on configuration B so frequently as on configuration A. Conclusion is that configuration A is better for testing than configuration B, because it allows to detect more environment related defects. In order to compare environments it is good to have a numerical metric that will allow to evaluate effectiveness of given configuration. It is also a requirement for many optimizing algorithms (especially evolutionary) to provide a fitness function to compare solutions.

A. Measure of environment sensitivity

In order to compare two environments for software testing we need to establish metric that would tell which configuration is better. Let a T be a set of n tests (test suite) consisting single tests t_i. Let T_k be a vector of test results executed in k iteration. Test can be either 1 (pass) or 0 (fail).

$$T_k = (t_1, t_2, t_3 ... t_n), t_i = 1 \vee 0 \tag{1}$$

E_j is an environment j. Testing function F_T is a function that assigns for each k iteration a vector of tests results T_k to environment C_j.

$$F_T(k, E_j) = (t_1, t_2, t_3, ..., t_n) \tag{2}$$

We can describe F_n for single iteration k in more convenient way as a matrix, where columns are tests and rows are environments (lets note number of configurations as m). t_{ji} is a result of test t_i on environment C_j.

$$F_T(k) = \begin{vmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{vmatrix} \tag{3}$$

First step in calculating sensitivity is to remove those tests that does not bring any information about environment differences. We remove those columns that satisfy condition:

$$\exists p \in [1, m] \forall x \in [1, n] \forall y \in [1, n] : t_{xp} = t_{yp} \tag{4}$$

This means that removed are only those tests that passed or failed in all configurations (remove columns of all 1 or all 0).

Then for each row vector we calculate how many times given test failed and normalize it by number of tests in vector (after removing some of them in first step).

$$Sens(C_j) = \frac{\sum_{x=1}^{n} (1 - t_{jx})}{n} \tag{5}$$

$$Sens(C_j) \in [0, 1] \tag{6}$$

Sensitivity value close to 0 means that given environment is not good for finding defects because all tests here pass. When sensitivity is 1 means that configuration is a good candidate for finding software errors because all tests fail on it whereas on at least one other configuration they pass. Of course, if all tests fail on given configuration we have to check if the problem is not with tests itself - for example there is a defect in testing code.

B. Properties of environment sensitivity

Let's define environment domination: environment A dominates environment B if:

$$\exists x : t_{Ax} < t_{Bx} \wedge \forall y \neq x : t_{Ay} \leq t_{By} \tag{7}$$

In other word: there is at least one test that failed on configuration A but passed on configuration B. This would mean that configuration A found a defect that was not discovered by configuration B.

Environment sensitivity has this property that:

$$A \; dominates \; B \Rightarrow Sens(C_A) > Sens(C_B) \tag{8}$$

This property is result of environment sensitivity definition. Let sum inequalities in second part of domination definition:

$$\sum_{k=1 \wedge k \neq i}^{n} t_{Ak} \leq \sum_{k=1 \wedge k \neq i}^{n} t_{Bk} \tag{9}$$

If we add $t_{Ai} < t_{Bi}$, weak inequality will become strong inequality:

$$\sum_{k=1}^{n} t_{Ak} < \sum_{k=1}^{n} t_{Bk} \tag{10}$$

Note that sensivity calculation requires removing tests that in every configuration failed or passed. This will also convert weak inequality into strong one. If we multiply both sides by -1 and add n:

$$n - \sum_{k=1}^{n} t_{Ak} > n - \sum_{k=1}^{n} t_{Bk} \tag{11}$$

Because $n = \sum_{k=1}^{n} 1$ then we can rewrite equation as:

$$\sum_{k=1}^{n} 1 - \sum_{k=1}^{n} t_{Ak} > \sum_{k=1}^{n} 1 - \sum_{k=1}^{n} t_{Bk} \tag{12}$$

$$\sum_{k=1}^{n} (1 - t_{Ak}) > \sum_{k=1}^{n} (1 - t_{Bk}) \tag{13}$$

Now divide both sides by n:

$$\frac{\sum_{k=1}^{n} (1 - t_{Ak})}{n} > \frac{\sum_{k=1}^{n} (1 - t_{Bk})}{n} \tag{14}$$

And now using environment sensitivity definition:

$$Sens(C_A) > Sens(C_B) \tag{15}$$

Introduction of environment domination allows to us to use existing multi-criteria optimization techniques to find Pareto-efficient solutions. This proof can also be used to quickly compare results of tests run on two configurations without calculating sensitivity itself. Of course it will only introduce order to the set of configurations but would not give any idea how much they differ.

C. Algorithm to generate configurations

Algorithm that will generate and evaluate environments must have several important properties:

1) Works on discrete solution spaces.
2) An ability to search unknown solution space (we have no additional information about local optima) .
3) Iterative schema of work, similar to iterative test execution.

Generating new environment and maintaining it is a costly operation. This means that algorithm must work with small data sets - typically 8-12 configurations. Tests should be run frequently, every code change or at least daily. This means that we may have enough iterations until we reach optimal solution. However we have to remember that each iteration means adding new configuration and this is a costly operation.

Algorithm 1 Simple algorithm for selecting most sensitive environments.

$E \Leftarrow GenerateAvailableEnvironments()$ {initial configuration pool}
$P \Leftarrow RandomSubset(E, n)$ {P represents current working set}
$P' \Leftarrow \emptyset$ {P' represents next set}
repeat
  $P'' \Leftarrow P'$ {P'' is set from previous iteration}
  $E \Leftarrow E - P$
  $R \Leftarrow RunTests(P)$
  $S \Leftarrow CalculateSensivity(R)$
  $S \Leftarrow SortBySensivity(S)$
  $P' \Leftarrow \emptyset$
  for $i = 0$ to k **do**
    $P' \Leftarrow P' \cup S[i]$
  end for
  $P \Leftarrow P' \cup RandomSubset(E, n - k)$
until E not empty and $P' \neq P''$

$GenerateAvailableEnvironments()$ creates a set of all possible environments we want to execute tests on. Function $RandomSubset(X, n)$ generates a random subset from set X of size n. Summarizing algorithm above: in each iteration we run test suite on n selected environments (working set). Using test results we calculate sensitivity for each of them. After that we select k best of them. From initial pool of environments we choose random ones to fill up working set so it will have n configurations again. Procedure is repeated until all configurations are used (initial pool is empty) or in next two consecutive iterations k best configurations is the same.

This algorithm tends to go though different solutions until it converge to optimal one. For environment compatibility testing this means that before we reach optimal set of configurations we will test much more of them and there is a possibility that we will find even more software defects, than using optimal set from the beginning.

III. Experiments

A. Direct application of environment sensitivity measure

To verify properties of environment sensitivity an experiment was conducted. We used a simple configuration with one operating system and a web browser installed on it. Operating systems used were Linux, Windows and Mac OS X. Browsers under test were Internet Explorer, Firefox, Chrome, Safari, Opera. Each browser was available in several different versions (depending on browser type). System used for experiment was built using small web server (Jetty) that was running a static web page. This web page was based on a popular HTML5 compatibility test site (www.html5test.com) and server both as test suite and application under test. Existing scripts were modified to send test results to data collection servlet running on the same web server (originally they were displayed on the screen). Schematic diagram of system used for experiment is shown on Figure 2.

When the test page was loaded it executed 242 true/false tests and sent results using JSON format back to server where they were stored in file along with information about browser and operating system type. The same page was run on each environment and test results were sent using separate script back to server that stored them for further analysis. Different configurations were provided by web browser compatibility testing cloud service (browsershots.org). We collected results for 46 different configurations. Expected number of environments should be higher, because some of the configurations has been not executed at all by the cloud (which is a defect in cloud service). However number of collected data is enough for analysis. In real life testing setups there are usually no more than several environments in constant use.

Sensitivity measure by definition is calculated relatively to other environments in tested configuration set. In most cases there is not enough resources (computing power, time, machines) to perform tests (and calculate sensitivity) for every environment in given configuration space. We wanted to check how much sensitivity measure will differ when it is calculated for small subset of configuration space against full configuration space. From all 46 environments we chosen randomly subsets of 5, 8 and 10 environments and calculated sensitivity for each configuration in it. We observed that sensitivity measure does not change more than 12 percent when it is calculated for a reasonable subset of initial test results (see Table I). If we use more than 8 environments it seems that average difference is less than 10 percent. We suppose that this behavior of measure is possible because the way environment reacts to tests is not dependent on other environments. This makes this sensitivity measure good candidate for fitness function in evolutionary programming.

B. Sensitivity measure as a fitness function

As stated before sensitivity measure can be used as fitness function so the next step was to check if proposed algorithm allows to quickly find best environments. Populations of sizes 8, 10 and 12 were tested. Each time algorithm was run 1000

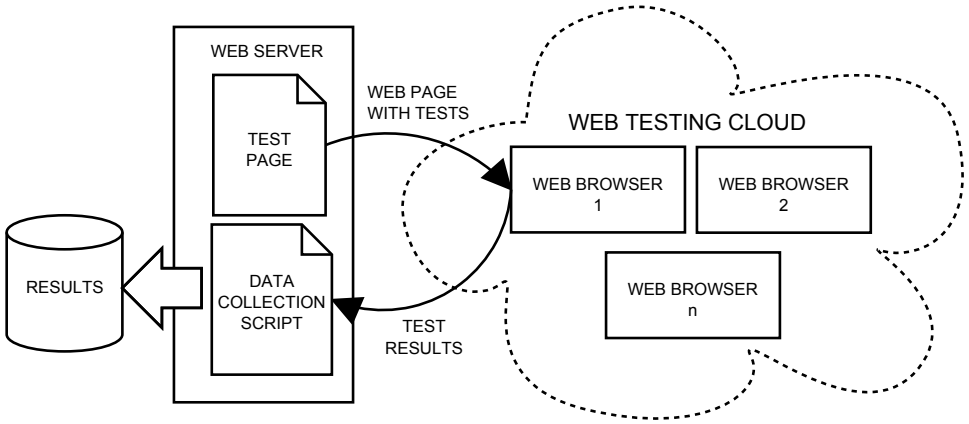


Fig. 2: Architecture of the system used for environment compatibility experiment.

TABLE I: Standard deviation of environment sensitivity calculated from random subsets from initial data.

	5 environments	8 environments	10 environments
100 random subsets	0.077	0.045	0.04
1000 random subsets	0.10	0.075	0.06
10000 random subsets	0.12	0.09	0.08

TABLE II: Averaged results after 1000 execution of environment selection algorithm. Difference is calculated against sensitivity calculated for all configurations together.

Population size	Max difference	Max iterations
n = 8, k = 4	15%	4.32 iterations
n = 10, k = 5	4.5%	4.09 iterations
n = 12, k = 6	3.8%	3.3 iterations

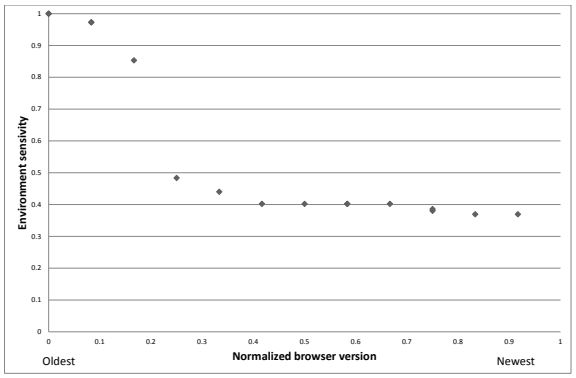


Fig. 3: Sensitivity of environment by browser version for Mozilla Firefox. Version numbers were normalized to be from 0 (oldest) to 1 (latest). You can see sudden improvement in HTML5 compatibility after third consecutive version.

times and results were averaged. They are presented in Table II. For population size of 10 algorithm delivered a stable set of environments in less than 5 iterations.

In Table III we can see browsers selected by algorithm in more than 10% of cases along with their average sensitivity. Columns Operating system, Browser type and Browser Version define environment. Frequency shows percentage of times given configuration was chosen in top k results in 1000 runs of algorithm (eg. 75% means that is was chosen in 750 times). Average sensitivity is arithmetic mean of sensitivity value calculated for environment in all runs. If we compare this table with sensitivity calculated for all environments in Appendix A Table IV they basically match each other.

C. Strategies for selecting environments for tests

Results also proved common sense that better to test on older versions of software because newer versions have lot of compatibility problems already fixed. This can be seen on Figure 3 and 4 where sensitivity of environment is presented versus browser version. We had to normalize version numbering to $[0, 1]$ because of different numbering schemes used by browser vendors. We can consider several strategies to reduce number of environments used for tests. Simplest one is to establish a cut off point below that every environment is discarded. On Figure 3 we see that good cut off point will be sensitivity with value 0.5 because it clearly separates set. However other good strategy will be to discard those some environments that have similar sensitivity value. From Figure 3 and Table IV we see that Firefox from version 6 to 15 have sensitivity between 0.3 and 0.4. This means that we can choose one or several of them based on own preference (or random choice) because they behave more or less similarly during tests.

Other important aspect is that environment sensitivity can provide an order of tests. If we start with environments with highest sensitivity and some tests will fail, we can stop, fix defect and start over again. In our test case, testing complicated web pages on latest browser versions will likely be successful,

TABLE III: Average sensitivity for environments using proposed algorithm (averaged after 1000 runs) for configuration set of size 10. Frequency show how many times given environment was chosen by algorithm in top k best. Only those configurations with frequency more than 10% are shown.

Operating system	Browser type	Browser version	Frequency	Average sensitivity
LINUX	FIREFOX	2.0.0.17	77%	0.960
LINUX	KONQUEROR	4.8	77%	0.889
MAC OS X	CAMINO2	2.1.2	76%	0.802
LINUX	FIREFOX	1.5.0.12	76%	0.983
WINDOWS	FIREFOX	2.0.0.12	74%	0.963
MAC OS X	SAFARI	4.0.5	50%	0.766
WINDOWS	CHROME	3.0.182.2	31%	0.777
WINDOWS	CHROME	4.0.223.11	24%	0.609
WINDOWS	OPERA	10.00	16%	0.388
LINUX	FIREFOX	7.0.1	12%	0.323
LINUX	FIREFOX	6.0.1	10%	0.326

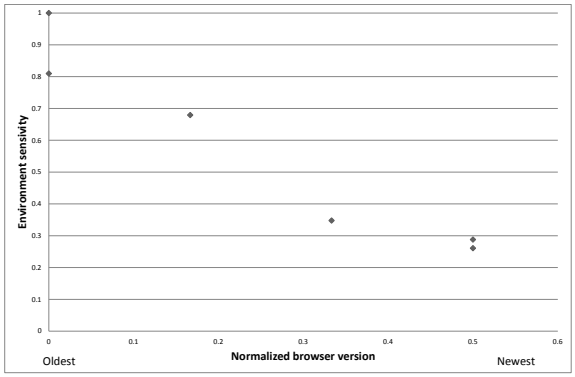


Fig. 4: Sensitivity of environment by browser version for Chrome. Version numbers were normalized to be from 0 (oldest) to 1 (latest). You can see improvement in HTML5 compatibility in latest versions. However it is not as steep as in Firefox browser.

because they are more HTML5 compatible. So better strategy will be to test on older versions and if they pass tests, then check on latest versions.

IV. Conclusions and future work

It seems that introduced environment sensitivity measure is a good way of measuring usefulness of environment for testing purposes. It provides analytical way to compare configurations and allows to use existing optimization techniques. For more complicated environments (that have several nodes in their CDG) we plan to use evolutionary algorithms. For presented browser testing case, cross-over and mutation operations were not feasible because they produced configurations that were not available in testing cloud. Introduction of environment domination (in Pareto sense) will allow to use existing methods used in multi-criteria optimization. Automated tests are usually run frequently in order to find out regression defects introduced during development. This causes tests to repeatedly oscillate between pass and fail states. We are now extending sensitivity model by introducing time line to take those changes into consideration and utilize historical information for more precise results.

We are also investigating possibility of using machine learning to correlate changes in application code base with historical test results to predict the best configuration and tests order to test on. This way when a new change is being introduced to software we can decide in which environment it should be tested in first place.

In our research we are planning to used multi-agent systems (See [5] and [6]) that will automatically deploy environments and optimize them for most efficient testing in terms of quality and resource consumption. Sensitivity is a useful measure to be used in algorithms that detect unusual behaviors like those mentioned in [7] and [8].

We are also considering introducing second measure based on probability that will cooperate with environment sensitivity that will allow us to better describe environment behavior and compare them in more than one category.

References

[1] L. Pobereznik, “Automatic generation and configuration of test environments,” in *Information Systems Architecture and Technology - Web Information Systems Engineering, Knowledge Discovery and Hybrid Computing*. Oficyna Wydawnicza Politechniki WrocÅĆawskiej, WrocÅĆaw, 2011, p. 303.

[2] I.-C. Yoon, A. Sussman, A. Memon, and A. Porter, “Direct-dependency-based software compatibility testing,” in *Proceedings of the twenty-second IEEE/ACM international conference on Automated software engineering*, ser. ASE ’07. New York, NY, USA: ACM, 2007, pp. 409–412. [Online]. Available: http://doi.acm.org/10.1145/1321631.1321696

[3] I.-C. Yoon, A. Sussman, A. Memon, and A. Porter, “Effective and scalable software compatibility testing,” in *Proceedings of the 2008 international symposium on Software testing and analysis*, ser. ISSTA ’08. New York, NY, USA: ACM, 2008, pp. 63–74. [Online]. Available: http://doi.acm.org/10.1145/1390630.1390640

[4] I. Yoon, “Compatibility testing for component-based systems,” Ph.D. dissertation, University of Maryland, 2010, hdl.handle.net/1903/11294.

[5] K. Cetnarowicz and V. Gruer, P. anf Hilare, "A formal specification of m-agent architecture," in *Proc. Multi-Agent Systems CEEMAS 2001*, L. N. i. A. I. v. . S.-V. Keplicz B., Nawarecki E., Ed., Berlin, Heidelberg, 2002, pp. 62–72.
[6] K. Cetnarowicz, "From algorithm to agent," in *Computational Science ICCS 2009, LNCS 5545 Springer Verlag*, 2009, pp. 825–834.
[7] K. Cetnarowicz and G. Rojek, "Behavior based detection of unfavorable resources," in *Proc. Computational Science - ICCS 2004*, G. S. P. e. a. L. N. i. C. S. v. . S.-V. Bubak, M; VanAlbada, Ed., Berlin, Heidelberg, 2004, pp. 607–614.
[8] K. Cetnarowicz and G. Rojek, "Behavior evaluation with actions' sampling in multi-agent system," in *Proc. Multi-Agent Systems and Applications CEEMAS 2005*, P. V. L. N. i. C. S. v. . S.-V. Pechoucek, M; Petta, Ed., Berlin, Heidelberg, 2005, pp. 490–499.

APPENDIX

In this section we present a table with sensitivity values calculated for different versions of popular browsers running on various operating systems. Sensitivity was calculated at once based on all test results from all available configurations.. In production it is usually not possible to keep so many testing environments, so only a small subset of them is used for daily testing and more of them are added when needed (for example before product release). You can compare results from this table with values from Table III.

TABLE IV: Sensitivity values calculated for all configurations (only non-zero values are shown). In this case sensitivity was calculated for all environments at once.

System	Browser	Browser version	Sensitivity
LINUX	FIREFOX	1.5.0.12	1.000
LINUX	FIREFOX	2.0.0.17	0.971
WINDOWS	FIREFOX	2.0.0.12	0.971
LINUX	KONQUEROR	4.8	0.902
MAC OS X	CAMINO2	2.1.2	0.821
MAC OS X	SAFARI	4	0.202
WINDOWS	CHROME	3.0.182.2	0.798
WINDOWS	CHROME	4.0.223.11	0.659
WINDOWS	OPERA	10.00	0.445
LINUX	FIREFOX	7.0.1	0.405
LINUX	FIREFOX	6.0.1	0.405
MAC OS X	FIREFOX	11.0	0.364
MAC OS X	FIREFOX	12.0	0.364
MAC OS X	FIREFOX	13.0.1	0.364
MAC OS X	FIREFOX	14.0.1	0.364
MAC OS X	FIREFOX	15.0.1	0.358
WINDOWS	FIREFOX	11.0	0.358
WINDOWS	FIREFOX	16.0	0.341
LINUX	SAFARI	5.0	0.312
WINDOWS	CHROME	5.0.375.125	0.306
WINDOWS	CHROME	6.0.453.1	0.243
LINUX	CHROME	6.0.472.63	0.214
MAC OS X	SAFARI	6.0.1	0.208
WINDOWS	CHROME	7.0.517.44	0.173
LINUX	CHROME	20.0.1132.47	0.075
LINUX	CHROME	22.0.1229.94	0.052
MAC OS X	CHROME	22.0.1229.94	0.046
WINDOWS	SAFARI	5.0	0.046
WINDOWS	OPERA	11.64	0.017

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1349–1352

An Evaluation of Data Race Detectors Using Bug Repositories

Jochen Schimmel, Korbinian Molitorisz, Walter F. Tichy
Karlsruhe Institute of Technology (KIT), Germany
{schimmel, molitorisz, tichy}@kit.edu

***Abstract*—Multithreaded software is subject to data races. A large number of data race detectors exists, but they are mainly evaluated in academic examples. In this paper we present a study in which we applied data race detectors to real applications. In particular, we want to show, if these tools can be used to locate data races effectively at an early stage in software development.**

We therefore tracked 25 data races in bug repositories back to their roots, created parallel unit tests and executed 4 different data race detectors on these tests. We show, that with a combination of all detectors 92% of the contained data races can be found, whereas the best data race detector only finds about 50%.

***Index Terms*—Data Races, Unit Testing, Multicore Software Engineering, Empirical Study**

I. Introduction

ALMOST all race detection approaches are evaluated in source code, that is freely available or in productive use. But can these detectors also be effectively used during software development before delivery and prevent shipping errors? We conducted an empirical study to answer this question. For our experimental setup we browsed bug repositories of open source applications for reports of data races. For each report we tracked back the revision history to the point, at which the defect has been checked into the code repository for the first time. We used this revision to evaluate which of the data race detectors would have been able to find the race at the distinct moment where it was unintentionally inserted.

A first key finding was, that it was almost impossible to simply apply a race detector on our evaluation programs: The application of most race detectors was impractical and the true data races were outbalanced by the huge number of false positives. Furthermore, most data race detectors available consume too much memory and computation time. Results tend to be impressive when applied to small programs, but with increasing sizes of real world applications, race detection approaches become increasingly impractical. Our solution to this problem was to create parallel unit tests[1] for all programs: A parallel unit test calls two (or more) methods under test in parallel within separate threads. In contrast to regular unit tests, they do not contain assertions. A data race detector decides on the test result. It executes parallel test cases rather than the program itself. By writing such parallel unit tests, we divided the programs into smaller fractions and focused error detection on the relevant portions, that were small enough to be handled by the race detectors. As a consequence, we could successfully apply all four race detectors to the bug repositories and locate 92% of the data races.

This paper is structured as follows: In section II, we introduce the sample applications we used as benchmark for the data race detectors and present the bugs we found in the respective repositories. Section III presents the four data race detectors we evaluate. We discuss the results of our study in section IV and detail on parallel unit tests in section V. The paper concludes with related studies in section VI and a conclusion of the key findings.

II. Sample Applications and Bug Repositories

Apache Tomcat is a web server written in Java and uses Bugzilla as bug tracking system. In Tomcat, each web request is handled by a separate thread and all of them access common data. No or incorrect synchronization of the common data leads to program stalls or incorrect data. In Bugzilla we tracked down 23 reports of data races in Tomcat due to synchronization errors.

Spring is an application development framework library for Java and uses Jira to track bugs. Spring contains framework classes, that can be executed both sequentially or in parallel. We tracked 24 data race reports in the bug database caused by altered program semantics when executing the parallel versions of the framework classes.

Eclipse is an integrated development environment for Java and other programming languages and tracks bugs using Bugzilla. Eclipse executes long-running computations in background threads to keep the user interface responsive. We tracked 18 synchronisation errors in Bugzilla concering long-running background threads.

Defect Classification: We categorize the defects according to their root into four different error patterns: (1) Atomicity violation, (2) wrong usage of Java library, (3) if-race and (4) bad optimization. A data race may account to more than one of the four error patterns. However, some of the defects we found are specific and do not apply to any of these categories.

Atomicity violations are data races caused by incorrect granularity of synchonization. Here, different memory location have data- or control flow dependencies. They form a logical unit and may only be changed atomically or in a transaction. Even if each location might be synchonized separately, the acceess to the whole unit is not.

The Java library contains thread-safe classes, i.e. they can be used in multithreaded applications without additional locks.

978-83-60810-53-8/$25.00

TABLE I
DATA RACES AND DETECTION RESULTS

Program Details			Bug Details					MTRAT	ConTest	Jinx	Jchord		Enriched PUT
Bug-ID	Program	Σ Methods	Impact	Atom.	Lib	if-Race	Opt.	Det.	Det. (%)	Det. (%)	Det.	FPs	Det. (%)
4418	Tomcat	1	Crash			x	x	x	70	60	x	1	90
31018	Tomcat	1	Crash		x	x			60	0			0
48790	Tomcat	1	Wrong Results			x		x	0	0	x		0
728	Tomcat	2	Crash				x	x	80	0	x	1	0
48177	Tomcat	1	Crash			x		x	100	90	x		100
46085	Tomcat	2	Deadlock					x	0	0	x	10	0
48172	Tomcat	2	Wrong Results					x	0	0	x	1	0
36173	Tomcat	2	Crash		x		x	x	0	0		2	0
SPR-5658	Spring	1	Crash			x			0	0			0
SPR-4932	Spring	1	Crash			x		x	20	20			10
SPR-4938	Spring	1	Deadlock	x	x	x	x		70	10		10+	90
INT-748	Spring	2	Crash		x				30	0		10+	30
SPR-3228	Spring	2	Crash		x				70	100			80
SPR-4672	Spring	1	Crash	x					0	0			0
SPR-2000	Spring	2	Crash					x	100	0			100
SPR-3432	Spring	1	Crash						0	0	x		0
INT-1072	Spring	2	Crash		x				80	90		10+	0
44809	Eclipse	2	Crash					x	20	0	x	10+	0
104294	Eclipse	1	Crash		x	x			100	100	x	10+	100
163685	Eclipse	2	Crash					x	0	0			0
296822	Eclipse	2	Wrong Results					x	0	0	x	1	0
31159	Eclipse	1	Wrong Results		x	x			0	50			100
272742	Eclipse	1	Crash					x	30	0	x		0
298648	Eclipse	1	Wrong Results		x	x			90	0	x	10+	0
36659	Eclipse	1	Crash			x			60	100	x		0
Σ (25 total)		1: 14 / 2: 11		2	9	11	4	13	15	9	13	12	9

The second error-pattern we define classifies code locations as defect in which classes are treated as thread-safe but that are not designed to be used like this.

Almost half of all errors we found are if-races: An if-race occurs, when a variable is checked for a certain value inside a conditional expression leading to a branch. In there, the variable is updated to a new value. For correct program semantics, both statements must be within the same lock because the thread could otherwise be interrupted in between and the variable is changed by another parallel operation.

The bad optimization pattern does not reflect a certain code design pattern, but rather describes different kinds of errors which come from the intention to improve program runtime but not its code behaviour. In fact, these code changes have unintended side effects. We could only identify these errors because of comments we found in the bug repositories.

We summarize our results in table I. One central finding is, that practically all data races can already be reproduced by a twofold parallel execution, although the programs may execute the parallel regions at a higher degree of parallelism. If the parallel regions execute in the wrong order at runtime, the defect manifests. 76% of all defects belong to this category. 20% require three or four operations, while only 4% require at least five operations in an unexpected order. With just one exception, we could successfully reproduce the defects using two threads. The column *Method sum* under *Program Details* shows, that roughly half of the data races are caused by a parallel execution of two different methods, whereas the other half is caused by parallel execution of the same method. This relates to the key finding by Shan Lu et al. [2], according to which most data races can be reproduced with only two methods. According to our finding, we can generalize the term of *two methods* to either depict the parallel execution of two separate methods as in task parallelism or to depict the parallel execution of the same method as in data parallelism.

III. DATA RACE DETECTORS

We present the four data race detectors used in our study. The selected detectors had to be freely available and were required to support Java code natively for comparability reasons. We included three dynamic and one static tool. Further studies should also include other prominent tools available such as Helgrind+ [3] or Racer-X [4].

MTRAT: Multi-Thread Run-time Analysis Tool for Java (MTRAT) is a dynamic detection tool for data races and deadlocks developed by IBM [5]. MTRAT uses a combination of the happens-before and lockset race detection algorithms. For this work, we used its Eclipse plugin for Windows in order to define which classes of an application to instrument. MTRAT can also instrument libraries at bytecode level, except for Java core libraries. For self-written code, MTRAT returns the line numbers where the error occured; for errors in libraries, only the class name is returned. MTRAT captures the execution path and makes the data race reproducable. Unlike other race detectors, it is unable to identify alternative control flows. For us to work with MTRAT, we manually wrote test cases, that induced the problematic control flow. Thus, the investigation of complex programs such as Eclipse is not feasable due to slowdown limitations.

ConTest is another dynamic tool developed at IBM alphaworks [5]. ConTest inserts sleep and yield instructions heuristically into Java bytecode to create different thread interleavings. When re-executing an application, ConTest varies the thread

scheduling to provoke data races and deadlocks. ConTest is available as an Eclipse plugin. It offers numerous options to adjust the interleaving heuristics. In contrast to other race detection tools, ConTest does not identify data races, it only provokes different interleavings, so ConTest raises the chance for a data race to occur. The developer finally has to identify the race by invalid program behaviour using assertions.

Jinx is a commercial tool to find errors in multithreaded applications [6]. It supports programs in Java, C/C++ and .NET-languages. Jinx is a dynamic race detector and executes a program several times, altering thread schedules. When a race is detected, Jinx can replay the problematic schedule. Like ConTest, Jinx relies on programs throwing exceptions as soon as a data race occurs.

Jchord is an open source static data race detector [7]. Jchord is a command line tool which expects the Java classes under test as input; it will also detect errors within compiled Java libraries, such as the Java core libraries.

IV. Results

Table I summarizes all 25 bugs and the data race detection results of all 4 evaluated tools. The column *Bug-ID* references the respective bug tracking system. We evaluated each of the race detectors with the same unit tests as program input. For ConTest and Jinx we had to extend the parallel test cases to include exceptions and assertions. We found, that by executing 9 defects could even be found without any data race detector. We call this extension *enriched parallel unit tests*. The results are shown in column *enriched PUT*.

MTRAT is unaware of atomicity violations and cannot find defects due to wrong library usage, as it does not check the Java core libraries. To verify the library limitation, we used an open source implementation of the Java library. With this change MTRAT would have found all 9 library defects. MTRAT exhibited false positives on one occasion only. According to its heuristics, MTRAT could have found 15 data races, 13 were in fact found.

ConTest alternates thread interleavings, so its results are not reliable. ConTest can only find data races, when they actually occur. We therefore executed ConTest 10 times and measured if the race was reported at least once. Another weakness is, that ConTest reports races on the basis of hand-written exceptions or assertions that fail, so we extended our benchmark to use enriched PUTs. With this, ConTest could identify 9 additional defects. A third drawback of ConTest is the lack of information to resolve the defect: It only executes the test case. It does not provide any information about what caused the data race or at which code line.

As Jinx is also unreliable we used the same reproduction logic and used our enriched PUTs as input. Jinx offers several *intensity levels* to improve defect detection, that showed no noticeable difference in our experiments. Jinx is able to detect errors due to wrong library usage, atomicity violations and did not produce false positives, but found only 36% of the errors. With 9 defects, the enriched PUTs found as many errors as Jinx, but 2 of them were only found by Jinx. The slowdown of Jinx is within a few seconds, so it may be used as a supportive tool.

Jchord is a static race detector. It can be applied to the regular evaluation applications, but we used it on our test cases for comparability reasons. Also, this extension reduced the execution time drastically: With the regular test cases Jchord required 4 minutes on average and in 4 cases it crashed with out of memory exceptions. With enriched PUTs each test case executes within a few seconds; From the remaining 21 bugs, 13 were found. Jchord is unable to find atomicity violations and was the only tool to produce a significant number of false positives. For 6 test cases it reported more than 10 false positives. This severly lowers its benefit for real-life scenarios.

V. Parallel Unit Tests

Our evaluation shows the efficiency of data race detection supported by parallel unit tests. If parallel unit tests are available, they can be used as input for different race detectors. Combining all 4 detectors, we could identify 92% of the bugs. A combination of the two best race detectors MTRAT and ConTest still found 84%. This shows that parallel unit tests may be a veritable approach to ease data race detection. Some race detectors like CHESS [8] are specificially designed for parallel unit tests. However, writing sound parallel unit tests is hard. Therefore, the exploration of automatic generation of parallel unit tests is an active research topic [9, 10]. The parallel test cases we wrote for this study conform to this research: A parallel test case is a test method calling at least two program methods in separate threads; the test method exits as soon as the threads have returned. A parallel test method does not alter the thread schedule or influence the program execution in any way - this is left to the data race detector that executes the parallel test method. Parallel unit tests do not contain assertions or throw exceptions deliberately, the decision whether a race is found or not is completely left to the detector. As we showed, some race detectors break with this definition of a parallel unit test, as they require assertions in the test case. If a parallel test case contains assertions to detect the presence or negative effects of data races, we call it enriched.

Figure 1 shows a sample with a parallel test case and an enriched version. The test case executes *inc()* concurrently in two threads. After they return, the test exits. The results and side effects of the test are not evaluated, this is left to the execution environment, i.e. the race detector. In the second case, the enriched test case waits for both methods and will report an error if the value of *val* is not 2. This test is able to detect malicious race behaviour, but it depends on the concrete thread schedule and the race detector influences the probability to provoke unintended behaviour. Using enriched PUTs, complexity is transfered from detector design to test development; this may be a good approach for bugs that are hard to detect, like atomicity violations. Here, semantic information on the programmer's intention is required to identify an error. Even in our small sample, we show that MTRAT cannot find them, whereas detectors using enriched tests such

```
class CInc {

private int val;

public void inc() {
  val++;
}

public int getVal() {
  return val;
}

}
```

(a) The sample class CInc.

```
class IncrementTest  {
static CInc inc = new CInc();

public static void Main() {
  Thread t1 = new Thread(
    new Runnable() {
    public void run() {inc.inc();}
  });
  Thread t2 = new Thread(
    new Runnable() {
    public void run() {inc.inc();}
  });
  t1.start(); t2.start();
}
```

(b) A parallel test case for CInc.

```
class IncrementTestX {
static CInc inc = new CInc();

public static void Main() {
  Thread t1 = new Thread(
    new Runnable() {
    public void run() {inc.inc();}});
  Thread t2 = new Thread(
    new Runnable() {
    public void run() {inc.inc();}});
  t1.start(); t2.start();

  try { t1.join(); t2.join(); }
  catch (InterruptedException e) { }
  if (inc.getVal() != 2)
    System.err.println("Error!");
}
```

(c) Enriched parallel test case for CInc.

Fig. 1. Code excerpt of the Bank Account Sample with its instrumented versions.

as Jinx can. Nevertheless, developing enriched, sound parallel test cases is harder than usual parallel test cases and to our current knowledge, no automatic generation approaches exist.

VI. Related Work

Bug evaluation has been performed before: Shan Lu et al. [2] evalute 105 synchronisation bugs from large applications for bug patterns. In contrast to our work, no data race detectors have been evaluated. In [11] and [12], different Java race detectors are evaluated. However, the used defects are from artificial sample applications, not from real bug repositories. They indicate that static race detectors produce too much false positives and are hard to use. In [13], programs written in C/C++ are evaluated.

VII. Conclusion

In this work, we searched bug repositories of four large Java applications for historic data races reported by users. We then tested 4 well-known data race detectors with the program revisions which contained the bugs for the first time. Seen individually, each of the four data race detectors found about 50% of the bugs. Together, the detectors found 92% of these bugs. In order to efficiently use the detectors, it is necessary to write specific test cases for data race detection. Our results indicate that a good, test based detection infrastructure combining different race detection approaches may help to find most data races early. However, writing good parallel test cases is hard and time-consuming. We therefore see our results as a motivation to automatically generate parallel unit tests. Different works heading in this direction have been mentioned. For future work, we plan to extend this study to more evaluation programs and other race detectors. As a combination of different tools seems promising, it would be interesting to know if a certain combination of detection strategies leads to optimal results. Furthermore, we want to search for data races using generated test cases from the works presented above.

Acknowledgment

The authors would like to thank Yana Stoeva for her support with this project.

References

[1] G. Szeder, "Unit testing for multi-threaded java programs," in *Proceedings of the 7th Workshop on Parallel and Distributed Systems*, 2009.

[2] S. Lu, S. Park, E. Seo, and Y. Zhou, "Learning from mistakes: a comprehensive study on real world concurrency bug characteristics," ser. ASPLOS XIII, 2008.

[3] A. Jannesari, K. Bao, V. Pankratius, and W. F. Tichy, "Helgrind+: An efficient dynamic race detector," ser. IPDPS '09, 2009.

[4] D. Engler and K. Ashcraft, "Racerx: effective, static detection of race conditions and deadlocks," *SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, pp. 237–252, Oct. 2003.

[5] *alphaWorks: Advanced Testing for Multi-Threaded Applications*, September 2010. [Online]. Available: http://www.alphaworks.ibm.com/tech

[6] *Corensic: Jinx*, November 2012. [Online]. Available: http://wiki.corensic.com/wiki

[7] *Jchord*, September 2010. [Online]. Available: http://code.google.com/p/jchord

[8] S. Q. Madanlal Musuvathi and T. Ball, "Chess: A systematic testing tool for concurrent software," Microsoft Research, Tech. Rep., Nov 2007.

[9] A. Nistor, Q. Luo, M. Pradel, T. R. Gross, and D. Marinov, "Ballerina: automatic generation and clustering of efficient random unit tests for multithreaded code," in *Proceedings of the 2012 International Conference on Software Engineering*, 2012.

[10] J. Schimmel, K. Molitorisz, A. Jannesari, and W. F. Tichy, "Automatic generation of parallel unit tests," in *ACM AST '13*, 2013.

[11] C. Artho, "Finding faults in multi-threaded programs," Masters thesis, Tech. Rep., 2001.

[12] A. K. Md Abdullah, Al Mamun, "Concurrent software testing: A systematic review and an evaluation of static analysis tools," 2009.

[13] S. Lu, Z. Li, F. Qin, L. Tan, P. Zhou, and Y. Zhou, "Bugbench: Benchmarks for evaluating bug detection tools," in *Workshop on the Evaluation of Software Defect Detection Tools*, 2005.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1353–1358

Test City metaphor as support for visual testcase analysis within integration test domain

Artur Sosnówka
West Pomeranian University of
Technology, ul. Żołnierska 49,
71-210 Szczecin, Poland
Email: arsosnowka@wi.zut.edu.pl

***Abstract*—the majority of formal description for software testing in the industry is conducted at the system or acceptance level, however most formal research has been focused on the unit level. This paper shows formal test selection and analyzes criteria for system or integration test based on visualization analysis for low level test cases. Visual analysis for low level test case selection is to be based on inputs from available Test Management system. The paper presents a use case for visual metaphor as a base for analysis testware for a test project in the industry.**

I. Introduction

Software development is dealing with growing complexity, shorter delivery times and current progress made in the hardware technology. Within the software lifecycle the biggest, however not directly seen part, is the maintenance. Number of used systems in the corporation is continuously increasing. During time progression users get trusted to the used software, so tolerated number of deviations is decreasing. As soon as software is put in the production environment, every big change or even small adaption of the source code can cause potential danger in the best case monetary in the worst case image or even human being losses. Nevertheless the maintenance is very often provided during the whole period through different groups of technicians or business partners. This makes the task of programming, understanding and maintaining of the source code for the system and its testware more complex and difficult. Testware management, especially for the high (HLTC) and low level test cases (LLTC) [8], which are focusing on old but still valid functionality keeps going to be not affordable, or omitted on purpose. This causes increasing maintenance costs to the limit, when new development can produce less cost and even be easier to implement than creation of the new functionality within the old system.

Required quality of the software is very often to be reached through quality assurance activities on several levels, starting from unit test, through system, integration and ending on acceptance tests. Artifacts produced during the test process required to plan, design, and execute tests, such as documentation, scripts, inputs, expected out-comes, set-up and clear-up procedures, files, databases, environment, and any additional software or utilities used in testing are named, according to ISTQB, testware [8]. Detection of the problems within a testware can save much effort and reduce necessary maintenance costs. Number of executed tests in the first or second year of software maintenance is not being a disruptive factor for the test projects. As soon as software is coming into the last phase, associated teams are very often moved to the other development projects or taken out of the company (e.g. consultants are being moved from customer to customer). To prove necessary quality after performed adaptations, growing complexity of the system is demanding high professional skills and understanding from people and organizations taken over the responsibility for the system.

Software quality is according to definition:

1. The degree to which a system, component or process meets specified requirements [21].
2. Ability of a product, service, system, component, or process to meet customer or user needs, expectations, or requirements [22].
3. Degree to which the system satisfies the stated and implied needs of its various stakeholders, and thus provides value [23].
4. Degree to which a system, component, or process meets customer or user needs or expectations [21].
5. The degree to which a set of inherent characteristics fulfills requirements [24].

Above given definition is obligating quality assurance teams to perform planned and systematic pattern of actions to provide adequate confidence to the product or item that it conforms to established technical requirements [2]. Execution of needed actions to provide at least same quality during the whole maintenance phase is a big cost factor. According to survey-analysis presented during the iqnite 2011 conference in Düsseldorf [19], almost 60% of the software projects are spending between 20 and 30% of its budget on Quality Management (QM) and testing activities.

Especially big and complex systems are providing large number of functions and demanding even larger number of objects within the testware. To provide 100% fulfillment the test team has to ensure that each function is not affected through the code adaptation and its site effects. Adaptation of the system demands adaptation of testware to fulfill quality requirement for the current system.

Even best managed testware, after few years of usage, is not free of objects which are old, obsolete, duplicated or

978-83-60810-53-8/$25.00

there are no HLTCs or LLTCs covering demanded functionality. Those objects are causing additional management effort and its existence does not increase expected quality needs.

Often developers and managers believe that a required change is minor and attempt to accomplish it as a quick fix. Insufficient planning, design, impact analysis and testing may lead to increased costs in the future. Over time successive quick fixes may degrade or obscure the original design, making modifications more difficult [7] and finishing in not acceptable, low quality of the system.

As long as we are accepting loose of the software and testware quality, its transparency, increasing maintenance costs, decreasing test efficiency, and continuous testware erosion is not a subject. However, in time of financial crisis and decreasing IT budgets, there is none of the project which can come over this dilemma. In the next chapters we would like to show results from pilot project which has been executed in the industry in order to prove usefulness for the approach of the visualization metaphor for testware reorganization.

II. Related Work

Since the early days of software visualization, software has been visualized at various levels of detail, from the module granularity seen in Rigi [13] to the individual lines of code depicted in SeeSoft [3]

The increase in computing power over the last 2 decades enabled the use of 3D metric-based visualizations, which provides the means to explore more realistic metaphors for software representation. One such approach is poly cylinders [20], which makes use of the third dimension to map more metrics. As opposed to this approach in which the representations of the software artifacts can be manipulated (i.e., moved around), our test cities imply a clear sense of locality which helps in viewer orientation. Moreover, our approach provides an overview of the hierarchical (i.e., package, test object) structure of the systems.

The value of a city metaphor for information visualization is proven by papers which proposed the idea, even without having an implementation. [15] Proposed this idea for visualizing information for network monitoring and later [14] proposed a similar idea for software production. Among the researchers who actually implemented the city metaphor, ([9]; [1]; [18]) represented classes are districts and the methods are buildings. Apart from the loss of package information (i.e., the big picture), this approach does not scale to the magnitude of today's software systems, because of its granularity.

The 3D visual approach closest in focus to ours is [10], which uses boxes to depict classes and maps software metrics on their height, color and twist. The classes' box representations are laid out using either a modified tree map layout or a sunburst layout, which split the space according to the package structure of the system. The authors address the detection of design principles violations or anti-patterns by visually correlating outlying properties of the representations, e.g., a twisted and tall box represents a class for which the two mapped metrics have an extremely high value. Besides false positives and negatives, the drawbacks of this approach is that one needs different sets of metrics for each design anomaly and the number of metrics needed for the detection oftentimes exceeds the mapping limit of the representation (i.e., 3). The detection strategies [12] were introduced as a mechanism to formulate complex rules using the composition of metrics-based filters, and extended later [11] by formalizing the detection strategies and providing aid in recovering from detected problems.

III. Visualization Metaphor

A visualization metaphor is defined as a map establishing the correspondence between concepts and objects of the application under test and a system of some similarities and analogies. This map generates a set of views and a set of methods for communication with visual objects in our case - test cases [6].

Lev Manovich has said: "an important innovation of computers is that they can transform any media into another". This gives us possibility to create a new world of data art that the viewer will find as interesting. It does not matter if the detail is important to the author; the translation of raw data into visual form gives a viewer possibility to get information which is the most important just for him. Hence, any type of visualization has specific connotations, which may become metaphoric when seen in context of a specific data source. Metaphor in visualization works at the level of structure, it compares the composition of a dataset to a particular conceptual construct, and the choice of any visualization is always a matter of interpretation.

Numerous currently existing visualization systems are divided into three main classes:

→Scientific visualization systems [4];

→Information visualization systems [5];

→Software visualization systems [16]

Although all visualization systems differ in purposes and implementation details, they do have something common; they manipulate some visual model of the abstract data and are translating this into a concrete graphical representation.

In this paper we are not aiming to present all possible visualization metaphors, as this is not the focus for our research. We would like to show basic and easy to understand "City metaphor" which is helpful for representation specific test data and allow easier test reorganization. After some of the previous research work which is however not in focus of this paper we settled our first attempt to the metaphor which is very widely presented in [17] and is a part of his PhD [17]. In its research and implementation for software source code classes are represented as buildings located in city districts which in turn represent packages, because of the following reasons:

→A city, with its downtown area and its suburbs is a familiar notion with a clear concept of orientation.

→A city, especially a large one, is still an intrinsically, complex construct and can only be incrementally explored, in the same way that the understanding of a complex system increases step by step. Using an all too simple visual

metaphor (such as a large cube or sphere) does not do justice to the complexity of a software system, and leads to incorrect oversimplifications: Software is complex; there is no way around this.

→Classes are the cornerstone of the object-oriented paradigm, and together with the packages they reside in, the primary orientation point for developers.

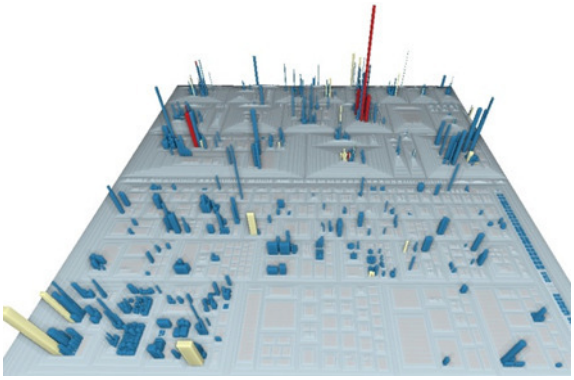


Fig. 1 Example of "Software City" representation of JBoss application server.

In our attempt we perform mapping between available LLTC and its basic metrics, perform testware reorganization and at the end provide easy to understand and manage overview about the current state of testware.

A. Test metrics

To be able to perform data visualization, defined set of the static and dynamic data has to be prepared. Based on the available information's for LLTC we are extracting following basic metrics, which we use for later mapping:

→Amount of LLTC
→Execution status for available LLTC
→Last modification date/age
→Number of executions

Dependent on the metrics type, those are to be taken as a data export through the available API from the test management tool or statistical data taken from the support or test organization.

Fetched metric can be mapped into the chosen visualization metaphor as:

→Data physical properties (color, geometry, height mapping, abstract shapes)
→Data granularity (unit cubes, building border or urban block related)
→Effect of Z axis mappings on the image of the city
→Abstraction of data and LOD are key issues
→Resulting "data compatible" urban models are much larger than the original VR urban models.

IV. Test Reorganization

In this paper we would like to show how useful can be usage of visualization based on the "Test City" metaphor. We would like to show how to perform test reorganization based on the very basic set of metrics available in the test project.

For our experimental work we have established a new system interacting with several Test Management applications placed on the market. The base idea of the system is an automation extraction and pre-evaluation of several different test metrics. Those metric are imported via available API connections from the Test Management tool and evaluated to get required set of metrics. The test metrics are provided as a text file, e.g. CSV (Comma Separated Values), and imported into visualization framework. Used visualization framework is based on the existing solution presented in [17] and allows us perform necessary analysis. The analysis result is taken as an input to the Test Management tool for Test-Set creation and evaluation.

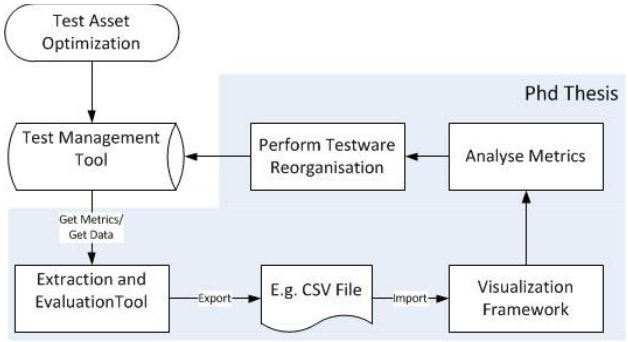


Fig. 2 - Block Structure created analysis system.

Within our research for one test project that contains over 4000 LLTC, we have performed analysis for basic and extended test metrics.

Visualization results for this test project with testware structure shown in the tables 1 and 2 are shown in the Figure 3, 4, 5 and 6. Parameters have been based on following test metrics:

1. Test execution age → mapped to the color.
2. Number of executions → mapped to the height.
3. Modification age → mapped to size.

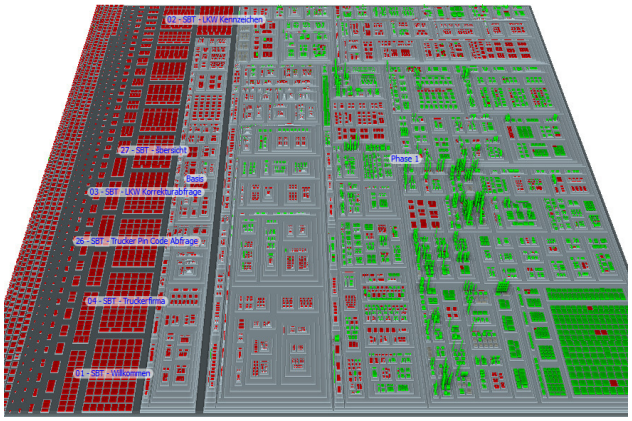


Fig. 3 Test-City based on LLTC for Test Project

To provide real reference to the analysed testware, the districts (as a square group) of the Test City are mapped to the structure created by test teams and managed with help of the Test Management system (e.g. Test folder or Test object).

Looking at the possible analysis for testware visualization according to the Figure 3 we can provide following input for the improvements:

1. There is a large number of old LLTC which has been executed later than threshold set to 370 days (e.g. red buildings – left site in the Figure 3). Most of them had a small height which gives as an information about low number of executions. Those LLTC shall be either archived, or completely removed from the Testware. LLTC not modified for longer than 1 year and rarely executed is with very high probability obsolete.

2. There are other areas in the middle and the top, which has to be taken as well under investigation (red buildings). Based on the height we can assume, most of them are obsolete; however moving to the archive is better option than leaving them within the testware.

3. Each green building is representing LLTC been most likely commonly used in the last 370 days. Large number of high and green buildings allows us to assume area of regression tests. Those LLTC has been used in the last period to assure certain quality of the product and shall not be moved to the archive or adapted within the first phase for testware reorganization.

Below, the tables shows the visualized artifacts in numbers.

Using a visualization we are able to show up hotspots within the testware domain. In order to localize objects within the testware we are focusing the interesting area with help of built in zoom function. Please see Figure 4 for an example

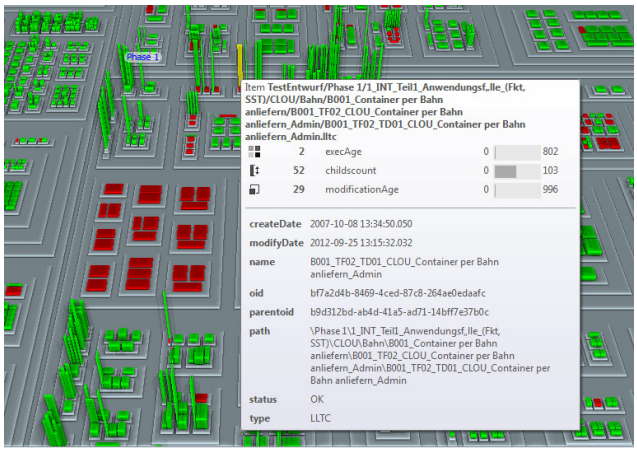


Fig. 4 Zoom on interested LLTC for Test Project

TABLE II.
TESTWARE QUANTITY FOR GIVEN TEST PROJECT

Object type	Quantity
LLTC	18473
Executions	38182

Without having a deep knowledge about the current testware and objects details we can provide the test managers with exact information regarding that LLTCs. Presented and used metrics are very basic but are giving very good start for testware reorganization and have been taken as a feedback for involved test managers.

TABLE I.
TESTWARE QUANTITY STRUCTURE

Execution Age	LLTC	(%)
0	11519	62,36
1-370	6526	35,33
>370	428	2,32

A. Reorganization results part 1

Based on the given output testware reorganization hase been performed. After finishing the first part we have executed testware domain processing and present the results. As an outcome for our work we have presented new Test-City shown in the Figure 5.

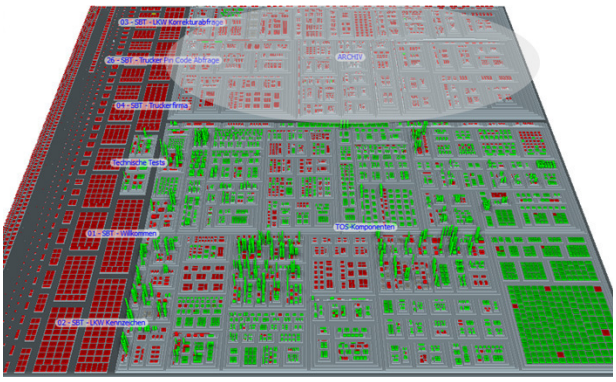


Fig. 5 Test City after first reorganization

There is a marked area where big part of obsolete LLTC has been moved. Most of the buildings are small and red colored, which shows correctness for our attempt.

Below the area we can see large district with several parcels of building which are green colored and relatively high. This allows us to assume commonly used regression tests.

Recursive execution our analysis has gave as result Test-City presented in Figure 6.

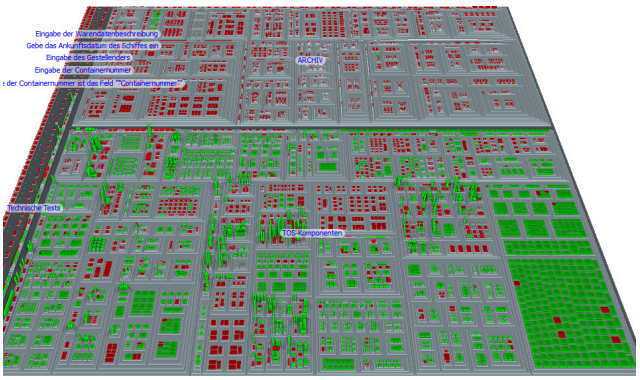


Fig. 6. Final structure for LLTCs within Test Project

There is visible well organized structure within the domain. Current analyze state has been taken as end Phase with used metrics. Deeper analysis can be performed based on other set of metrics which will allow even smaller size of testware database afterwards.

V. Feedback from Test Managers

Created results have been presented to the involved experienced Test managers within the organization and their feedback has been checked. Following results has been achieved:

→There is no false positives, all ugly layouts represents real problems
→No false negatives, no beauty layout should be ugly
→Unique global overview on the testware landscape
→Identify of hotspots ("there was always a question")
→Identify cluster of issues (e.g. regression test)
→Identify cluster of stagnation

The feedback has proven our first impression we got by looking at the testware visual representation. Even if the system looks well-organized, in spite of the numerous disharmonious artifacts: we see a districts, where the test which were executed more than 365 days ago are localized and districts of increased number of high building, even skyscrapers, in which several very important and common tests are defined.

The skyscrapers are giving us the impression how many of existing LLTC have been executed very often. Their color shows execution age as an important factor for testware reorganization.

Within very short time we were able to locate and show large number of obsolete and suspicious LLTCs. Identified hotspots and pain points based on very basic test metrics has been confirmed by the personal working for longer time with the testware, even without our deeper knowledge for the system itself. Necessary data for LLTC adaptation and/or reorganization has been exported based on zooming information at interesting areas/districts given to the test managers and used for next iteration.

Testware reorganization has been done within integration test domain and has brought minimization of used LLTC within a database. This saves in long term necessary maintenance costs and gives well overview about the current testware status.

VI. Conclusion

Test case management, test analysis and test creation are the most important tasks within the whole test management process. It is very hard to concentrate the analysis on small set of the LLTC as it is not getting potential win against the requirement spectrum. Possible loss of testware quality can be threated only as additional cost factor and each activity steering against is helping to keep those on expected level. Performed visualization has shown, how easy in use and efficient can be presented method for testware analysis. Finding an obsolete LLTC based on available metrics is very comfortable and does not require deep system knowledge, even if analyzed system seems to be very complex. This saves needed time, resources and allows problem presentation not only on technical but as well on management level. Presented results have been used for further deeper analysis and reorganization activities.

Additionally we have observed person performing analysis is tending to point its view on maximum two metrics in time and not searching for further information on the third one. This behavior was partly driven via visualization framework and its available mapping attributes and partly human laziness.

Our future directions will focus on the points listed below:

1. Extension for more APIs to Test Management tools available on the market.
2. Comparison for analysis outcome when using same metrics but different Visualization Metaphors.
3. Visualization for metrics within the timeline.
4. Extend number of evaluated metrics, especially to find out duplicate tests.

References

[1] Charters, S. M., Knight, C., Thomas, N., Munro, S., 2002: *Visualisation for informed decision making; from code to components*. In Proceedings of SEKE 2002, 765–772, ACM Press.

[2] Dickinson, W., 2001, *The Application of Cluster Filtering to operational testing of Software*. Doctoral dissertation. Case Western Reserve University.

[3] Eick, S., Graves, T., Karr, A., Marron, J., Mockus, S., 1998: *Does code decay? Assessing the evidence from change management data*. IEEE Transactions on Software Engineering 27, 1, 1–12.

[4] Friendly, M., 2008, *Milestones in the history of thematic cartography, statistical graphics, and data visualization*, http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf

[5] González, V., Kobsa, A., 2003, *Benefits of Information Visualization Systems for Administrative Data Analysts*, Proceedings. Seventh International Conference, 331-336, Information Visualization, IV 2003.

[6] Huffaker, B., Hyun, Z., Luckie, M., 2010, *IPv4 and IPv6 AS Core: Visualizing IPv4 and IPv6 Internet Topology at a Macroscopic Scale in 2010*, http://www.caida.org/research/topology/as_core_network/

[7] IEEE, 1059-1993 - *IEEE Guide for Software Verification and Validation Plans*, http://standards.ieee.org/findstds/standard/1059-1993.htm

[8] ISTQB, ISTQB® Glossary of Testing Terms, 2012, http://www.istqb.org/downloads/finish/20/101.html

[9] Knight, C., Munro, M. C. S., 2000: *Virtual but visible software.* 2000 IEEE Conference on Information Visualization, 198–205 , IEEE CS Press.

[10] Langelier, G., Sahraoui, H. A., Poulin, P. S., 2005: *Visualization-based analysis of quality for large-scale software systems.* In Proceedings of ASE 2005, 214–223, ACM Press.

[11] Lanza, M., Marinescu, R. S., 2006:. *Object-Oriented Metrics in Practice*. Springer

[12] Marinescu, R. S, 2004: *Detection strategies: Metrics-based rules for detecting design flaws.* In Proceedings of ICSM 2004, 350–359, IEEE CS Press

[13] Muller, H., and Klashinsky, S., Rigi, 1988: *a system for programming-in-the-large*. In Proceedings of ICSE 1988, 80–86, ACM Press.

[14] Panas, T., Berrigan, R., and Grundy, J. S., 2003: *A 3d metaphor for software production visualization*. IV 2003 - International Conference on Computer Visualization and Graphics Applications, 314, IEEE CS Press.

[15] Santos, C. R. D., Gros, P., Abel, P., Loisel, D., Trichaud, N., and Paris, J. P. S., 2000: *Mapping information onto 3d virtual worlds*. In Proceedings of the IV International Conference on In-formation Visualization 2000, 379–386.

[16] Stasko, J.T., Patterson, C., 1992, *Understanding and characterizing software visualization systems*, Proceedings., 1992 IEEE Workshop, 3 – 10.

[17] Wettel, R., 2010, *Software Systems as Cities*, Doctoral Dissertation, Faculty of Informatics of the Università della Svizzera Italiana

[18] Wettel, R., Lanza, M., 2008: *Visually Localizing Design Problems with Disharmony Maps*, SoftVis '08 Proceedings of the 4th ACM symposium on Software visualization, ACM Press

[19] http://www.iqnite-conferences.com/iqnite-en/about.aspx

[20] Marcus, A., Feng, L., Maletic, J., 2003, *3d representations for software visualization*. In Proceedings of SoftVis 2003, 27–36, ACM Press.
[21] IEEE 829-2008 IEEE Standard for Software and System Test Documentation, 3.1.2
[22] ISO/IEC/IEEE 24765:2010 Systems and software engineering— Vocabulary
[23] ISO/IEC 25010:2011 Systems and software engineering--Systems and software Quality Requirements and Evaluation (SQuaRE)--System and software quality models, 3.1
[24] PMI Institute*, A Guide to the Project Management Body of Knowledge (PMBOK(R) Guide) -- Fourth Edition*, 2009, ISBN: 978-1933890517

International Workshop on Cyber-Physical Systems

PROLIFERATION of computers in everyday life requires careful investigation of approaches related to the specification, design, implementation, testing, and use of modern computer systems interfacing with real world and controlling their environment. Cyber-Physical Systems (CPS) are physical and engineering systems closely integrated with their networked environment. Modern airplanes, automobiles, or medical devices are practically networks of computers. Sensors, robots, and intelligent devices are abundant. Our life depends on them. Cyber-physical systems transform how we interact with the physical world just like the Internet transformed how we interact with one another.

The event is a continuation and extension of 2006-2010 Real-Time Software FedCSIS workshops. The objective of the workshop is to assemble and develop a community with main interest in cyber-physical systems.

Topics

Due to an extensive scope of the topics, the workshop will accept papers in the following areas:

- Control Systems
  - embedded/networked/intelligent
  - wireless sensing/actuation
  - adaptive/predictive
- Scalability/Complexity
  - modularity
  - design methodology
  - legacy systems
  - tools
- Interoperability
  - concurrency
  - models of computation
  - networking
  - heterogeneity
- Validation and Verification
  - assurance
  - certification
  - simulation
- Cyber-security
  - intrusion detection
  - resilience
  - privacy
  - attack vectors
- Applications of CPS
  - robotics
  - transportation
  - military
  - medical
  - consumer
  - manufacturing
  - power systems
- CPS Education
  - curriculum development
  - web-based laboratories
  - academic courses
  - pedagogy issues

Event Chairs

Grega, Wojciech, AGH University of Science and Technology, Poland

Kornecki, Andrew J., Embry Riddle Aeronautical University, United States

Szmuc, Tomasz, AGH University of Science and Technology, Poland

Zalewski, Janusz, Florida Gulf Coast University, United States

Program Committee

Broy, Manfred, Technische Universitaet Muenchen, Germany

Caplinskas, Albertas, Vilnius Institute of Mathematics and Informatics, Lithuania

Crespo, Alfons, Universitat Politecnica de Valencia, Spain

Golatowski, Frank, University of Rostock, Germany

Gomes, Luis, Universidade Nova de Lisboa, Portugal

Halang, Wolfgang A., Fernuniversitaet, Germany

Hilburn, Thomas B., Embry Riddle Aeronautical University, United States

Kacprzyk, Janusz, Systems Research Institute PAN, Poland

Laplante, Phillip A., PennState, United States

Malec, Jacek, Lund University, Sweden

Motus, Leo, Tallinn University of Technology, Estonia

Nadjm-Tehrani, Simin, Linköping University, Sweden

Nigro, Libero, Universite della Calabria, Italy

Rozenblit, Jerzy W., University of Arizona, United States

Rysavy, Ondrei, Brno University of Technology, Czech Republic

Sanden, Bo, Colorado Technical University, United States

Schagaev, Igor, London Metropolitan University, United Kingdom

Sveda, Miroslav, Brno University of Technology, Czech Republic

Trybus, Leszek, Politechnika Rzeszowska, Poland

Vardanega, Tullio, University of Padova, Italy

Zoebel, Dieter, University Koblenz-Landau, Germany

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1361–1368

Modelling Java Concurrency: An Approach and a UPPAAL Library

Franco Cicirelli, Angelo Furfaro, Libero Nigro, Francesco Pupo
Laboratorio di Ingegneria del Software
Università della Calabria, DIMES
I-87036 Rende (CS) - Italy
Email: f.cicirelli@dimes.unical.it, a.furfaro@dimes.unical.it, l.nigro@unical.it, f.pupo@unical.it

***Abstract*—To effectively cope with correctness issues of concurrent and timed systems, the use of formal tools is mandatory. This paper proposes an original approach to modeling and exhaustive verification of Java-based concurrent systems which relies on the popular UPPAAL model checker. More precisely, a library of UPPAAL timed automata (TA) reproducing the semantics of major Java concurrent and synchronization mechanisms was developed, which fosters a smooth transition from specification down to implementation. The library includes such common control structures like semaphores and monitors, both classic and Java specific. The paper describes the developed TA library and shows its practical use by means of examples. Finally, an indication of on-going and future work directions is drawn in the conclusion.**

I. INTRODUCTION

CURRENT and prospective availability of powerful multi-core (in the CPU) and many-core (in the Graphical Processing Unit or GPU) computing architectures, and the growing acceptance of Java as a key technology for building time-dependent embedded systems, challenges software developers to the construction of concurrent programs which can greatly benefit from the high-performance computing potential of such parallel machines. Concurrent algorithm design, though, is a well-known difficult task due to human inability to check, either through peer-review or by experimental tests, the correctness of a parallel program where multiple threads of control evolve simultaneously according to complex interleaving of their actions. Race conditions, deadlocks, starvations and so forth are common risks deriving from an improper use of locks.

The work described in this paper argues that to properly design and implement concurrent and time-dependent software systems, the use of formal tools is mandatory which can enable a *reasoning* on concurrency, which is of utmost importance both in an educational or industrial context. This paper describes current status of a research project on modeling and verification (M&V) of concurrent and timed systems which was preliminarily proposed in [1]. The approach is centered on Java as the target implementation language and UPPAAL [2], [3] as a popular, mature and efficient timed automata (TA) [4] based toolbox, which makes it possible to model check complex systems [5], [6]. Although the developed concurrent structures are Java-based, they can easily be ported to other concurrent languages as well. More precisely, this paper describes current shape of a UPPAAL catalog of concurrent control structures, which was significantly improved and expanded with respect to the initial version reported in [1].

This paper contribution can be related to the solutions proposed e.g. by Hamberg & Vaandrager in [7] and to the well-known approach FSP/LTSA [8]. Our work shares with [7] the use of the UPPAAL model checker and some common semaphore and monitor control structures. However, the catalog described in this paper is original, more general and efficient, and fosters different concurrent programming styles. In addition, proposed mechanisms were mainly inspired by Java concurrency features. The FSP/LTSA approach is based on a process algebra specification of a concurrent system (FSP or Finite State Processes), automatically transformed into an equivalent Labelled Transition System (LTS) expression which is model checked in the toolbox LTSA (LTS Analyzer). A system specification must finally be implemented into Java. However, the FSP specification language does not favor the expression of FIFO based concurrent control structures (e.g. of a semaphore). Moreover, FSP/LTSA adopts a discrete time model which can complicate the verification of realistic models. A semantic gap exists between an FSP specification and a corresponding implementation in Java of a system model. Obviously, in general, an implementation cannot be proved to be a faithful concretization of a specification, but a reduction in the above semantic gap, as proposed in this work, can *help* achieving a correct implementation.

The paper is structured as follows. First basic concepts of UPPAAL are summarized. Then a running modeling example is introduced. The paper goes on by describing the developed TA catalog for modeling concurrent Java programs. Then the library is practiced through the chosen example. The discussion puts into evidence a general approach for modeling a Java thread-safe class. Finally, an indication of on-going and future work is given in the conclusion.

II. AN OVERVIEW TO UPPAAL

A system [2] is the parallel composition of multiple timed automata modeled as *template processes*, which can have parameters, can be instantiated, and consist of *atomic actions*. Parallel composition means that UPPAAL is capable of analyzing all the possible action interleavings of the component processes.

978-83-60810-53-8/$25.00

TA synchronize to one another by CSP-like channels (*rendezvous*) which carry no data values. Asynchronous communication is provided by broadcast channels where a single sender can engage in a synchronization with a (possibly empty) group of receivers. The sender of a broadcast signal in no case is blocked. Locations (states) of an automaton are linked by a set of *edges* (transitions). Time is handled by means of *clock* variables. Clocks can only be reset and compared against to a nonnegative integer constant. All the clocks of a model increase automatically at the same rate of advancement of the (hidden and dense) system time. UPPAAL extends basic TA with integer (and boolean) variables and arrays of integers, clocks and channels. Declarations can be global (shared by all the TA in a model) or local to a TA. In latest versions of the toolbox, C-like functions and structures are permitted.

Edges can be annotated by three (optional) components: (i) a *guard*, (ii) a *synchronization* action (? for input and ! for output) on a channel, and (iii) an *update* consisting of a set of clock resets and variable assignments. The update of an output command is executed before that of the matching input command.

A clock *invariant* can be attached to a location as a *progress* condition. The timed automaton can remain into the location as long as its invariant gets not violated. UPPAAL offers also *committed* and *urgent* locations which must be exited immediately (without passage of time), and *urgent channels* whose synchronizations must be fired as soon as possible. Committed locations have priority with respect to urgent locations.

UPPAAL consists of a graphical editor, a simulator and a verifier (model checker). The simulator executes a specification and visually documents the reached execution state by traversing the model state graph. The simulator is useful for model debugging and for examining a diagnostic trace (counter example) built by the verifier. For exhaustive property assessment, the verifier must be used which tries to build the reachability graph of the model, where execution states are organized into equivalence classes based on *time zones* (clock inequalities system).

Safety (e.g., absence of deadlocks) and bounded liveness (e.g. an end-to-end time constraint) properties can be verified by reachability analysis using a subset of TCTL formulas [2]. Admitted formulas (see below) refer to local state properties, i.e. boolean expressions over predicates on locations and integer variables and clock constraints.

$E <> \varphi$ means "*Possibly* φ" (a state can be reached in which φ holds).

$A[]\ \varphi$ means "*Invariantly* φ" (in all states φ holds).

$E[]\ \varphi$ means "*Potentially Always* φ" (a path exists where φ holds in all reached states).

$A <> \varphi$ means "*Always Eventually* φ" (equivalent to: $not\ E[]\ not\ \varphi$).

$\varphi - - > \psi$ means "φ *always leads to* ψ" (equivalent to: $A[]\ (\varphi\ imply\ A <> \psi)$).

III. A MODELING EXAMPLE

A classic yet representative concurrent example which can be modeled and verified using UPPAAL is the Dining-Philosophers problem (see e.g. [9], [10], [11]). N philosophers (e.g. $N = 5$), seat around a table which has a never ending big plate of spaghetti. Philosophers are equipped by a own plate and a single fork (at its left). Philosophers spend their life by thinking and, when they become hungry, try to get the two forks at its left and its right so as to take some spaghetti and then switching to eating. A thinking phase consumes from 2 to 10 time units. An eating requires from 4 to 12 time units. Forks are kept by the philosopher for the whole duration of its eating. When the philosopher finishes eating, it puts forks (hopefully after cleaning them) on the table and turns to thinking again. The availability of forks can now make some adjacent colleague get them and pass to eating as well. The problem is to ensure that the system is live (no deadlock occurs) and that it is bounded the waiting time a hungry philosopher experiments before achieving the forks (absence of starvation).

Fig. 1 shows a "native" model for philosopher i, which directly depends on the basic UPPAAL features. A global array of boolean `fork`, initialized to all true, holds the status of the fork resources. Forks relevant to the i-th philosopher have indexes i (left) e $(i + 1)\%N$ (right). Adjacent philosophers have identifiers respectively $(i + 1)\%N$ (left colleague) and $(i + (N - 1))\%N$ (right colleague).

The model is safe: a philosopher either picks both forks or none, but it is incorrect from the point of view of starvation. A hungry philosopher waits for forks in the `WAITING` location. When the fork status changes, a signal over the broadcast channel `check` is sent which allows all interested philosophers to review their status and possibly switch to the `EATING` location. On a system with N instances of the basic TA, the following queries can be issued:

```
1. A[] !deadlock                          (satisfied)
2. A[] forall(i:pid) Philosopher(i).EATING imply
   !Philosopher((i+1)%N).EATING &&
   !Philosopher((i+(N-1))%N).EATING       (satisfied)
3. Philosopher(0).THINKING-->Philosopher(0).EATING
                                          (not satisfied)
```

Query 2. confirms only not adjacent philosophers can be eating simultaneously. Native UPPAAL models tend to be concise and efficient (in space and time) for model checking. However, a native model has to be intuitively implemented e.g. in Java, relying on the reasoning on the problem solution allowed by model analysis. Ultimately, the action vocabulary of the source model (atomic actions, broadcast signals etc.) has to be transformed in the vocabulary of the target language (synchronized blocks and wait/notifyAll operations, or semaphores etc.). All of this can create problems in achieving a correct implementation. To shorten the semantic gap between modeling and implementation, source model design can be driven by implementation aspects. The following describes a library of UPPAAL TA which furnishes reusable templates for common concurrent control structures.

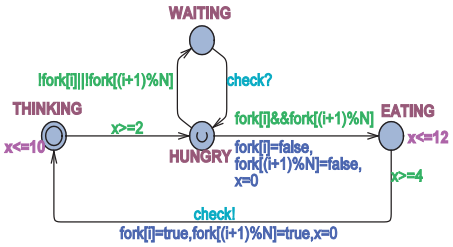


Fig. 1. A UPPAAL native model for the problem of Dining-Philosophers

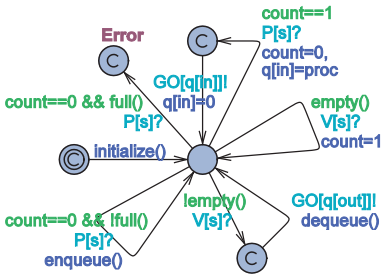


Fig. 2. BINARYSEMAPHORE automaton

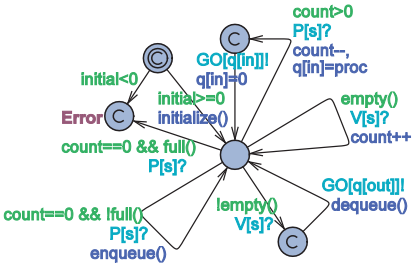


Fig. 3. SEMAPHORE automaton

IV. A TA CATALOG FOR CONCURRENT SYSTEMS

An original library of UPPAAL TA was developed which includes classic binary/counting semaphores, Java inspired semaphore, built-in monitor structure of Java, lock/condition monitor of Java, Hoare monitor, Active Oberon inspired monitor [12], exchangers, barriers etc. Other synchronizers can be added.

A. Semaphore structures

Fig. 2 and 3 respectively show a binary semaphore and a counting semaphore automata whose design tries to balance ease of use with efficient analysis.

Semaphore processes in Fig. 2 and 3 are strong in that they ensure FIFO management of waiting processes. Template parameters include the unique id of the semaphore, the initial number of permits and the expected queue size. A violation of the queue size determines the `Error` location is entered and the verification is deadlocked. Classic `P/V` operations are implemented as channel arrays `P[.]/V[.]` whose dimension mirrors the number of semaphores used in the model. A `P` operation to a semaphore `s`, is requested by a synchronization `P[s]!`. The requesting process is assumed to follow the pattern (see also Fig. 10) of putting into a global (`meta`) variable `proc` its unique process id at the time of `P[s]!`. Variable `proc` is used only during the atomic action of `P[s]!`, with the receiving semaphore which frees it immediately by storing the `proc` value in a local variable. Being a meta variable, `proc` does not contribute to the state part of the model. A further channel array `GO[.]`, whose dimension is the number of processes in the model, is used for blocking the requesting process until the semaphore assigns a permit to it. As a consequence, a `P[s]!` operation issued by process `p` should always be followed by `GO[p]?` synchronization. It is worth noting that the use of `GO` is implicit in the operation `P` in a programming language, but in UPPAAL it serves the purpose of transforming a *strict rendezvous* (`P[.]!`) into an *extended rendezvous* which terminates when the semaphore completes the handling of the `P` operation and allows the requesting process to unblock. A `V[s]!` request never blocks the requesting process and normally does not require the `proc` mediation.

With respect to the proposal in [7], our semaphores use less variables. For instance, the identity of the requesting process during a `P` operation which finds green a binary semaphore, is temporarily stored in the surely empty internal queue of the semaphore. The modeler often experiments that even by dropping one single redundant variable can avoid state explosion during the construction of the state graph, thus facilitating model checking.

Fig. 4 portrays the `JSemaphore` automaton which was inspired by the behavior of `java.util.concurrent Semaphore` class. Differences from classic semaphores concern the possibility of acquiring/releasing atomically a number of permits greater than 1. In addition, a `fair` parameter can be used to request a FIFO behavior of acquire requests. The use of `JSemaphore` rests on the channel arrays `Acquire[.]`, `Release[.]`, `PermitsAvailable[.]`, `GO[.]`, and the use of two global variables: `proc` and `perm`. The `perm` variable stores, at the time of an `Acquire[s]!` or `Release[s]!`, the number of involved permits, and contains the number of available permits of the semaphore following a `PermitsAvailable[s]!` operation. A `GO[p]?` synchronization must follow an `Acquire[s]!` or a `PermitsAvailable[s]!` command. More precisely, it is at the time of `GO[p]?` unblocking that `perm` is filled of the semaphore permits number.

It should be noted that both classic and Java specific semaphore TA are useful in practical concurrency modeling. Whereas a burst of release operations on a `JSemaphore` instance used as a mutex, will increase the permits number arbitrarily, in the case of a `BinarySemaphore` a burst of `V`'s can never augment the internal count beyond 1.

B. Monitor structures

Although widely used, semaphores are often viewed as a low level concurrent abstraction mechanism, where a misuse

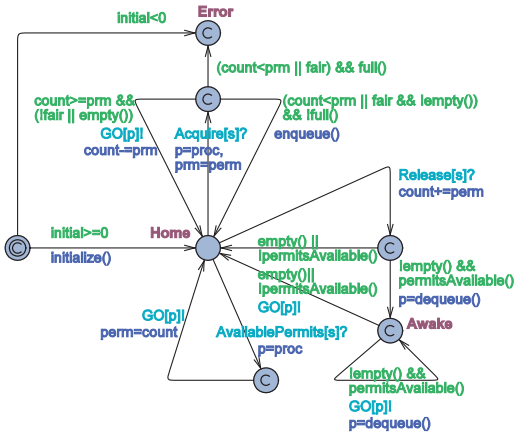


Fig. 4. JSEMAPHORE automaton

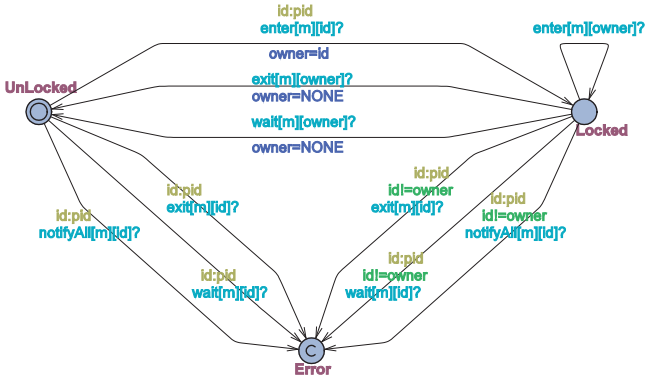


Fig. 5. JMONITOR automaton

of `P/V` operations can easily lead to a deadlock. Monitors, on the other hand, represent a higher level concurrent control structure which naturally acts as a guardian of an abstract data type, e.g. encapsulated into a Java class. Monitors are a key for achieving thread-safe classes by offering control over: mutual exclusion among methods (synchronized blocks or critical sections of code) and suspension/signaling from within a critical section. Different kinds of monitors are defined in the literature, which are characterized by different programming styles and guarantees/obligations which are assigned to both processes and the control structure.

Java adopts the Lampson&Redell [13] monitor structure with broadcast signaling, where suspended processes in a synchronized block are responsible of re-checking a condition in a while-loop to see, at each awaking, if the condition requires coming back to waiting or instead the process can go on because the condition is satisfied. Broadcast signaling is not blocking for the signaler process. An awaken process has to compete in reacquiring the lock for it to actually resume execution.

The Hoare monitor (e.g. [9], page 234) has a different signaling mechanism: when a process (*signaler*) changes the status of the data structure so that a (possibly) waiting process (*signalee*) on a condition can be awaken because the condition holds, control is immediately transferred to the signalee (together with the lock) which is thus the only process which can then proceed. The signaler, on the other hand, is put to wait on an urgent queue from where it gets unblocked as soon as the monitor is up to become free.

A discussion about Lampson&Redell vs. Hoare monitors can be found in [9] at page 240 where it is argued, besides any runtime implication and number of context switches, that Lampson&Redell monitor can be superior in the most general case.

An example of a monitor which facilitates the developer by transferring responsibilities from the programming level to the control structure, was adopted in the Active Oberon language [12]. Here the programmer has only to deal with the logic of conditions which, as long as they do not hold, prescribe a process has to wait. Signaling and process awaking is hidden in the control structure.

In the following, a series of developed monitor TA is presented.

Fig. 5 depicts the `JMonitor` automaton which allows to model concurrent objects according to the Java built-in monitor. A monitor instance can be operated using such channel arrays as `enter[mid][pid]`, `exit[mid][pid]`, `wait[mid][pid]`, `notifyAll[mid][pid]` which accommodate for the possible existence of multiple monitor instances in a model. Types `mid` and `pid` respectively are integer sub-ranges of unique identifiers for monitors and processes used in the model. For instance, `enter[m][p]!/exit[m][p]!` are used by a process `p` to explicitly enter/exit to/from a synchronized block based on monitor identifier `m`. Similarly, `wait[m][p]!/notifyAll[m][p]!` serve respectively to suspend the requesting process p until its condition holds (in a while loop), and to awake all the processes suspended on monitor `m`.

Every Java object owns a lock which can be used as a monitor. The lock holds one implicit condition, whose meaning is only known to the modeler/programmer. The lock object is associated with a *wait-set* where both entering processes which find the lock closed, or processes within a synchronized block (based on the lock object) but whose condition prescribes waiting, are put (although the two kind of waiting processes are clearly distinguished to one another) and suspended. Processes which are suspended for a `wait` operation can only be awaken by a `notifyAll` operation which does not free the lock. Other processes awake as the lock/monitor is up to be abandoned (at an `exit` or `wait` operation). In the proposed implementation, the wait-set is purposely realized implicitly. Processes requesting `enter` simply are blocked if the monitor is already locked.

Processes which execute `wait` are supposed to move into a location (see `WAITING` in Fig. 11) from which they can only exit following a relevant `notifyAll` signal. Towards

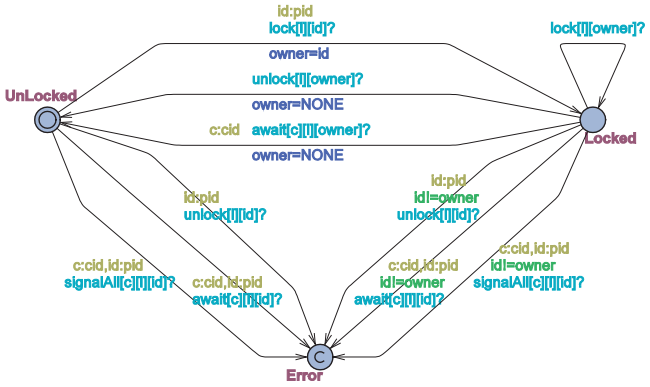


Fig. 6. LOCK automaton

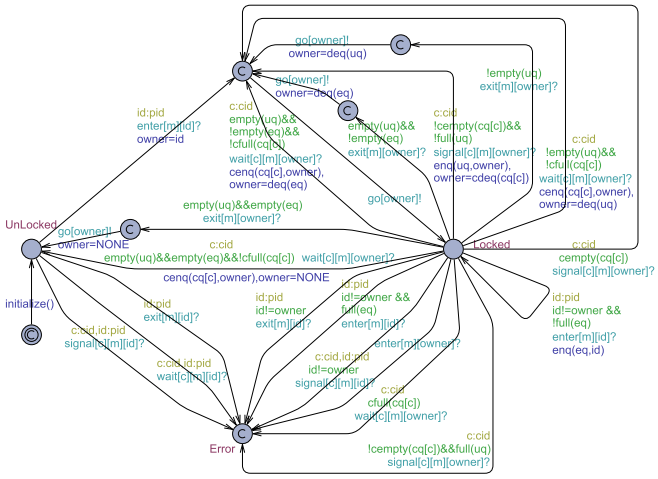


Fig. 7. HOAREMONITOR template

this, channels `notifyAll[.][.]` are declared as broadcast channels.

The automaton in Fig. 5 maintains the identity of the monitor `owner`, which is used both to realize reentrancy and to check for erroneous operations, which in Java correspond to raising an `IllegalMonitorStateException`.

The implicit realization of the wait-set complies with the Java specification and lets processes which try to enter the monitor and awaken processes to be handled non deterministically and thus without any privilege. The design pattern also requires that an awaken process from a wait location has to explicitly compete in reacquiring the lock (this operation is hidden in the Java `wait()` method of class `Object`). The design pattern makes it possible to implement also a *timed wait*. In this case, from the wait location (now provided of a clock invariant) the process can also exit when the clock goes beyond a given time limit (*timeout*), thus competing for the lock before checking the condition.

In reality the Java built-in monitor also offers a `notify` operation to awake *one* unspecific process suspended in the wait-set. For generality reasons the automaton in Fig. 5 only implements the `notifyAll` (broadcast) operation because, as discussed e.g. in [14] at pages 181-183, the use of `notify` can cause what is known as the `Lost-Wakeup-Problem`.

Since Java 5, the `java.util.concurrent` package also provides a refinement of the built-in monitor through the lock/condition control structure. In this version, it is possible to introduce both a lock object and a certain number of condition objects linked to the lock object. As a consequence, processes can be suspended on the different conditions and the signaling mechanism can be directed to all the processes waiting on a certain condition.

Fig. 6 depicts the Lock automaton which realizes the locking mechanism (and its reentrancy) and also handles the relevant conditions.

Monitor operations are captured by two dimensional `lock[lid][pid]`, `unlock[lid][pid]` array channels, whose first dimension is related to the sub-range of lock unique identifiers used in a model, and whose second dimension is tied to the process unique identifiers, and three dimensional `await[cid][lid][pid]`, `signalAll[cid][lid][pid]` array channels where the first dimension consists of the unique condition identifiers of a given lock identifier.

As in the case of the `JMonitor` automaton, the enter wait-set and condition wait-set are realized implicitly, as well as `signalAll[cid][lid][pid]` channels are declared as broadcast channels.

The same conventions discussed for `JMonitor` apply here: a waiting process on a condition `c` is supposed to wait into a suitable location of the automaton, from where the process can exit following a `signalAll` or a timeout. It then has to compete for reacquiring the lock and is in charge of re-checking the relevant condition.

As it is common, the Hoare monitor can be achieved on top of semaphores, in particular binary semaphores. If N are the conditions of the monitor, $N + 2$ semaphores are to be used: one as mutex, another for the urgent mechanism of signalers and N for conditions. In Fig. 7, a slightly different but equivalent and more efficient automaton implementation is proposed which rests on $N + 2$ queues. The monitor can be used through the matrices of channels `enter[mid][pid]`, `exit[mid][pid]`, `wait[cid][mid][pid]`, `signal[cid][mid][pid]`, `go[pid]` where `mid`, `cid` and `pid` are respectively the integer sub-ranges of monitor unique identifiers, relevant condition unique identifiers, unique process identifiers. The pattern of use does not necessarily depend on the while-loop required by built-in Java monitor. A process waiting on a condition is, in general, guaranteed that the condition holds when it is signaled. The monitor is assumed to be not reentrant. As a rule, a synchronization on the `go[.]?` channel must follow each invocation (!) of `enter`, `exit`, `wait` or `signal` operation.

A simplification of the Hoare monitor is provided by the Active Oberon monitor (see Fig. 8) where the signaling operation is removed. The burden (and the risks) of proceeding by an awaken process whose condition cannot possibly hold,

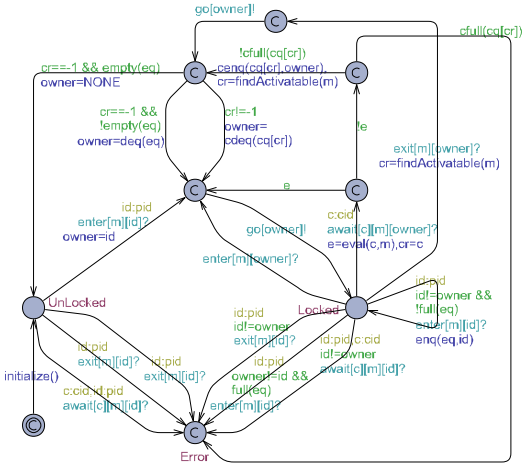


Fig. 8. AOMONITOR template

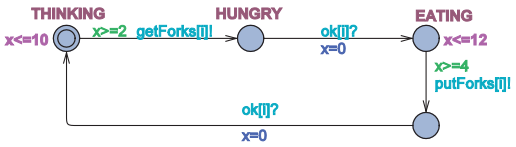


Fig. 9. The PHILOSOPHER automaton

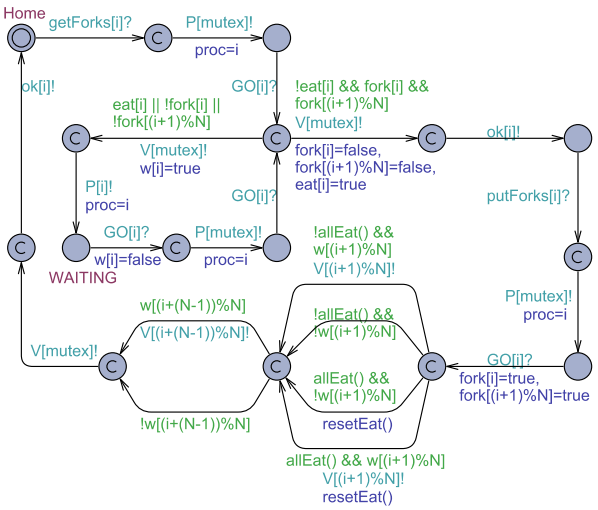


Fig. 10. Manager automaton based on semaphores

are eliminated by the control structure which manages an implicit signaling and the transfer of the lock to an awaken process. Therefore, the resultant modeling/programming style becomes more concise with respect to the Hoare monitor.

The `AOMonitor` automaton (Fig. 8) is supposed to be reentrant. Its use depends on the matrices of channels `enter[mid][pid]`, `exit[mid][pid]`, `await[cid][mid][pid]`, `go[pid]`. A `go[.]?` synchronization is required after each invocation (!) of `enter`, `exit` or `await` operation. Any waiting room is realized as a FIFO queue.

The modeler has to introduce a global `bool eval(cid,mid)` function, model specific, which receives a condition id and its monitor id and returns true if the logical boolean expression which is associated with the given condition of the given monitor, holds; otherwise `eval()` returns false.

The `findActivatable(mid)` function used in Fig. 8 scans the list of conditions of the given monitor and returns, if there is one, the identifier of the first condition which is found satisfied; the function returns -1 if the search fails. The result of `findActivatable` is used to transfer the control (together with the monitor lock) to the oldest process waiting on the given condition.

In order to avoid starvation, the control structure maintains the last index of success on the list of conditions so as to start the next lookup from the next position and cyclically.

V. PUTTING THE LIBRARY INTO ACTION

Usefulness of the developed library was assessed through several examples. In the following, the use of the library is demonstrated by applying it to the Dining-Philosophers problem. Modularity issues suggest separating the application processes from the details of concurrency control which in Java are embedded into a thread-safe class. Therefore it is convenient to organize the Philosopher model as shown in Fig. 9 and introducing a `Manager` model which exposes a suitable interface to philosopher processes and hides the synchronization constraints. In particular, the channel arrays `getForks[pid]`, `putForks[pid]` e `ok[pid]` are assumed to define the `Manager` interface. Since each philosopher can block in the manager model, N identical instances of the `Manager` are created, each one corresponding to a distinct philosopher. All the manager instances, though, share global data e.g. the boolean array `fork[.]` about free/occupied status of forks. Both `Philosopher` and `Manager` have one single parameter i (of type `pid`) which furnishes the identity of the philosopher. Following a `getForks[i]!` or `putForks[i]!` operation, the philosopher expects an `ok[i]?` synchronization confirming that the requested operation was carried out. Again, in a Java implementation the `ok` signal is redundant because the extended rendezvous is automatically provided by `getForks/putForks` methods of the `Manager` class.

A. Manager based on semaphores

In Fig. 10 is depicted a `Manager` automaton which uses $N+1$ binary semaphores.

One semaphore is used for mutual exclusion (`mutex`, initialized to 1). The remaining N semaphores, one per philosopher, are waiting rooms or conditions (always kept to 0). Condition identifiers coincide with philosopher identifiers. The model in Fig. 10 uses two further boolean arrays: `w[pid]` and `eat[pid]`. The former serves to know if a

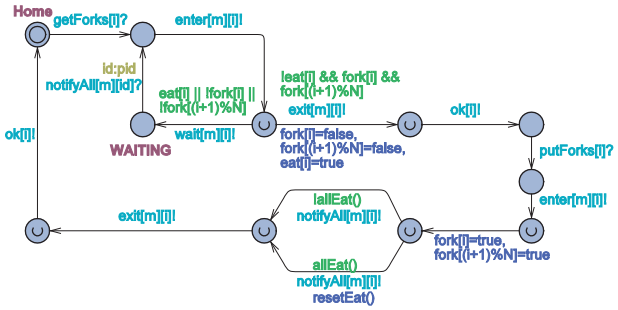


Fig. 11. Manager automaton based on Java monitor

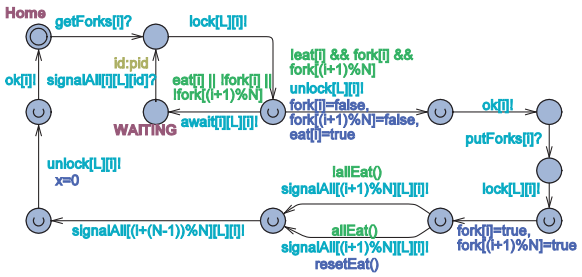


Fig. 12. Manager automaton based on lock/condition monitor

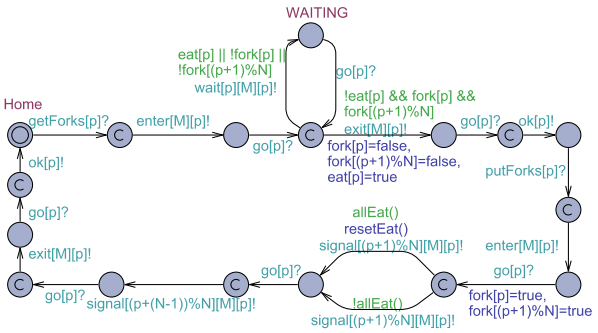


Fig. 13. Manager based on the Hoare monitor

given philosopher is waiting for forks. The latter is used for implementing a simple strategy for avoiding starvation. Philosophers are supposed to eat at turns. A turn finishes when all philosophers have eaten. A `getForks` request is not responded either because some fork is unavailable or the philosopher has already eaten in the current turn. The latest philosopher who eats resets the array `eat` so as to start a new turn.

All the three queries suggested in section III, are now satisfied. Query 3 concerning proving absence of starvation deserves some further comments. It is a liveness property. UPPAAL is most apt to verify safety and bounded liveness properties. General liveness can be difficult to assess. In a normal location, in fact, an automaton can stay an arbitrary amount of time. To help checking liveness properties, Urgent/Committed locations or urgent channels should be used. In Fig. 10, the use of committed locations was preferred.

A bounded liveness property for the model of Fig. 9 and Fig. 10 concerns the worst case amount of time a philosopher stays in the `WAITING` location of its manager, waiting for the other colleagues to complete the current turn. Using $N = 5$ philosophers, and adding a decoration clock y to the `Manager` model, which is reset at each `getForks[i]?` request, the following query was issued to the UPPAAL verifier:

```
A[] Manager(0).WAITING imply Manager(0).y<=64
```

The query was found satisfied, but changing the upper bound to 63 the query no longer holds. This result corresponds to a $(10-(2+4))*(N-1)$ remaining thinking time for the other partners, and then a $12*(N-1)$ worst case eating time of remaining colleagues.

As a final remark, except for the `GO[pid]` and `ok[pid]` channels which are required only in the UPPAAL models, the automata in Figures 9 and 10 can directly be expressed in Java code.

B. Manager based on JMonitor

Using the built-in Java monitor, a `Manager` model like that shown in Fig. 11 can be achieved. With respect to the semaphore based solution, it requires less space and time for the analysis.

A similar model to that in Fig. 11 was built using the lock/condition monitor, which is slightly more efficient due to reduced partial order following a signaling operation. The new model is portrayed in Fig. 12.

It should be noted that since enter/lock requests can be delayed by the monitor being already locked, such operations should not exit from urgent locations. In addition, because of broadcast signaling, there is no need to pay for the `w[pid]` boolean array used in the semaphore based solution.

C. Manager based on Hoare monitor

It was interesting achieving a Hoare monitor based model for manager, to compare expressiveness and guarantees during signaling with Java built-in or lock/condition based versions. The model is portrayed in Fig. 13. As expected, this particular example does not allow to exploit the normal guarantees of the Hoare monitor: i.e. that a signaled process waiting on a condition is sure its condition holds when awaken by a signal. In fact, putting forks is only a partial fulfillment for the precondition of adjacent philosophers to be able to get forks. As a consequence, without going to put much burden on the signaler processes, the right solution consists in envisioning the while-loop also in a signalee so as to come back to waiting if the precondition for awaking does not actually hold. Without the while-loop, UPPAAL confirms the model is incorrect.

Model of Fig. 13 is more expensive in terms of space and time for verification with respect to models in Fig. 11 and Fig. 12. This is due to the use of queues for conditions and related bookkeeping data.

D. Manager based on Active Oberon monitor

This version of the `Manager` model is illustrated in Fig. 14.

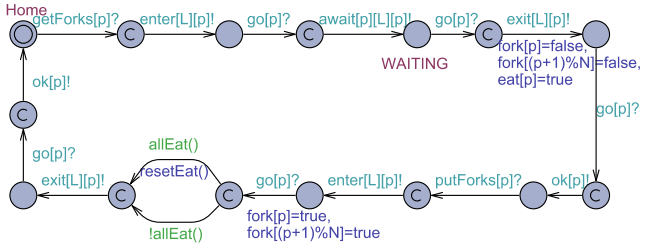


Fig. 14. Manager automaton based on the Active Oberon monitor

The following is the `eval()` function which was prepared for working with the Dining-Philosophers problem:

```
bool eval( cid c, mid m ){
  if( !eat[c]&&fork[c]&&fork[(c+1)%N] )
    return true;
  return false;
}//eval
```

The solution in Fig. 14 is the most concise and easy to follow from the modeler/programmer point of view. Its analysis performance, though, is similar to the Hoare monitor version because the model has almost the same data requirements.

Verification experiments with $N = 5$ philosophers were carried out on a Win 8, 12GB, Intel Core i7-3770K, 3.50GHz. To figure out efficiency of the various models, the query which checks for the absence of deadlocks lasts in the worst case in about 25sec with a RAM peak of about 100MB.

VI. Conclusion

The UPPAAL library of timed automata (TA) proposed in this paper is effective for modeling and verification (M&V) of Java-based concurrent and timed programs. It includes both semaphores and monitor control structures. The Java built-in monitor or its refinement based on lock/condition are often preferable both from the M&V perspective and the implementation viewpoint. The Active Oberon monitor offers the most concise level for modeling and implementing a thread-safe class.

Being not primitive in Java, Hoare monitor and Active Oberon monitor classes were achieved on top of semaphores. By the way, a `BinarySemaphore` class was also realized which is weakly bisimilar to the automaton in Fig. 2.

A nondeterministic point e.g. in the Hoare monitor class accompanies the implementation of the `wait` operation which in general must free the lock and put the requesting process to sleep. Whereas UPPAAL atomic actions hide the problem (e.g. through a committed location), in a concrete implementation the alea point could be handled by ensuring that before relinquishing the lock the wait operation starts a new time-slice. This could be achieved by using the Thread `yield()` method. However (see discussion in [15] at page 287) the `yield()` method can often behave as a no operation. A better provision could be a `Thread.sleep(1)` if 1 millisecond is the time resolution of the underlying operating system.

The library is currently in use in an undergraduate course on systems programming and the response of students is positive.

A major benefit of the catalog and of the UPPAAL model checker rests on the possibility of favoring a *reasoning on concurrency*.

On-going and future work are geared to:

- Optimizing the library so as to improve the efficiency of model checking activities.
- Extending the library with other concurrency control structures, e.g. based on the concept of software transactional memory [14] which delivers a different and attractive style of concurrent programming.
- Extending the approach based on the UPPAAL model checker to M&V of lock-free concurrent objects [14] which are often perceived as a grand challenge for an exploitation, in parallel and embedded software systems, of the computing potential of current and future multi-core/many core machines.

Acknowledgment

Authors are grateful to Christian Nigro for his contribution during the design and realization of the UPPAAL library proposed in this paper and its support in Java.

References

[1] F. Cicirelli, L. Nigro, and F. Pupo, "Modelling and verification of concurrent programs using UPPAAL," in *Proc. of 25th European Conference on Modelling and Simulation*, Krakow, Poland, 2011, pp. 525–533.
[2] R. Alur and D. Dill, "A theory of timed automata," *Theoretical Computer Science*, vol. 126, no. 2, pp. 183–235, 1994.
[3] G. Behrmann, A. David, and K. Larsen, "A tutorial on UPPAAL," in *Formal Methods for the Design of Real-Time Systems*, ser. LNCS 3185, M. Bernardo and F. Corradini, Eds. Springer, 2004, pp. 200–236.
[4] "UPPAAL, on-line," www.Uppaal.org.
[5] F. Cicirelli, A. Furfaro, and L. Nigro, "Model checking time-dependent system specifications using Time Stream Petri Nets and UPPAAL," *Applied mathematics and computation*, vol. 218, no. 16, pp. 8160–8186, 2012.
[6] F. Cicirelli, A. Furfaro, L. Nigro, and F. Pupo, "Development of a schedulability analysis framework based on PTPN and UPPAAL with Stopwatches," in *Proc. of the IEEE/ACM 16th International Symposium on Distributed Simulation and Real Time Applications (DS-RT'12)*, Dublin, Ireland, 2012, pp. 57–64.
[7] R. Hamberg and F. Vaandrager, "Using model checkers in an introductory course on operating systems," *SIGOPS Oper. Syst. Rev.*, vol. 42, no. 6, pp. 101–111, 2008.
[8] J. Magee and J. Kramer, *Concurrency: state models & Java programs*. John Wiley & Sons, Ltd., 2006.
[9] W. Stallings, *Operating Systems: Internals and Design Principles*. Upper Saddle River, NJ, USA: Prentice Hall Press, 2005.
[10] A. Silberschatz, P. Galvin, and G. Gagne, *Operating System Concepts*, 8th ed. Wiley Publishing, 2008.
[11] A. S. Tanenbaum, *Modern Operating Systems*. Upper Saddle River, NJ, USA: Prentice Hall Press, 2001.
[12] P. Reali, "Active oberon language report," http://bluebottle.ethz.ch/languagereport/index.html, 2002.
[13] B. Lampson and D. Redell, "Experience with processes and monitors in Mesa," in *Proc. of the seventh ACM symposium on Operating systems principles*, Pacific Grove, California, USA, 1979, pp. 43–44.
[14] M. Herlihy and N. Shavit, *The Art of Multiprocessor Programming*, revised version of first ed. Elsevier Science Limited, 2012.
[15] B. Joshua, *Effective Java*, 2nd ed. Addison Wesley, 2008.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1369–1376

Synthesis of Implementable Control Strategies for Lazy Linear Hybrid Automata

Luigi Di Guglielmo
Department of Computer Science
University of Verona, Italy

Sanjit A. Seshia
Department of EECS
University of California, Berkeley, USA

Tiziano Villa
Department of Computer Science
University of Verona, Italy

***Abstract*—In the last few years hybrid automata have been widely applied in the modeling and verification of hybrid systems, but their related formal verification techniques usually rely on un-implementable assumptions to which a concrete control strategy cannot adhere. For this reason, once a hybrid model of the system has been proved to be correct with respect to the desired properties, it would be valuable to derive a correct-by-construction implementable control strategy for such a model. This work discusses a new methodology and a corresponding tool-chain that allows to synthesize an implementable control strategy for the class of hybrid automata named Lazy Linear Hybrid Automata (LLHA). LLHA model the discrete time behavior of control systems containing finite-precision sensors and actuators interacting with their environment under bounded delays.**

I. Introduction

HYBRID systems are dynamical systems whose behaviors cannot be characterized faithfully using either discrete or continuous models. They consist of a discrete part that operates in a continuous environment, and for this reason, they are sensitive not only to time-driven phenomena but also to event-driven ones. The presence of mixed dynamics makes the formal treatment of this kind of systems considerably hard.

Hybrid automata (HA) [1] are a powerful formalism for modeling hybrid systems. It extends the usual definition of finite state automata with continuous variables that evolve according to dynamics characterizing each discrete state. In the last few years, a wide spectrum of algorithmic techniques has been studied to solve the problems of simulation and verification for hybrid automata. Current state-of-art tools can verify hybrid systems with complex nonlinear dynamics [2], [3], or linear systems with a large number of continuous variables [4], thus becoming of interest also for real test-cases and application domains.

Another phase of the design flow where the interplay of continuous and discrete behaviors makes things complicated is the refinement and implementation phase. Indeed, formal verification techniques for hybrid automata usually rely on un-implementable assumptions, such as the *synchrony hypothesis*, i.e., the capability of performing any computation in zero time units and forcing a change in the dynamics of the hybrid system with no delays. As a consequence, the verification results on the correctness of the ideal model of the system cannot be directly applied to a real implementation [5].

For this reason, new semantics for hybrid automata, which do no rely on synchrony or other unrealistic assumptions, have been proposed in the literature [6], [7]. By formally verifying the correctness of the system using such semantics, it is possible to synthesize an implementable control strategy for the analyzed hybrid system, i.e., determine the *performance and latency bounds* to be satisfied by any conservative concrete hardware/software device that implements the system.

In [6], [8], the authors propose the Almost-ASAP semantics, a formal semantics that imposes a controller to react within a bounded delay, i.e., Δ, when a synchronization or a control action has to take place. The authors use reachability analysis [1] to look for the largest value Δ for which the controller is still correct w.r.t. the properties that the original instantaneous model has to enforce. Such a Δ-relaxed controller represents an implementable control strategy if $\Delta > 0$. The main limitation of this approach is that the problem of synthesizing such a value Δ may not be decidable.

The work in [7] proposes a similar approach for synthesizing implementable control strategies. The authors try to derive an implementable control strategy from the original instantaneous model by shrinking the guards of the latter so that, by assuming bounded reaction delays for synchronization and control actions, all behaviors of the former are non-blocking (i.e., shrinkability problem). This means that all timing requirements satisfied by the instantaneous control strategy, such as critical deadlines, are strictly respected by the derived one. The authors have shown that deciding the shrinkability problem can be checked in EXPTIME.

This work focuses on the problem of automating the synthesis of implementable control strategies for a relevant class of hybrid automata, named *Lazy Linear Hybrid Automata* (LLHA). Such a kind of automata takes into account all the typical implementation aspects (e.g., discrete time behaviors, finite precision of sensors and clock, sensing and actuation delays), thus, once they have been proved correct, they provide an implementable control strategy for the hybrid system.

The paper is organized as follows. Section II introduces the fundamental definitions and semantics of LLHA which motivate the assumptions at the base of the proposed methodology for synthesis of implementable control strategies described in Section III. Section IV describes some case studies to which the methodology has been applied. Finally, Section V is devoted to concluding remarks.

978-83-60810-53-8/$25.00

II. BACKGROUND

In the following the class of LLHA is described. A LLHA is meant to be a model of a closed-loop system consisting of a digital controller interacting with a continuous environment [9]. The controller samples the state of the continuous environment at periodic discrete-time instants. The state of the environment consists of the values of the continuous variables as observed by the sensors. These values are digitized with finite precision and reported to the controller that may decide to switch the state of the environment. In such a case, the controller generates suitable output signals that, once transmitted to the actuators, will effect the desired change. Sensors will report the values of the current variables and actuators will change the evolution of the continuous variables with bounded delays.

A. The LLHA formal definition

Definition 1 (Lazy Linear Hybrid Automaton)**.** A finite precision Lazy Linear Hybrid Automaton (LLHA) is a tuple $\langle X, Q, init, inv, flow, E, jump, Act, P, D, \epsilon, B\rangle$. The components of a LLHA are as follows:

- *Variables*. A finite set $X = \{x_1, \ldots, x_n\}$ of real-valued variables. $\dot{X}$ stands for the set $\{\dot{x_1}, \ldots, \dot{x_n}\}$ of dotted variables and X' stands for the set $\{x'_1, \ldots, x'_n\}$ of primed variables.
- *Control modes*. A finite set Q of control modes. $Q_0 \subseteq Q$ denotes the set of initial modes.
- *Initial condition*. A labeling function $init$ that assigns to each control mode $q \in Q_0$ an initial predicate. The initial predicate $init(q)$ is a convex (non-)linear formula over the variables in X.
- *Invariant condition*. A labeling function inv that assigns to each control mode $q \in Q$ an invariant predicate. The invariant predicate $inv(q)$ is a convex (non-)linear formula over the variables in X.
- *Flow condition*. A labeling function $flow$ that assigns to each control mode $q \in Q$ a flow predicate. For each $i \in \{1, \ldots, n\}$, let $\dot{X}^i_q \subset \mathbb{Q}$ be the set of legal flow rates for the variable x_i in the control mode q. The flow predicate $flow(q)$ is of the form $(\dot{x}_1 \in \dot{X}^1_q) \wedge \ldots \wedge (\dot{x}_n \in \dot{X}^n_q)$.
- *Control switches*. A set E of edges (q, q') from a source mode $q \in Q$ to a target mode $q' \in Q$.
- *Jump condition*. A labeling function $jump$ that assigns to each control switch $e \in E$ a predicate. Each jump predicate $jump(e)$ from the control mode q to q', is given by the conjunction of a guard and a reset condition. The *guard* is given by a convex (non-)linear formula over the variables in X. The *reset* condition is given by the identity predicate over the variables in $X \cup X'$ (e.g., $x'_i = x_i$).
- *Actions*. A finite set Act of actions that the automaton uses either for internal synchronization or for synchronizing with other communicating automata. An edge labeling function $action : E \to Act$ assigns an action to each control switch.
- *Period*. P represents the sampling interval of the controller, i.e., control mode switches take place at times T_0, T_1, T_2, ... where $T_{k+1} = T_k + P$.
- *Delay parameters*. $D = \{g, \delta_g, h, \delta_h\} \subset \mathbb{Q}$ is the set of delay parameters such that $0 \leq g \leq g + \delta_g < h \leq h + \delta_h \leq P$, where g denotes the actuation delay, h denotes the sensing delay and δ_g, δ_h represent the uncertainty in actuation and sensing delay, respectively.
- *Precision*. ϵ_i is the precision of measurement of variable x_i.
- *Range*. $B_i = [B_{i_{min}}, B_{i_{max}}] \subset \mathbb{R}$ is the allowed range of the variable x_i such that $B_{i_{min}}$, $B_{i_{max}} \in \mathbb{Q}$ and $B_{i_{min}} < B_{i_{max}}$.

To keep the notation compact, in what follows, $q \xrightarrow{a,\varphi} q'$ is used to denote that there exists a control switch $e = (q, q')$ with $q \neq q'$, $a = action(e)$ and $\varphi = jump(e)$ in A.

Unlike the conventional definition of linear hybrid automata [10], invariants and guards in LLHA can be non-linear (i.e., polynomial). The flows in linear hybrid automata are represented using rectangular formulas which denote closed intervals of the form $[l, r] \subset \mathbb{R}$ with $l, r \in \mathbb{Q}$ and $l < r$. Under the assumption of finite precision, in a LLHA such rectangular formulas denote finite sets of rational values modeling the rate of change of the different continuous variables.

B. The LLHA formal semantics

Let A be a lazy linear hybrid automaton as defined above. The following definitions are required to specify the behavior of A in terms of a transition relation.

Definition 2 (Valuation for continuous variables)**.** Let $X = \{x_1, \ldots, x_n\}$ be a set of continuous variables. A valuation V for the variables in X is a member of $\mathbb{R}^n$ such that V assigns a real value $V(i)$ to each variable x_i.

Definition 3 (State of a LLHA)**.** A *state* of a lazy linear hybrid automaton A is a triple $(q, V, \hat{q})$ where q, $\hat{q}$ are control modes and V is a valuation. q is the control mode holding at the current time instant and $\hat{q}$ is the control mode that held at the previous time instant. V captures the actual values of the variables at the current instant. The state $(q, V, \hat{q})$ is feasible if and only if $V(i) \in [B_{min_i}, B_{max_i}]$ for every i.

Intuitively, the state of a LLHA stores information about current and previous control modes (i.e., q and $\hat{q}$, respectively) due to the fact that, as a result of a mode change, the change of rates of continuous variables will occur with bounded delays. As a consequence, the evolution of continuous variables in a mode q will depend not only on the flow predicates of q, but also on the flow predicate of the previous control mode $\hat{q}$.

The initial state is, by convention, the triple $(q_{init}, V_{init}, q_{init})$. It is assumed without loss of generality that the initial state is feasible. Let S_A denote the set of states of A.

For convenience, in what follows, it is assumed that the rate of change of continuous variables is constant in each control mode. Thus, each flow predicate $flow(q)$ can be described

by vector $\rho_q \in \mathbb{Q}^n$ that specifies the rate $\rho_q(i)$ at which each variable x_i evolves when the automaton is in the control mode q.

Definition 4 (Transition relation of a LLHA)**.** $\Longrightarrow \subseteq S_A \times (Act \cup \{\tau\}) \times S_A$ is such that:

- Let $(q, V, \hat{q})$, $(q', V', \hat{q}') \in S_A$ be states and $a \in Act$. Then $(q, V, \hat{q}) \overset{a}{\Longrightarrow} (q', V', \hat{q}')$ if and only if $\hat{q}' = q$ and there exist a control switch of the form $q \overset{a,\varphi}{\longrightarrow} q'$ in A and $t_1 \in \mathbb{Q}^n$, $t_2 \in \mathbb{Q}^n$ such that $\forall i \in [1, n]$, $t_1(i) \in [g, g+\delta_g]$, $t_2(i) \in [h, h+\delta_h]$ and the following conditions are satisfied:
  1) Let $v_i = V(i) + \rho_{\hat{q}}(i) \cdot t_1(i) + \rho_q(i) \cdot (t_2(i) - t_1(i))$ for each i. Then $(\langle v_1 \rangle, \ldots \langle v_n \rangle)$ satisfies φ and each $\langle v_i \rangle$ represents the digitized value of the variable x_i that has been rounded using the value of ϵ_i.
  2) $V'(i) = V(i) + \rho_{\hat{q}}(i) \cdot t_1(i) + \rho_q(i) \cdot (P - t_1(i))$ for each i.
- Let $(q, V, \hat{q})$, $(q', V', \hat{q}') \in S_A$ be states. Then $(q, V, \hat{q}) \overset{\tau}{\Longrightarrow} (q', V', \hat{q}')$ if and only if $q' = \hat{q}' = q$ and there exists $t_1 \in \mathbb{Q}^n$ and $\forall i \in [1, n]$, $t_1(i) \in [g, g + \delta_g]$ such that:
  1) $V'(i) = V(i) + \rho_{\hat{q}}(i) \cdot t_1(i) + \rho_q(i) \cdot (P - t_1(i))$ for each i.

The lazy semantics of linear hybrid automata means that if a control mode switch took place at time T_k, then the delay in actuating a change in flow rates lies between $[T_k + g, T_k + g + \delta_g]$. Similarly, a control decision made at time T_k is based on the variables values read by the controller at some time in the interval $[T_{k-1} + h, T_{k-1} + h + \delta_h]$. The parameters δ_g and δ_h represent the bounded uncertainty in actuation and sensing delay, respectively. The precision ϵ_i depends on the accuracy of the sensors measuring x_i from the continuous dynamical system. Guards and state invariants are evaluated on the digitized values $\langle x_i \rangle$ of the variables x_i that have been rounded using the value of ϵ_i. The parameter B, instead, reflects the range of values which can be taken by a state variable associated with a fixed width register.

From the semantics defined above, it is possible to derive the notions of trajectory of a LLHA and the reachability relation between states.

Definition 5 (Trajectory of a LLHA)**.** Let A be a lazy linear hybrid automaton and let $(q, V, \hat{q})$ be a state of A. A trajectory of A from $(q, V, \hat{q})$ is a sequence of states $(q_i, V_i, \hat{q}_i)$ with $i > 0$, such that $(q_0, V_0, \hat{q}_0) = (q, V, \hat{q})$ and $(q_{i-1}, V_{i-1}, \hat{q}_{i-1}) \overset{\alpha}{\Longrightarrow} (q_i, V_i, \hat{q}_i)$ for some $\alpha \in Act \cup \{\tau\}$.

Example 1. Figure 1(i) sketches a lazy linear hybrid automaton A with two control modes q_1 and q_2. Let i and j be such that $i, j \in \{1, 2\}$ and $i \neq j$. Each invariant condition $inv(q_i)$ is defined as a subset I_i of $\mathbb{R}$ and the LLHA can stay in the control mode q_i if the valuation of the variable x satisfies the invariant condition. The jump condition of a control switch $e_{ij} = (q_i, q_j)$ is specified by a guard set G_{ij} and a reset function that, by definition of LLHA, is always

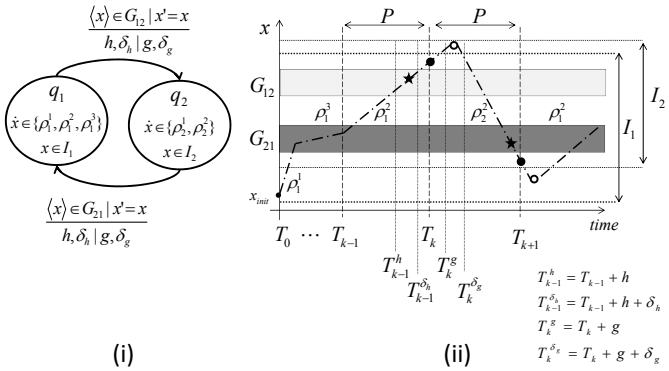


Fig. 1. Lazy linear hybrid automaton example.

the identity function. The control switch e_{ij} is enabled only if the digitized value $\langle x \rangle$ detected by the sensor belongs to G_{ij}. Moreover, sensing and actuation delays (i.e., h, δ_h and g, δ_g, respectively) are associated to the control switch. Finally, each flow condition $flow(q_i)$ constrains the evolution of the continuous variable x to one of the possible rates ρ_i^n allowed in the mode (e.g., $\{\rho_1^1, \rho_1^2, \rho_1^3\}$ in q_1).

Figure 1(ii) sketches part of a trajectory of such a LLHA starting from the initial state (q_1, x_{init}). In the example, the trajectory keeps following the dynamics $flow(q_1)$ until the time instant T_k. In fact, the control switch e_{12} is not enabled as soon as the trajectory reaches the guard set G_{12} because of the semantics of LLHA: a jump condition can be evaluated only at periodic time points and by considering the digitized values detected by the sensor at some instant (marked with $\star$) in the interval $[T_{k-1}^h, T_{k-1}^{\delta_h}]$. As shown in the figure, at T_k, the invariant condition of the mode q_1 is still satisfied, thus, the LLHA can either switch to q_2 or continue with the dynamics of q_1. Let assume that the automaton performs a control switch (marked with $\bullet$) and moves to q_2. When the LLHA switches from q_1 to q_2, it resets the continuous variable x according to the predicate specified by the jump condition, i.e., the identity function. Thus, in this case, the trajectory starts from the same state reached at T_k. Notice that the trajectory keeps following the dynamics of q_1 due to the presence of an actuation delay (i.e., g, δ_g) on the control switch. In fact, only at some time (marked with $\circ$) in the interval $[T_k^g, T_k^{\delta_g}]$ the trajectory changes according to the rates specified by the flow condition of q_2 (i.e., $\dot{x} \in \{\rho_2^1, \rho_2^2\}$). Then the trajectory follows that flow rate until the invariant I_2 is violated or the jump condition G_{21} is satisfied allowing the automaton to jump back in the mode q_1.

Definition 6 (Reachability relation between states of a LLHA)**.** Let A be a lazy linear hybrid automaton. A state $(q, V, \hat{q})$ reaches a state $(q', V', \hat{q}')$ if there exists a finite trajectory of states $(q_i, V_i, \hat{q}_i)$, with $0 \leq i \leq n$, such that $(q_0, V_0, \hat{q}_0) = (q, V, \hat{q})$ and $(q_n, V_n, \hat{q}_n) = (q', V', \hat{q}')$. $\mathcal{RC}(q, V, \hat{q})$ is used to denote the set of states reachable from $(q, V, \hat{q})$. $\mathcal{RC}$ is used to denote the set of all the possible states reachable from the initial ones.

III. Synthesis of Implementable Control Strategies for LLHA

The main contributions of this work can be summarized as follows:

- it proposes a Bounded Model Checking (BMC) [11] formulation for the problem of synthesizing implementable control strategies for LLHA that reduces such a problem to the state reachability problem on LLHA. A previous BMC formulation has been given in [12]. While that work assumes that precision and delay parameters are given, the present paper models them as parameters that must be synthesized. Then, by verifying the safety properties as reachability queries, it is possible to identify values for such parameters which make the control strategy implementable, i.e., the control strategy is able to handle the continuous plant by following discrete-time and finite-precision behaviors.
- it proposes a synthesis procedure that, starting from a set of feasible values for the different parameters, identifies for each of them the maximum values which enable a LLHA to satisfy its required safety properties.

The following sections describe all the details of the proposed approach.

A. Problem definition

The synthesis of an implementable control strategy for a LLHA consists of determining if there exist legal values for the sampling period (i.e., P), and upper bounds for sensing and actuation delays (i.e., $T_{SD} = h + \delta_h$ and $T_{AD} = g + \delta_g$, respectively) for which the control strategy modeled in the LLHA is able to satisfy the safety properties that the hybrid system has to ensure.

Let A be a LLHA such that $\mathcal{S}_A$ is the set of the possible states, $\mathcal{INIT}$ be a predicate that constrains the initial state, $\mathcal{TR}$ be the transition relation that models the lazy behavior of A and φ_{safe} be a function that tests whether the safety properties for the hybrid system hold in a given state. The synthesis problem summarized above can be formalized by a Quantified Boolean Formula (QBF), i.e., a formula in which propositional variables can be either quantified existentially or universally, as follows:

$$\begin{aligned}&\exists P, T_{SD}, T_{AD}, \forall\, n \in \mathbb{N}, \forall S_i \in \mathcal{S}_A : \mathcal{INIT}(S_0) \wedge \\ &\bigwedge_{i=0}^{n} \mathcal{TR}(S_i, S_{i+1}, P, T_{SD}, T_{AD}) \rightarrow \bigwedge_{i=0}^{n} \varphi_{safe}(S_i)\end{aligned} \tag{1}$$

Intuitively, the formula states that there exist suitable values for P, T_{SD} and T_{AD} for which at any step i, the state S_{i+1}, reachable from a previous state S_i, satisfies the safety property φ_{safe}.

An efficient way to solve this problem consists of deriving from Formula (1) a BMC problem on A. Such a BMC problem focuses on identifying the existence of *bad states*, i.e., states S_i violating the safety properties and reachable from the initial state of A:

$$\begin{aligned}&\mathcal{BMC}(A, \varphi_{safe}, n, P, T_{SD}, T_{AD}) \equiv \mathcal{INIT}(S_0) \wedge \\ &\bigwedge_{i=0}^{n} \mathcal{TR}(S_i, S_{i+1}, P, T_{SD}, T_{AD}) \wedge \bigvee_{i=0}^{n} \neg\varphi_{safe}(S_i)\end{aligned} \tag{2}$$

The identification of suitable values for the parameters n, P, T_{SD} and T_{AD} which cause the unsatisfiability of Formula (2), will prove the validity of Formula (1) w.r.t. the chosen P, T_{SD} and T_{AD}. Notice that, given the values for n, P, T_{SD} and T_{AD}, the satisfiability of Formula (2) may be proved or disproved by applying a Satisfiability Modulo Theory (SMT) [13] decision procedure on its propositional part.

Notice that the transition relation $\mathcal{TR}$ in the $\mathcal{BMC}$ formula may be unrolled a finite number n of times, where n is the *reachability diameter* [14] of the LLHA, i.e., the minimal number of steps for reaching all its reachable states. Thus, the formula checks if a bad state $S_{i \leq n}$ is reachable from the initial state S_0. Unfortunately, such a number n may require a very high number of copies of the transition relation in the $\mathcal{BMC}$ formula making the verification unfeasible due to memory problems.

Due to lack of space, the symbolic BMC encoding will be described in an extended version of the paper. In what follows a synthesis procedure is proposed for identifying the maximum suitable values of sampling period (P), sensing and actuation delays (T_{SD},T_{AD} respectively) to let the LLHA A satisfy the safety specification φ_{safe}.

B. Synthesis procedure

The definition of the BMC formula described in the previous section is based on a set of parameters whose values affect the correctness of the LLHA model. The synthesis engine aims at identifying the maximum values of such parameters for which the discrete-time and finite-precision behaviors specified by the LLHA are able to satisfy φ_{safe}.

In particular, the parameters reported into the formula $\mathcal{BMC}$ are the following:

- *sampling period* P. It specifies the periodicity at which it is possible to evaluate the guards for performing a mode switch;
- *sensing delay upper-bound* T_{SD}. It specifies the maximum delay admitted for notifying the controller that a mode switch can be performed (i.e., sensor latency);
- *actuation delay upper-bound* T_{AD}. It specifies the maximum delay admitted for changing the rate due to a mode switch (i.e., actuator latency).

At the moment, the precision ϵ of the observed values is not explicitly modeled as a parameter. Instead, it is assumed that it is fixed at some suitable level of granularity and that the constant values reported in the predicates that model guard and invariant conditions have been scaled accordingly to be represented as integers[1].

[1]Remember that the underlying structure used for the symbolic representation of variables is the bit-vector.

For reducing the time required in identifying the suitable values for the parameters summarized above, the user is asked to specify a desired sampling period P that the control strategy has to adopt. Then a synthesis procedure will automatically retrieve the maximum values for T_{SD} and T_{AD} that preserve the safety of the model according to the specified sampling period.

The procedure identifies the intended values by using a bisection method on a finite interval of feasible values for the parameters. According to the LLHA semantics, the actuation delay has to be smaller than the sensing delay and the sensing delay has to be smaller than the sampling period. Thus, it is necessary to search the suitable values of T_{AD} and T_{SD} in some intervals $\mathcal{I}_1 = [a, b]$ and $\mathcal{I}_2 = [c, d]$ such that $b < d$ and $b + d < P$. This is due to the fact that, in the BMC encoding, the upper bound T_{AD} for the actuation delay is considered as the most distant time instant from a sampling period T_i at which a control switch has occurred and, as a consequence, its starting search space should be given by the interval $[0, b]$. Similarly, in the BMC encoding, the upper-bound T_{SD} for the sensing delay is considered as the most distant time instant from a sampling period T_j, subsequent to T_i, at which a control decision may be taken. Thus, the starting search space for T_{SD} should be given by the interval $\mathcal{I}_2 = [0, d]$.

Algorithm 1 reports the pseudo-code implementing the synthesis procedure for parameters, and $BMC^n(P, T_{AD}, T_{SD})$ denotes the BMC encoding of the transition relation of the LLHA A unrolled n times (n is the reachability diameter).

At each step, the procedure divides the current subintervals of $[a, b]$ and $[c, d]$ in two by computing their midpoints mid_1 and mid_2. Then, using a SMT solver (i.e., $\mathcal{SMT}$), it verifies whether the formula $BMC^n(P, mid_1, mid_2)$ is valid by fixing $T_{AD} = mid_1$ and $T_{SD} = mid_2$. Now, according to the verification results, the method registers the current midpoints as candidate solutions and selects the subintervals to be used in the next step. In particular, if mid_1 and mid_2 make the formula valid, they become *candidate* maximal values for T_{AD} and T_{SD}, resp., and the new intervals of search will be $[mid_1, b]$ and $[mid_2, d]$, i.e., the procedure will check the validity of the formula on new values greater than the current midpoints. Otherwise, the procedure has to look for smaller ones. At first, it checks the formula validity by reducing only the current value for actuation delays. It computes the new candidate for T_{AD} (i.e., mid_{new}) and, by preserving the previous candidate mid_2 for T_{SD}, verifies the validity of the formula $BMC^n(P, mid_{new}, mid_2)$. If the verification returns a positive answer, then mid_{new} and mid_2 are recorded as new candidate maximal solutions for T_{AD} and T_{SD}, resp., and the new intervals of search will be $T_{AD} \in [mid_{new}, mid_1]$ and $T_{SD} \in [mid_2, d]$. On the contrary, the non-validity of the formula underlines that also a smaller sensing delay is required. For this reason the search continues on the intervals $[a, mid_1]$ and $[c, mid_2]$. In this way the intervals that contain the satisfying values of the parameters are reduced in width at least by 50% at each step. The process is continued until the maximum number N of iterations is reached.

Algorithm 1: The synthesis procedure of parameters for LLHA-based control strategies.

procedure find_values(BMC^n, P, a, b, c, d, N)
input: the BMC^n formula, the sampling period P, initial intervals [a,b] and [c,d] of feasible values for T_{AD} and T_{SD}, resp., and the maximum number N of iterations
output: maximal values of T_{AD} and T_{SD} for which the control strategy satisfies φ_{safe}, otherwise $T_{SD} = 0$ and $T_{AD} = 0$

```
1  it = 0;
2  T_SD = 0;
3  T_AD = 0;
4  mid_1 = ⌊(a + b)/2⌋;
5  mid_2 = ⌊(c + d)/2⌋;
6  while (it < N) do
7      if SMT(BMC^n(P, mid_1, mid_2)) --> valid then
8          a = mid_1;
9          T_AD = mid_1;
10         c = mid_2;
11         T_SD = mid_2;
12     else
13         b = mid_1;
14         mid_new = ⌊(a + b)/2⌋;
15         if SMT(BMC^n(P, mid_new, mid_2)) --> valid then
16             a = mid_new;
17             T_AD = mid_new;
18             c = mid_2;
19             T_SD = mid_2;
20         else
21             d = mid_2;
22     mid_1 = ⌊(a + b)/2⌋;
23     mid_2 = ⌊(c + d)/2⌋;
24     it = it + 1;
25 return (T_SD, T_AD);
```

IV. Experimental Results

This section reports the results obtained by applying the proposed LLHA parameter synthesis approach on four case studies. All experiments have been performed on a workstation with Intel Xeon 2.53 GHz processors and 16GB RAM. The hybrid models of the case studies have been described by means of the CIF [15] language. The *cif2uclid* tool has been implemented to automatically derive, from such models, the LLHA descriptions and the corresponding BMC encodings which have been automatically synthesized into equivalent SMT formulas by using the UCLID [16] modeling environment. Several SMT solvers have been used to verify the models and identify the maximum values for the sensing and actuation delay parameters appearing in the LLHA models. In particular, for each case-study the performances of the following SMT solvers have been compared: Beaver [17], Boolector[2] [18], and Yices [19]. Notice that any available SMT solver could be used as verification and parameter synthesis engine.

[2]MiniSat and PicoSat have been used as the underlying SAT engines.

A. Train-Gate Controller

The train gate controller ensures that the gate is closed when the train is approaching it. The train is assumed to move at a constant speed v on a circular track of length $d_{far-away}$ and the gate begins to close at a constant angular speed u when the train is at d_{max} distance from the gate. Once the train has moved d_{max} distance away from the gate, the gate begins to open again. The system is shown in Figure 2. The distance d of the train is measured in meters, the angle a of the gate in degrees and the time in seconds. The set of parameter values used in the running example is as follows: $v = 20m/s$, $u = 10°/s$, $d_{far-away} = 20000m$, $d_{max} = 400m$ and $d_{safe} = 160m$.

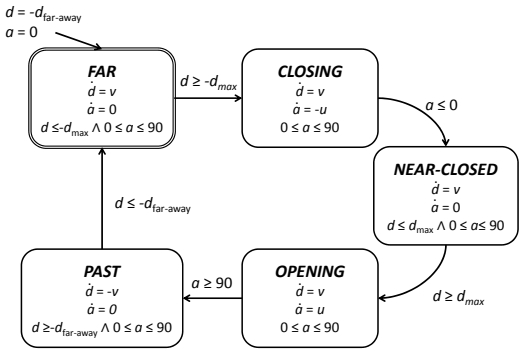


Fig. 2. LLHA model of the Train-Gate controller.

The system is considered safe, i.e., the train is never closer to the gate than d_{safe} unless the gate is completely closed, only if the following safety property is satisfied: $-d_{safe} \leq d \leq d_{safe} \rightarrow a <= 0$.

Such a property is used during the synthesis phase for identifying the maximum values for the sensing and actuation delay parameters (T_{SD} and T_{AD}, respectively) that are reported into the LLHA modeling the system. In particular, the parametric LLHA has been automatically generated by using a digitizing precision $\epsilon = 10^{-3}$ and a control switch period $P = 10^{-2}s$. Then, the parameter synthesis approach has determined that the coarse values for the sensing and actuation delays which ensure the correctness of such a LLHA are $T_{SD} = 4 \cdot 10^{-3}s$ and $T_{AD} = 2 \cdot 10^{-3}s$. The time required for synthesizing such values is reported in Table I.

TABLE I
SYNTHESIS TIMES USING DIFFERENT SMT SOLVERS.

$\mathcal{SMT}$	T_{SD} **s-Space**	T_{AD} **s-Space**	**# Bisect.**	**Time (s)**
Beaver	$[0; 5 \cdot 10^{-3}]$	$[0; 4 \cdot 10^{-3}]$	15	123.336
Boolector	$[0; 5 \cdot 10^{-3}]$	$[0; 4 \cdot 10^{-3}]$	15	101.544
Yices	$[0; 5 \cdot 10^{-3}]$	$[0; 4 \cdot 10^{-3}]$	15	49121.94

In particular, column $\mathcal{SMT}$ reports the name of the compared SMT solvers; columns T_{SD} *s-Space* and T_{AD} *s-Space* report the initial search spaces used for identifying suitable values for the sensing and actuation delays, respectively. Column *# Bisect.* shows the maximum number of bisection iterations allowed for synthesizing the parameter values and, finally, column *Time* reports the total time (in seconds) spent for the synthesis process.

B. Room Heating Controller

The room heating controller ensures that the temperature of a room is kept into a comfort interval by turning *on* and *off* the heater installed into the room. Figure 3 depicts the model of the system. Intuitively, the automaton is composed by two control modes *on* and *off*, representing the status of the heater. The variable x denotes the room temperature (measured in $°C$). When the heater is *off*, the temperature of the room falls according to any rate specified by the rectangular constraint $\dot{x} \in [-b, -a]$. Instead, when the heater is *on* the temperature of the room rises following any rate specified by the constraint $\dot{x} \in [b, c]$. The heater is turned on as soon as the falling temperature reaches x_{low}: the automaton moves to the control mode *on* and the temperature rises starting at a value $x \leq x_{low}$. The heater is turned off as soon as the temperature reaches x_{high}: the automaton moves to the control mode *off* and the temperature starts falling again at a value $x \geq x_{high}$. This control strategy guarantees that the temperature of the room will remain between x_{min} and x_{max} starting at the initial temperature x_{init} such that $x_{low} < x_{init} < x_{high}$. The set of parameter values used in the running example is as follows: $a = 2 \cdot 10^{-1}$ $°C/s$, $b = 3 \cdot 10^{-1}$ $°C/s$, $c = 4 \cdot 10^{-1}$ $°C/s$, $x_{min} = 17.8°C$, $x_{low} = 18°C$, $x_{high} = 22°C$, $x_{max} = 22.5°C$ and $x_{init} = 19°C$.

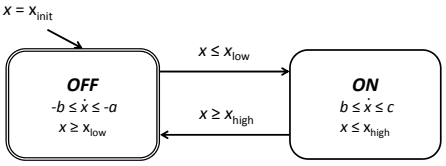


Fig. 3. The LLHA model of the room heating controller.

The property $x_{min} < x < x_{max}$ is used for synthesizing the maximum values of the sensing delay T_{SD} and the actuation delay T_{AD} in the parametric LLHA modeling the finite-precision lazy heating controller. Such a model has been generated from the one depicted in Figure 3 by choosing a digitizing precision $\epsilon = 10^{-4}$ and a control switch period $P = 10^{-2}s$. The synthesis procedure found that the values $T_{SD} = 11 \cdot 8^{-3}s$ and $T_{AD} = 10 \cdot 10^{-3}s$ guarantee that the lazy controller keeps the temperature into the comfort bounds (i.e., x_{min} and x_{max}). The time required for synthesizing such values is reported in Table II.

C. Watertank Controller

The watertank system is centered on a water tank, which is characterized by an uncontrolled outbound water flow, while the inbound water flow is controlled by the aperture of a valve. The controller acts on the aperture of the valve y in order to keep the water level x in a safe interval $x_{min} < x <$

TABLE II
SYNTHESIS TIMES USING DIFFERENT SMT SOLVERS.

$\mathcal{SMT}$	T_{SD} s-Space	T_{AD} s-Space	# Bisect.	Time (s)
Beaver	$[0; 4 \cdot 10^{-2}]$	$[0; 3 \cdot 10^{-2}]$	15	343.888
Boolector	$[0; 4 \cdot 10^{-2}]$	$[0; 3 \cdot 10^{-2}]$	15	1584.864
Yices	$[0; 4 \cdot 10^{-2}]$	$[0; 3 \cdot 10^{-2}]$	15	90903.836

TABLE III
SYNTHESIS TIMES USING DIFFERENT SMT SOLVERS.

$\mathcal{SMT}$	T_{SD} s-Space	T_{AD} s-Space	# Bisect.	Time (s)
Beaver	$[0; 5 \cdot 10^{-2}]$	$[0; 4 \cdot 10^{-2}]$	15	531.56
Boolector	$[0; 5 \cdot 10^{-2}]$	$[0; 4 \cdot 10^{-2}]$	15	2413.02
Yices	$[0; 5 \cdot 10^{-2}]$	$[0; 4 \cdot 10^{-2}]$	15	107236.98

x_{max}. The system model is shown in Figure 4. The water level is measured in deciliters, the valve aperture in degrees and the time in seconds. The set of parameter values used in the running example is as follows: $a = 1dl$, $b = 2dl$, $c = 4dl$, $d = 7dl$, $v = 20°/s$, $x_{high} = 850dl$, $x_{low} = 550dl$, $y_{min} = 0°$, $y_{max} = 360°$, $x_{max} = 870dl$ and $x_{min} = 540dl$.

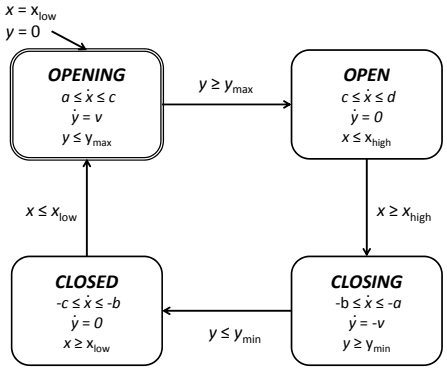


Fig. 4. LLHA model of the watertank controller.

In the model, the water level begins to increase according to a rectangular constraint $\dot{x} = [a, c]$ when the valve starts to open at constant angular speed v. As soon as the valve reaches its full aperture, the incoming flow reaches its maximum value, filling faster the water tank. Once the water level crosses an upper threshold x_{high}, the valve starts to close in order to avoid a water overflow. Once the valve is completely closed, no inbound water flow is present and the water level keeps decreasing. When the water level reaches its lower threshold x_{low}, the valve begins to open again.

The property $x_{min} < x < x_{max}$ is used for synthesizing the maximum values of the sensing delay T_{SD} and the actuation delay T_{AD} in the parametric LLHA modeling the finite-precision lazy watertank controller. Such a model has been generated from the one shown in Figure 4 by choosing a digitizing precision $\epsilon = 10^{-3}$ and a control switch period $P = 10^{-1}s$. At the end of the synthesis phase, the verification has determined that the values $T_{SD} = 23 \cdot 10^{-3}s$ and $T_{AD} = 11 \cdot 10^{-3}s$ guarantee that the lazy controller keeps safely the water level into the x_{min} and x_{max} bounds. The time required for synthesizing such values is reported in Table III.

D. Automated Highway Control System

Automated Highway Control System (AHS) is an arbiter which ensures that there is no collision between cars running on a highway by imposing legal speed ranges. The linear hybrid automaton representing the case of four cars is shown in Figure 5. This example is a small variant of the original model reported in [20]. The distance between cars is measured in km, time in hours and speeds in km/h. The set of parameter values used in the running example is as follows: $rl = 20km/h$, $b = 30km/h$, $c = 40km/h$, $d = 50km/h$, $e = 60km/h$, $ru = 70km/h$, $f = 100km/h$, $\alpha_{min} = 2 \cdot 10^{-3}km$, $\alpha_{max} = 1km$ and $\alpha'_{min} = 5 \cdot 10^{-4}km$.

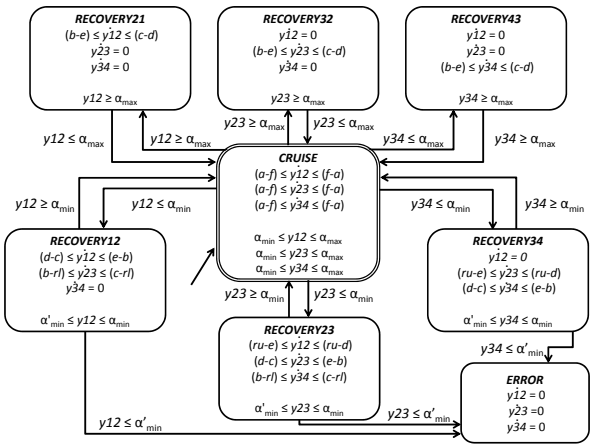


Fig. 5. LLHA model of the Automated Highway Control System.

To avoid collisions, the arbiter specifies speed limits (i.e., $[a, f]$) for each vehicle. When two vehicles i and j come within a distance $y_{ij} \leq \alpha_{min}$ of each other, there exists a possible collision event. The arbiter asks the approaching car to slow down by reducing the speed into the interval $[b, c]$, and asks the leading car to speed up by keeping a speed into the interval $[d, e]$; it also requires that all other cars not involved in the possible collision slow down to a constant recovery mode velocity rl for cars behind the critical region and ru for cars in front of the critical region. When the distance between the two vehicles involved in the possible collision exceeds α, the arbiter model goes back to the dynamics of the cruise mode. Moreover, the arbiter keeps all the vehicles below a maximal distance α_{max} of each other. When two vehicles i and j exceed such a distance (i.e., $y_{ij} \geq \alpha_{max}$), the arbiter asks the leading car to slow down by reducing the speed into the interval $[b, c]$ and asks the approaching car to speed up by keeping a speed into the interval $[d, e]$; it also requires that all other cars keep the current distance constant (i.e., speeding up for cars behind the critical region and slowing down for cars in front of the critical region). When the distance between the

two vehicles decreases below α_{max}, the arbiter model goes back to the dynamics of the cruise mode.

The only safety property to be satisfied by the model is that the control mode is never the *error* mode.

Again, once the parametric LLHA model has been extracted by choosing a digitizing precision $\epsilon = 10^{-5}$ and a clock period $P = 10^{-2}$, the safety property is used as a constraint for identifying the coarse values of the sensing and actuation delay parameters (T_{SD} and T_{AD}, respectively). In this case study, the synthesis phase determined that the values $T_{SD} = 218 \cdot 10^{-5}h$ and $T_{AD} = 112 \cdot 10^{-5}h$ guarantee the safety of the system, i.e., the lazy controller is able to avoid cars' collision. The time required for synthesizing such values is reported in Table IV.

TABLE IV
SYNTHESIS TIMES USING DIFFERENT SMT SOLVERS.

$\mathcal{SMT}$	T_{SD} **s-Space**	T_{AD} **s-Space**	**# Bisect.**	**Time (s)**
Beaver	$[0, 5 \cdot 10^{-3}]$	$[0, 4 \cdot 10^{-3}]$	15	1472.75
Boolector	$[0, 5 \cdot 10^{-3}]$	$[0, 4 \cdot 10^{-3}]$	15	1844.05
Yices	$[0, 5 \cdot 10^{-3}]$	$[0, 4 \cdot 10^{-3}]$	15	188523.11

V. CONCLUSION

The development of methodologies for the synthesis of implementable control strategies for models based on hybrid automata is a new and valuable research area. This work focused on defining a new methodology which enables the synthesis of implementable control strategies for the interesting subclass of lazy linear hybrid automata. To support the methodology, a tool, i.e., *cif2uclid*, and a synthesis procedure were implemented in order to provide a complete toolchain for synthesizing the implementable control strategy in a systematic way. *cif2uclid* is able to extract from a hybrid model a corresponding parametric-LLHA description, that is automatically synthesized into an equivalent SMT formula by using the UCLID modeling environment. The verification of such a formula retrieves constraints for the parameters which guarantee that the control strategy is implementable, i.e., the verification retrieves the performance and latency bounds which make the control strategy realizable by a concrete hardware/software device. The synthesis procedure may use any available SMT solver.

The proposed procedure for the automatic synthesis of parameters that guarantee implementability of control strategies is supported by a complete toolchain and improves the state-of-art with respect to previous proposals, however scalability of the overall approach is still a serious issue, when the objective is to study test cases of industrial strength. This may require from one side further restrictions to the class of allowed hybrid automata, still preserving enough expressivity for practical purposes, and from the other side advances in the computational engines and how they are used (for instance, gaining efficiency by using incrementally the SMT solvers).

REFERENCES

[1] T. Henzinger, "The Theory of Hybrid Automata," in *Proc. of IEEE Symposium on Logic in Computer Science (LICS)*, pp. 278 – 292, 1996.

[2] S. Ratschan and Z. She, "Safety Verification of Hybrid Systems by Constraint Propagation Based Abstraction Refinement," *ACM Transactions in Embedded Computing Systems*, vol. 6, no. 1, 2007.

[3] L. Benvenuti, D. Bresolin, P. Collins, A. Ferrari, L. Geretti, and T. Villa, "Assumeguarantee verification of nonlinear hybrid systems withAriadne," *International Journal of Robust and Nonlinear Control*, p. doi: 10.1002/rnc.2914, 2012.

[4] G. Frehse, C. Le Guernic, A. Donzé, S. Cotton, R. Ray, O. Lebeltel, R. Ripado, A. Girard, T. Dang, and O. Maler, "SpaceEx: Scalable Verification of Hybrid Systems," in *Proc. of International Conference on Computer Aided Verification (CAV)*, pp. 379–395, 2011.

[5] D. Bresolin, L. D. Guglielmo, L. Geretti, R. Muradore, P. Fiorini, and T. Villa, "Open Problems in Verification and Refinement of Autonomous Robotic Systems," in *Proceedings of the 15th EUROMICRO Conference on Digital System Design (DSD 2012)*, pp. 469–476, 2012.

[6] M. Wulf, L. Doyen, and J. Raskin, "Almost ASAP Semantics: from Timed Models to Timed Implementations," *Formal Aspects of Computing*, vol. 17, no. 3, pp. 319 – 341, 2005.

[7] O. Sankur, P. Bouyer, and N. Markey, "Shrinking Timed Automata," in *Proc. of IARCS Annual Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pp. 90–102, 2011.

[8] D. Bresolin, L. D. Guglielmo, L. Geretti, and T. Villa, "Correct-by-Construction Code Generation from Hybrid Automata Specification," in *Proceedings of First IEEE Workshop on Design, Modeling and Evaluation of Cyber Physical Systems (CyPhy'11)*, pp. 1660–1665, 2011.

[9] M. Agrawal and P. Thiagarajan, "The Discrete Time Behavior of Lazy Linear Hybrid Automata," in *Proc. of International Conference on Hybrid Systems: Computation and Control (HSCC)*, pp. 55–69, 2005.

[10] T. Henzinger and P. Kopke, "Discrete-Time Control For Rectangular Hybrid Automata," in *Proc. of International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 582–593, Springer, 1997.

[11] A. Biere, A. Cimatti, E. Clarke, O. Strichman, and Y. Zhu, "Bounded Model Checking," *Advances in Computers*, vol. 58, pp. 117–148, 2003.

[12] S. Jha, B. A. Brady, and S. A. Seshia, "Symbolic Reachability Analysis of Lazy Linear Hybrid Automata," in *Proc. of International Conference on Formal Modeling and Analysis of Timed Systems (FORMATS)*, pp. 241–256, 2007.

[13] C. Barrett, R. Sebastiani, S. A. Seshia, and C. Tinelli, "Satisfiability Modulo Theories," in *Handbook of Satisfiability* (A. Biere, H. van Maaren, and T. Walsh, eds.), vol. 4, ch. 8, IOS Press, 2009.

[14] D. Kroening and O. Strichman, "Efficient Computation of Recurrence Diameters," in *Proc. of International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI)*, pp. 298–309, 2003.

[15] C. Sonntag, R. Schiffelers, D. van Beek, J. Rooda, and S. Engell, "Modeling and Simulation using the Compositional Interchange Format for Hybrid Systems," in *International Conference on Mathematical Modelling (MATHMOD)*, pp. 640–650, 2009.

[16] R. E. Bryant, S. K. Lahiri, and S. A. Seshia, "Modeling and Verifying Systems using a Logic of Counter Arithmetic with Lambda Expressions and Uninterpreted Functions," in *Proc. of International Conference on Computer Aided Verification (CAV)*, pp. 78–92, 2002.

[17] S. Jha, R. Limaye, and S. Seshia, "Beaver: Engineering An Efficient SMT Solver for Bit-Vector Arithmetic," in *Proc. of International Conference on Computer Aided Verification (CAV)*, pp. 668–674, 2009.

[18] R. Brummayer and A. Biere, "Boolector: An Efficient SMT Solver for Bit-Vectors and Arrays," *Tools and Algorithms for the Construction and Analysis of Systems*, pp. 174–177, 2009.

[19] B. Dutertre and L. De Moura, "The Yices SMT Solver." http://yices.csl.sri.com/tool-paper.pdf, 2006.

[20] S. K. Jha, B. H. Krogh, J. E. Weimer, and E. M. Clarke, "Reachability for Linear Hybrid Automata Using Iterative Relaxation Abstraction," in *Proc. of International Conference on Hybrid Systems: Computation and Control (HSCC)*, pp. 287–300, 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1377–1380

Towards deductive-based support for software development processes

Radosław Klimek
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
rklimek@agh.edu.pl

***Abstract*—The work relates two initial disciplines of the Rational Unified Process (RUP), i.e. Business Modeling and Requirements Engineering, to support them in an integrated way through deductive-based formal verification using temporal logic. On the other hand, Cyber-Physical Systems (CPS), which should be an effective orchestration of computations and physical processes, need careful development and formal verification to ensure they influence software reliability, trustworthiness and cost in a positive way. A method for building both business models and requirements models, including their logical specifications, is proposed and presented step by step. Applying the presented concepts bridges the gap between the benefits of deductive reasoning for correctness analysis and the difficulties in obtaining complete logical specifications.**

I. Introduction

CYBER-Physical Systems (CPS) are understood as integrations of computation with physical processes [1] and often refer to embedded systems. A CPS is designed as a network of interacting elements with physical input and output and focus on both technology and mathematical abstractions. CPS need to improve the development processes, in order to raise the level of abstraction, and to formally verify designs. On the other hand, the Rational Unified Process (RUP) provides a disciplined approach to assignment of tasks and responsibilities within software processes. Most iterations within RUP phases result in an executable deliverable. RUP consists of perspectives, disciplines, etc. [2]. The work focuses on the first two disciplines which are Business Modeling (BM) and Requirements Engineering (RE). BM involves higher level managing people. RE involves higher level software engineers. BM facilitates discovering RE. On the other hand, considering BM and RE together can result in a synergic effect. Formal methods enable the precise formulation of important artifacts, eliminating ambiguity during the software development process [3]. Deductive inference enables the analysis of infinite computation sequences and is an essential part of everyday life and scientific work. On the other hand, the important question for deductive approach is the lack of automatic methods for obtaining logical specifications understood as (large) sets of temporal logic formulas. Thus, the automation of this process seems justified and particularly important.

The motivation for the work is the lack of tools for deductive-based formal verification of RUP-like processes. Another motivation is the lack of tools for automatic extraction of logical specifications from software models. The contribution of the work is a method for automatic generation of logical specifications considered as sets of temporal logic formulas. Relatively simple yet illustrative examples of the approach are provided.

Work by Morimoto [4] contains a survey of formal verification methods for business processes. It discusses automata, model checking, communicating sequential processes, Petri nets, Markov networks, and all these issues are discussed in the context of business process management and web services. Work by Brambilla et al. [5] contains some aspects of workflows and temporal logic, but formulas are mostly created manually and formal verification is not discussed widely. In work by Kazhamiakin [6], a method based on formal verification of requirements using temporal logic and model checking approach is proposed, and a case study is discussed.

II. Basic assumptions

The idea of workflow patterns is crucial for this work. They constitute a kind of primitives to enable development of software models and modeling logical specifications. A set of temporal logic formulas is linked to every pattern. The basic issues related to temporal logics and their syntax and semantics are discussed in many works, e.g. [7]. The considerations in this work are limited to the *linear-time temporal logic* LTL, and attention is focused on the *propositional linear time logic* PLTL. The *elementary set* $pat()$ of formulas over atomic formulas a_i, where $i > 0$, which is also denoted $pat(a_i)$, is a set of temporal logic formulas $f_1, ..., f_m$ such that all formulas are syntactically correct. The example of an elementary set is $pat(a, b, c) = \{a \Rightarrow \Diamond b, b \Rightarrow \Diamond c, \Box\neg(a \wedge b \wedge c)\}$ which is a three-element set of formulas created over three atomic formulas. The *logical expression* W_L is a structure similar to the well-known regular expressions and allows to express the complex and nested workflow model in a literal notation. For example, $Seq(Split(a, b, c), Cond(d, e, f))$ shows the sequence of a parallel split followed by conditional execution of some tasks, and the meaning of all patterns is intuitive and not formally defined.

Every pattern has a predefined and countable set of linear temporal logic formulas. Thus, all acceptable patterns constitute a set of *predefined design patterns* Π. The example of such a set for business models is shown in Fig. 1. Software workflows should be modeled using predefined patterns only.

978-83-60810-53-8/$25.00

Most elements of this predefined set, i.e. comments, two temporal logic operators, classical logic operators, are not in doubt. The slash allows to place more than one formula in a single line. The $entry$, or shortly en, and $exit$, or shortly ex, expressions are collections of atomic formulas which describe, informally speaking, the potential logical entry points end logical exit points for a particular pattern. The entry and exit formulas represent a pattern as a whole. f_1, f_2 etc. are atomic formulas for a pattern. They constitute a kind of formal arguments for a pattern.

The *logical specification* L consists of all formulas derived from the logical expression W_L using the algorithm $\mathcal{A}$, i.e.

$$L(W_L) = \{f_i : i \geq 0 \wedge f_i \in \mathcal{A}(W_L, P)\} \quad (1)$$

where f_i is a PLTL formula. The generation algorithm $\mathcal{A}$ has two inputs. The first one is a logical expression W_L which is a kind of a variable, i.e. it varies when a model (workflow) is subjected to any modification by software engineers. The second one is the predefined set P for business models or for activity models which is a kind of constant, i.e. it is once defined and then widely used. The output of the algorithm is a logical specification understood as a set given by a formula 1. However, generation of logical specifications is not a simple summation of formula collections from predefined design patterns. The algorithm ($\mathcal{A}$) is as follows:

1) at the beginning, the logical specification is empty, i.e. $L = \emptyset$;
2) the most nested pattern or patterns are processed first; then, less nested patterns are processed one by one, i.e. patterns that are located more towards the outside;
3) if the currently analyzed pattern consists of atomic formulas only, the logical specification is extended, by summing sets, by formulas linked to the currently being analyzed pattern $pat()$, i.e. $L = L \cup pat()$;
4) if any argument is a pattern itself, then the logical disjunction of its $entry$ and $exit$ formulas is substituted in the place of the pattern as an argument.

III. Business modeling

Business modeling BM allows to understand the structure and behavior of the organization in which a system is to be deployed. It also ensures stakeholders to have a common understanding of the target organization. *Business Process Modeling Notation* BPMN is a standard graphical notation provided by the Business Process Management Initiative (BPMI) for the modeling of business processes. BPMN is becoming the dominant modeling notation, bridging the gap between business process design and process implementation. The main goal of BPMN is to provide a notation that is understandable by all business users, from business analysts to technical developers, and finally, to business people who will manage and monitor those processes [8]. An important parts of BPMN are 21 patterns which are introduced by van der Aalst et al. [9]. Gradually building in complexity, process patterns were broken down into six categories: Basic Control Flow Patterns, Advanced Branching, Structural, Multiple Instances, State Based, and Cancellation. For example, the basic control flow patterns are divided into five particular patterns: Sequence, Parallel Split, Synchronization, Exclusive Choice, and Simple Merge, the meaning of which is defined in terms of temporal logic formulas in Fig. 1.

```
                                /* ver. 27.04.2013
/* Basic Control Patterns
Sequence(f1,f2):
entry=f1 / exit=f2
f1 => <>f2 / ~f1 => ~<>f2 / []~(f1 & f2)
ParallelSplit(f1,f2,f3):
en=f1 / ex=f2,f3
f1 => <>f2 & <>f3 / ~f1 => ~<>f2 & ~<>f3
[]~(f1&(f2|f3))
Synchronization(f1,f2,f3):
en=f1,f2 / ex=f3
f1 & f2 => <>f3 / ~f1 | ~f2 => ~<>f3
[]~((f1|f2)&f3)
ExclusiveChoice(f1,f2,f3):
en=f1 / ex=f2,f3
f1 => (<>f2 & ~<>f3)|(~<>f2 & <>f3)
~f1 => ~<>f2 & ~<>f3
[]~(f2 & f3) / []~(f1&(f2|f3))
SimpleMerge(f1,f2,f3):
en=f1 / ex=f2,f3
f1|f2 => <>f3 / ~f1 & ~f2 => ~<>f3
[]~(f1&f2) / []~((f1|f2)&f3)

/* ..... [other] Business Patterns
```

Fig. 1. A predefined set of patterns P for BPMN models

Let $\Pi_1 = \{Sequence, ParallelSplit, Synchronization, ExclusiveChoice, SimpleMerge, ...\}$ be a predefined set of patterns for business models, with aliases Seq, $Split$, $Synch$, $Choice$, and $Merge$, respectively. The meaning of patterns is formally defined in Fig. 1. Although the above set contains only some patterns to present the main ideas of the work, other workflow patterns together with their temporal logic formulas could be defined in future research.

The business model for flight dispatch is considered below. The BPMN model is shown in Fig. 2. When the decision to realize the flight is taken, then some processes are carried out concurrently. Some of them relate to the preparation of the aircraft and the crew, while others – to passenger handling and loading of baggage. Taking off must be preceded by a permission to start with all of its internal actions. Every business process should be decomposed into a business use case and a series of activities. The logical expression W_L is

$$Seq(Split(Initiate, Split(PreFlightPreparation, AircraftPreparation, PassengerHandling), FlightCoordination), Synch(Synch(CrewPreparation, BaggageLoading, AircraftTaxiing), PermissionStart, TakingOff))$$

or after the substitution of propositions (business processes) as capital letters of the Latin alphabet: A – Initiate, B – PreFlightPreparation, C – AircraftPreparation, D – Passenger-Handling, E – FlightCoordination, F – CrewPreparation, G – BaggageLoading, H – AircraftTaxiing, I – PermissionStart, and J – TakingOff. A logical specification L for the above logical expression is built in such a way that patterns are processes in the following order: most nested $Split$, most

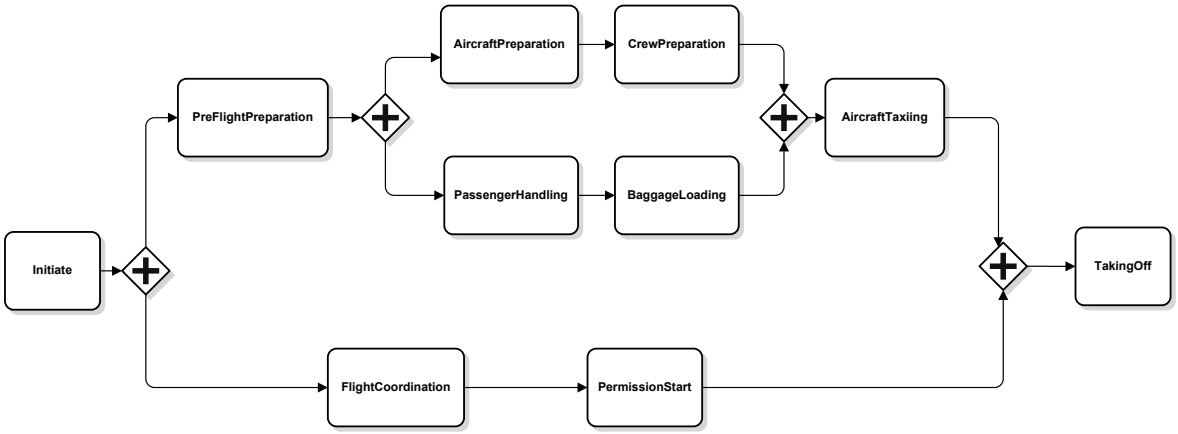


Fig. 2. A business model for a flight dispatch

nested $Synch$, outer $Split$, outer $Synch$, and Seq. The resulting logical specification contains formulas

$$\begin{aligned} L = \{&B \Rightarrow \Diamond C \wedge \Diamond D, \neg B \Rightarrow \neg\Diamond C \wedge \neg\Diamond D,\\ &\Box\neg(B \wedge (C \vee D)), F \wedge G \Rightarrow \Diamond H, \neg F \vee \neg G \Rightarrow \neg\Diamond H,\\ &\Box\neg((F \vee G) \wedge H), A \Rightarrow \Diamond(B \vee C \vee D) \wedge \Diamond E,\\ &\neg A \Rightarrow \neg\Diamond(B \vee C \vee D) \wedge \neg\Diamond E,\\ &\Box\neg(A \wedge ((B \vee C \vee D) \vee E)), (F \vee G \vee H) \wedge I \Rightarrow \Diamond J,\\ &\neg(F \vee G \vee H) \vee \neg I \Rightarrow \neg\Diamond J,\\ &\Box\neg(((F \vee G \vee H) \vee I) \wedge J),\\ &(A \vee C \vee D \vee E) \Rightarrow \Diamond(F \vee G \vee I \vee J),\\ &\neg(A \vee C \vee D \vee E) \Rightarrow\\ &\neg\Diamond(F \vee G \vee I \vee J),\\ &\Box\neg((A \vee C \vee D \vee E) \wedge (F \vee G \vee I \vee J))\} \end{aligned} \quad (2)$$

Formal *verification* is the act of proving the correctness of a system. Liveness and safety are a taxonomy of system properties. *Liveness* means that the computational process achieves its goals (something good eventually happens). *Safety* means that the computational process avoids undesirable situations (nothing bad ever happens). The liveness property for the model can be

$$A \Rightarrow \Diamond J \quad (3)$$

which means that **if the initiation is executed then sometime in the future the aircraft take off**, or formally $Initiate \Rightarrow \Diamond TakeOff$. When considering the property, the entire formula to be analyzed using, for example, the semantic tableaux method is

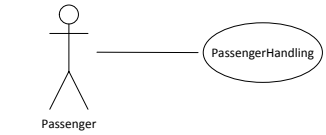
$$C(L) \Rightarrow (A \Rightarrow \Diamond J) \quad (4)$$

where $C(L)$ is a logical conjunction of all formulas that belong to the logical specification L, c.f. formula 2. In a similar way, safety formulas are considered, e.g. $\Box\neg(\neg PermissionStart \wedge TakingOff)$ or even formula $\Box\neg(PermissionStart \wedge \neg CrewPreaparation.c \wedge \neg BaggageLoading.c)$ the meaning of which seems understandable, and the $.c$ suffix means the logical condition associated with an activity. However, the presentation of a full inference tree for these cases exceeds the size of the work. In the case of the semantic tableaux method for logical reasoning, when the falsification of the semantic tree is received, the open branches are obtained and provide information about the source of the error. This is another advantage of the method.

IV. Requirements modeling

Once the business model is built, a requirements model is developed. The transition between the models is done in such a way that a single business process is mapped to a single business use case [2]. A business use case is always actor-centric. A business use case *scenario* is a description that illustrates the behavior from the actors's point of view. Every scenario allows to identify and extract atomic activities. Afterwards, the *activity diagram* enables the modeling of workflows for atomic activities using predefined workflow patterns. It supports choice, concurrency and iteration. The activity diagram shows how an activity depends on others. Nesting of activities is permitted. The following design patterns for activities are introduced [10], [11]: *sequence*, *concurrent fork/join*, *branching* and *loop-while* for iteration. Thus, the predefined set of patterns for activity workflows is $\Pi_2 = \{Sequence, Concurrency, Branching, LoopWhile\}$, and aliases are: Seq, $Concur$, $Branch$, and $Loop$, respectively.

The business use case "PassengerHandling" is discussed below. This is one of the processes received from a business model shown in Fig. 2. The use case scenario is in Fig. 3. The use case scenario is in Fig. 4. When the passenger holding a valid ticket arrives at the airport, the airport check-in, i.e. counter or self, must be made to confirm the passenger's presence, and then a boarding pass is issued. When the passenger has hold baggage, then it must be registered. (In Europe) when the passenger is outside the non-Schengen countries, then border and custom controls need to be performed. The last step before boarding is the security control. Not to introduce the simple branching (if-then) as another pattern for the activity



Use Case: PassengerHandling
Preconditions: Passenger has a valid ticket without check-in
Scenario: 1. Passenger's "Check-in"or"SelfCheck-in" 2. If necessary then "HoldBaggage" 3. If non-Schengen then "BorderControl" and "CustomControl" 4. Passenger's "SecurityControl" 5. Passenger's "Boarding"
Postconditions: Passenger completed all formalities related to the flight

Fig. 3. A business use case (top) and its scenario (bottom)

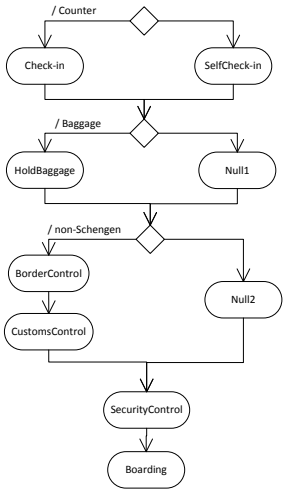


Fig. 4. An activity diagram for the scenario from the Fig. 3

diagram, the empty activity ("Null") is introduced instead a full branching (if-then-else). The Null activity does not consume time, i.e. after reaching the activity it is always immediately completed.

After the substitution of propositions (atomic activities) as small letters of the Latin alphabet: a – Counter, b – CheckIn, c – SelfCheckIn, d – Baggage, e – HoldBaggage, $n1$ – Null1, f – nonSchengen, g – BorderControl, h – CustomControl, $n2$ – Null2, i – SecurityControl, and j – Boarding, then the expression W_L is

$$Seq(Seq(Branch(a,b,c),Branch(d,e,n1)),\\ Seq(Branch(f,Seg(g,h),n2),Seq(i,j)))$$

A logical specification L for the above logical expression is build in the same way as it is shown in the previous section. An example of the liveness property for the model can be

$$e \Rightarrow \Diamond j \tag{5}$$

which means that **if the hold baggage for a passenger is registered then sometime in the future the passenger is boarding**, or more formally $HoldBaggage \Rightarrow \Diamond Boarding$.

V. Conclusion

A method for a deductive-based support developing software models for the first two RUP disciplines is proposed. Also, a method for automatic generation of logical specifications is defined.

Future works might result in development of CASE software which supports deduction-based formal verification of software development processes for CPS systems. Considering Concurrent Communicating Lists [12] is encouraging for strengthen and join these directions of research.

Acknowledgment

This work was supported by the AGH UST internal grant no. 11.11.120.859.

References

[1] E. A. Lee, "Cyber physical systems: Design challenges," in *Proceedings of the 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing*, ser. ISORC 2008. IEEE Computer Society, 2008, pp. 363–369. [Online]. Available: http://dx.doi.org/10.1109/ISORC.2008.25

[2] P. Kruchten, *The Rational Unified Process: An Introduction*, ser. The Addison-Wesley Object Technology Series. Addison Wesley, 2004.

[3] J. Woodcock, P. G. Larsen, J. Bicarregui, and J. Fitzgerald, "Formal methods: Practice and experience," *ACM Computing Survey*, vol. 41, no. 4, pp. 19:1–19:36, 2009.

[4] S. Morimoto, "A survey of formal verification for business process modeling," in *Proceedings of the 8th International Conference Computational Science (ICCS 2008), June 23–25, 2008, Kraków, Poland, Part II*, ser. Lecture Notes in Computer Science, M. Bubak, G. D. van Albada, J. Dongarra, and P. M. A. Sloot, Eds., vol. 5102. Springer–Verlag, 2008, pp. 514–522.

[5] M. Brambilla, A. Deutsch, L. Sui, and V. Vianu, "The role of visual tools in a web application design and verification framework: A visual notation for ltl formulae," in *Proceeding of the 5th International Conference on Web Engineering (ICWE 2005), July 27-29, 2005, Sydney, Australia*, ser. Lecture Notes in Computer Science, D. Lowe and M. Gaedke, Eds., vol. 3579. Springer–Verlag, 2005, pp. 557–568.

[6] R. Kazhamiakin, M. Pistore, and M. Roveri, "Formal verification of requirements using spin: A case study on web services," in *Proceedings of 2nd International Conference on Software Engineering and Formal Methods (SEFM 2004), 28–30 September 2004, Beijing, China*, 2004, pp. 406–415.

[7] E. Emerson, *Handbook of Theoretical Computer Science*. Elsevier, MIT Press, 1990, vol. B, ch. Temporal and Modal Logic, pp. 995–1072.

[8] OMG, "Business process modeling notation specification, version 1.2," January 2009, OMG Document dtc/2009-01-03, Tech. Rep., 2009.

[9] W. M. van der Aalst, A. ter Hofstede, B. Kiepusewski, and A. Barros, "Workflow patterns," *Distributed and Parallel Databases*, vol. 4(1), pp. 5–51, 2003.

[10] R. Klimek, "Proposal to improve the requirements process through formal verification using deductive approach," in *Proceedings of 7th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2012), 29–30 June, 2012, Wrocław, Poland*, J. Filipe and L. Maciaszek, Eds. SciTePress, 2012, pp. 105–114.

[11] ——, "From extraction of logical specifications to deduction-based formal verification of requirements models," in *Proceedings of 11th International Conference on Software Engineering and Formal Methods (SEFM 2013), 25–27 September 2013, Madrid, Spain*, ser. Lecture Notes in Computer Science, R. Hierons, M. Merayo, and M. Bravetti, Eds., vol. 8137. Springer Verlag, 2013, pp. 61–75.

[12] K. Kułakowski and T. Szmuc, "Modeling robot behavior with ccl," in *Simulation, Modeling, and Programming for Autonomous Robots*, ser. Lecture Notes in Computer Science, I. Noda, N. Ando, D. Brugali, and J. Kuffner, Eds. Springer, 2012, vol. 7628, pp. 40–51. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-34327-8_7

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1381–1387

Studying Interrelationships of Safety and Security for Software Assurance in Cyber-Physical Systems: Approach Based on Bayesian Belief Networks

Andrew J. Kornecki
Embry Riddle Aero University
600 S. Clyde Morris Blvd
Daytona Beach, Florida, USA
Email: kornecka@erau.edu

Nary Subramanian
The University of Texas at Tyler
3900 University Blvd
Tyler, Texas, USA
Email: nsubramanian@uttyler.edu

Janusz Zalewski
Florida Gulf Coast University
10501 FGCU Blvd
Ft. Myers, Florida, USA
Email: zalewski@fgcu.edu

***Abstract*— The paper discusses mutual relationships of safety and security properties in cyber-physical systems (CPS). Generally, safety impacts the system's environment while environment impacts security of a CPS. Very frequently, safety and security of a CPS interact with each other either synergistically or conflictingly. Therefore, a combined evaluation of safety and security that considers their interrelationships is required for proper assessment of a CPS. Bayesian Belief Networks (BBN) can be used for this evaluation where factors related to safety and security of a CPS are assumed to be randomly distributed. The result of this evaluation is an assessment that is non-deterministic in nature but gives a very good approximation of the actual extent of safety and security in a CPS. Using a case study of a SCADA system in an oil pipeline control, the authors present a BBN approach for assessing mutual impacts of security and safety violations. This approach is compared with the Non-Functional Requirements approach (NFR), used previously, which is largely qualitative in nature. This study demonstrates that the BBN approach can significantly complement other techniques for joint assessment of safety and security in CPS.**

I. Introduction

MODERN industrial computer systems are a complex combination of hardware and software. In addition, with the proliferation of the Internet, they are all becoming interconnected, which gave rise to the term cyber-physical systems (CPS), reflecting the fact that embedded computers are interfaced to physical devices and make them accessible in the cyberspace.

The ease of interconnectivity raises a number of previously unknown issues in the design and operation of safety-critical CPS, which are now exposed to security vulnerabilities and related threats. Thus, relevant problems are being addressed by respective professional communities. For example, recent discussions between aviation professionals engaged in the work of RTCA Special Committee SC205 [1] dedicated to the software aspects of airborne systems certification (safety focus) and SC216 [2] dealing with aviation systems security, brought us an interesting perspective. The two committees came up with two sets of guidelines for industry developing aviation systems discussing these issues somehow independent from each other.

Thus, industry faces enormous challenges when designing and implementing software-intensive safety and security related systems exposed to abundant networking environments. The critical observation of this paper is that some aspects of integration of complementary views existing in specific domains are inadequate and exhibit lack of required system and process thinking.

The paper presents a perspective on joint, integrated treatment of safety and security properties in cyber-physical systems, with a potential for quantitative analysis of their interrelationships to provide software assurance, i.e., to achieve a required level of confidence that software systems and services function in the intended manner, are free from accidental or intentional vulnerabilities, provide security capabilities appropriate to the threat environment, and recover from intrusions and failures [3]. Our conjecture is that security and safety can be addressed jointly to measure their mutual impact on system trustworthiness and on each other.

The rest of the paper is structured as follows. Section 2 outlines some previous studies on joint treatment of safety and security, Section 3 introduces the case study of an oil pipeline control system, Section 4 discusses our approach, based on Bayesian belief networks, and Section 5 derives some conclusions.

II. Safety and Security

A. Common Perspective

From the technical perspective, in cyber-physical systems, critical system properties, such as security, safety, reliability, etc., cannot be treated in isolation from each other. In industrial applications, with a control system in charge of the technological process, typically safety was considered a critical property. Computer systems were

This project has been funded in part by a grant SBAHQ-10-I-0250 from the U.S. Small Business Administration (SBA). SBA's funding should not be construed as an endorsement of any products, opinions, or services. The second and third authors gratefully acknowledge the AFRL 2011 and 2012 Summer Faculty Fellowships, in Rome Labs.

designed such that the behavior of computer software or hardware would not endanger the environment in a sense that equipment's failure would cause death, loss of limbs or large financial losses.

On the other hand, the security of industrial computer control systems was typically limited to the physical plant access and off-line protection of data. With the miniaturization of computing devices, growing sophistication of control, and with the advent of the Internet, multiple functions of industrial control systems have become accessible online, which opened doors to enormous security threats. Thus, to increase trustworthiness of industrial computer systems, security concerns have to be taken into account and the mutual relationships of safety and security have to be studied and reconciled.

B. Background

Several industries have attempted to address related issues, for example, railways [4], chemical [5], off-shore [6], automation [7], nuclear [8], and industrial control [9]. Since the publication of a seminal paper by Burns et al. [10], around three dozen papers have been published discussing jointly safety and security issues, recently summarized in [11]. Since then, a more comprehensive review of related issues has been published [12].

Boyes, based on his 25 years of industry experience, discussed the problems of vulnerability of critical infrastructure due to the increasing interactions with external networks [13]. The question posed is whether or not the safety system built on top of the control system is not only safe but also secure. He identified situations when security violations may lead to safety violation and thus related incidents resulting even in some fatalities. He observed that security issues must be considered in safety implementation in any process plant, just as safety issues must be considered when administering conventional information technology security issues.

However, there seem to be only a few studies that aim at assessing both properties in a comprehensive manner, including an impact, which one might have on another in the same system. For example, the OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation) framework [14] provides a checklist-based approach to evaluate safety and security in an organization; however, explicit analysis of tradeoffs between these properties is left to the judgment of evaluators.

Metrics-based approaches can be used to compute safety and security quantitatively: for example, Fenton's [15] causal/explanatory model which uses factors to determine metrics can perhaps be applied in the context of cyber-physical systems as well. Likewise, ATAM, the Attribute Tradeoff and Analysis Method [16], develops a utility tree to capture factors involved in analyzing a design. Again, the tradeoff analysis is mostly implicit. The NFR Approach, where NFR stands for Non-Functional Requirements, allows explicit joint analysis of safety and security properties [17]-[18], by using a goal-orientation. This approach is essential for the current research and described in detail in the next section.

III. NON-FUNCTIONAL REQUIREMENTS APPROACH

The Non-Functional Requirements (NFR) approach is a goal-oriented technique that can be applied to determine the extent to which specific objectives are achieved by a design. The NFR considers properties of a system such as reliability, maintainability, and usability, and could equally well consider functional objectives and constraints for a system. Thus the NFR approach can be applied to evaluate whether a specific design satisfies safety and security requirements for the system.

The NFR approach uses a well-defined ontology that includes softgoals, contributions, and propagation rules [17]. The graph that captures the softgoals, their decompositions, and the contributions is called the Softgoal Interdependency Graph (SIG). The approach relies on a qualitative assessment based on the concept of the contribution "satisficing" positively or negatively the softgoals resulting in determination of the network softgoals to be satisficed or denied.

Subramanian and Zalewski recently applied the NFR [18] to evaluate safety and security of an example cyber-physical system: a typical oil pipeline control SCADA based system (Figure 1) at the Center for Petroleum Security Research (CPSR) [19]. Such system consists of the Master Station and Remote Terminal Units (RTU) connected directly to field instruments measuring pressure and rate of flow of the oil. The field instruments also contain shutoff valves that can change the rate of flow or the pressure. The RTU's communicate with a central master via Ethernet, satellite, cable, cellular phones, or fiber optics.

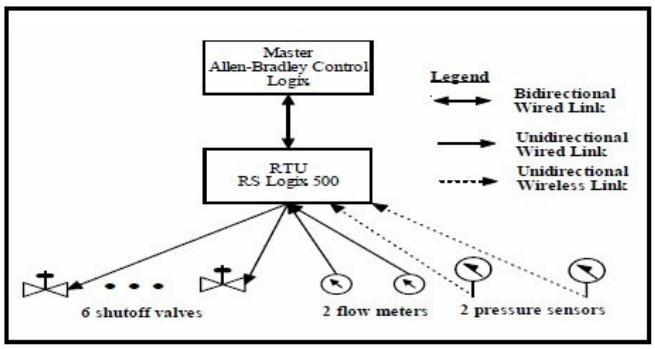


Fig. 1 Example of application (oil pipeline flow control)

In the selected example, safety requirements combine operational and maintenance safety. For operational safety, pressure, structural integrity, and correct distribution are

monitored. For maintenance safety, the flow must be diverted to alternate line, leaving the flow-free portion of the pipeline not monitored for operational safety. Security requires that only authorized personnel are to control the system, all events are logged for audit, and encrypted data are used for wireless transmissions.

The results of the study [18] showed that the NFR approach is effective in joint qualitative assessment of security and safety properties, allowing for simultaneous evaluation of impacts lower level variables might have on these system level properties. In the current project, we are using the same case study and apply the technique known as Bayesian Belief Networks (BBN) to address issues of mutual relationships of safety and security, and their impact on each other.

IV. Bayesian Belief Network Approach

A. Background

A Bayesian Belief Network (BBN) is a graphical model representing the conditional probability distribution of a set of random variables. The technique has been used in the last two decades in multiple industrial applications for decision making under uncertainty, including safety assessment [20]. Since its theoretical background has been well described elsewhere [21], in this paper we provide only a brief overview of BBN principles.

The BBN is based on a formula for belief updating from evidence (E) about a hypothesis (H) using conditional probability measurements of the prior truth of the statement updated by posterior evidence:

$$P(H|E) = (P(E|H) * P(H)) / P(E) \qquad (1)$$

The BBN is described by a directed acyclic graph of nodes and arcs. The nodes can assume specific states with apriori defined likelihood or with a certainty (if there is evidence of their actual state). The arcs represent relations among the variables in terms of likelihood of being in a specific state depending on a state of their ancestors.

An arc from node A to node B means that variable B depends directly on variable A (and A is called a parent of B). If the variable represented by a node has a known state then the node is said to be observed as an evidence node. A node can represent a variable, a measured parameter, or a hypothesis.

A Bayesian network is specified by an expert providing an initial assessment of likelihood that the nodes are in a specific state as well as the likelihood of descendant node being in a specific state, assuming states of its parent nodes. The network is then used to perform inference after some evidence about the state of specific nodes is entered. The predictive mode allows the user to determine likelihood of the outcome, i.e., top-level node being in certain state, assuming specific evidence one may have on its preceding nodes. The network allows also use a diagnostic mode of reasoning. Introducing the evidence of a resulting event leads to estimates regarding the causes of this event.

There are several tools supporting development of BBN with a list compiled in [22]. MSBNx has been used in this project [23].

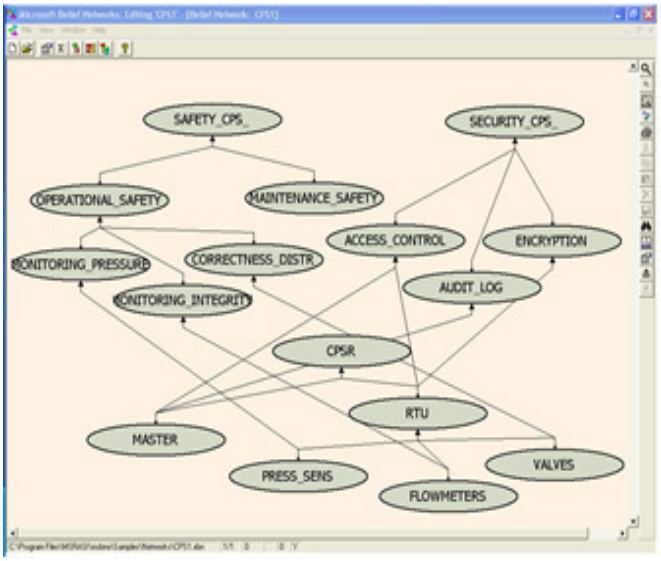


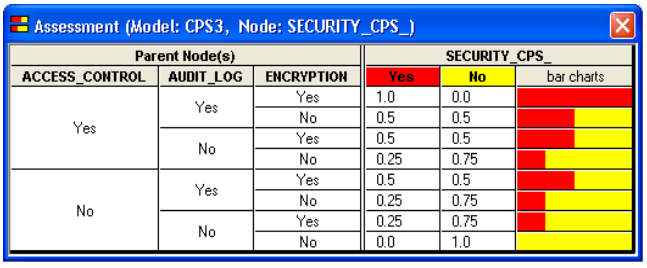
Fig. 2 BBN of the example CPS - oil pipeline control

B. Preliminaries

A BBN model was built for a case study of an oil pipeline control. In reality, safety may be affected by valve fault, pump failure, pressure build-up, leakage, blockage of pipes, and other factors. Security may be affected by lack of authentication and authorization, excessive privileges, wireless transmissions, lack of encryption, connection between the enterprise IT and SCADA networks, lack of audit logs, improper personnel training, poor physical security and the like. However, we consider only a few of these factors to illustrate our ideas. Figure 2 shows a diagram presenting the belief network with safety and security as the top nodes. Here, safety depends on both operational and maintenance safety. Operational safety, in turn, depends on correct monitoring of the pressure (depending on proper work of the pressure sensors) and integrity of the pipeline (depending on correctness of the pressure meters), as well on correctness of flow distribution (proper operation of shutoff valves). Security depends on controlling access, maintaining audit logs, and assuring encryption of transmission. Both Master Controller and RTU's are responsible for access control. Master maintains audit logs while RTU sends data, which may or may not be encrypted. Correct operation of the entire system depends on correctness of the hierarchy of underlying hardware and software.

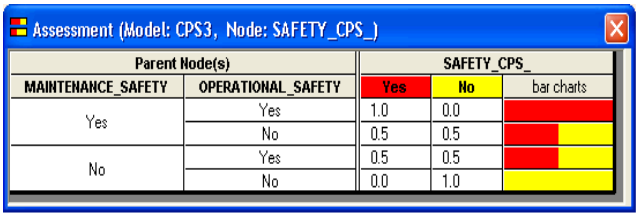
The computation is initialized with the likelihoods reflecting the probability of correct (State 0-YES) or incorrect (State 1-NO) operation. The dependency relations have been also initialized by assuming that incorrect operation of the parent node impacts the descendant node. Two cases were analyzed: all components operated with a specified likelihood of correctness: 90% and 99%.

The example dependency relationships for top-level safety and security nodes are shown in Figures 3 and 4. Likelihood level of the system security and safety properties are determined by dependency relationships based on the specific evidence of the state of the events affecting these properties.



Parent Node(s)			SECURITY_CPS_		
ACCESS_CONTROL	AUDIT_LOG	ENCRYPTION	Yes	No	bar charts
Yes	Yes	Yes	1.0	0.0	
		No	0.5	0.5	
	No	Yes	0.5	0.5	
		No	0.25	0.75	
No	Yes	Yes	0.5	0.5	
		No	0.25	0.75	
	No	Yes	0.25	0.75	
		No	0.0	1.0	

Fig. 3 Dependency relations for Security node



Parent Node(s)		SAFETY_CPS_		
MAINTENANCE_SAFETY	OPERATIONAL_SAFETY	Yes	No	bar charts
Yes	Yes	1.0	0.0	
	No	0.5	0.5	
No	Yes	0.5	0.5	
	No	0.0	1.0	

Fig. 4 Dependency relations for Safety node

Figures 5 and 6 present the example results of inference in a nominal state i.e., assuming the case that the likelihood of correct operation of all base nodes (master controller, flow and pressure sensors, and shutoff valves) is 90%.

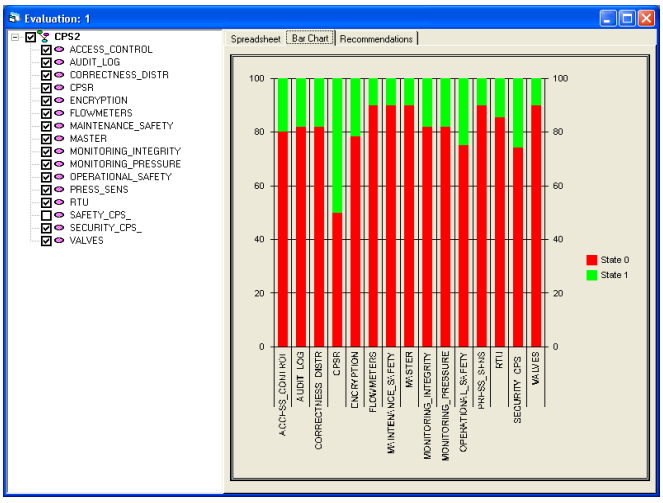
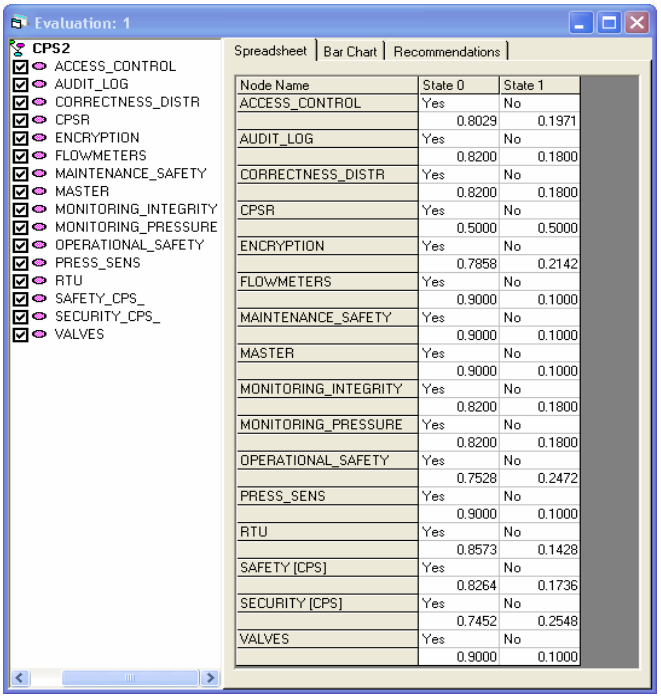


Fig. 5: A nominal state of the system in a bar-chart format

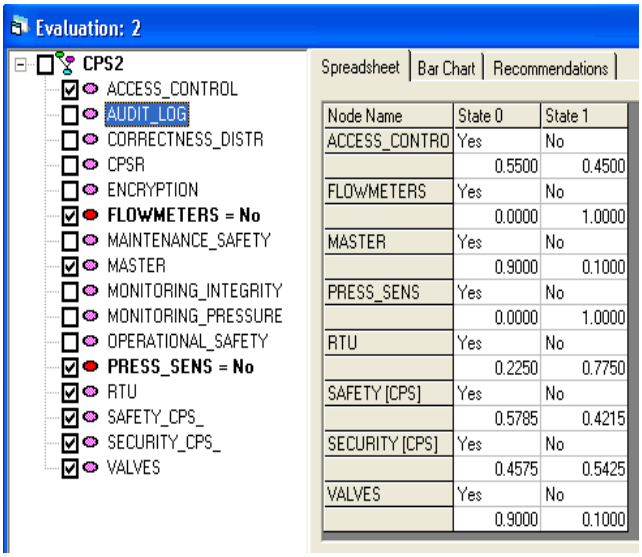


Node Name	State 0	State 1
ACCESS_CONTROL	Yes	No
	0.8029	0.1971
AUDIT_LOG	Yes	No
	0.8200	0.1800
CORRECTNESS_DISTR	Yes	No
	0.8200	0.1800
CPSR	Yes	No
	0.5000	0.5000
ENCRYPTION	Yes	No
	0.7858	0.2142
FLOWMETERS	Yes	No
	0.9000	0.1000
MAINTENANCE_SAFETY	Yes	No
	0.9000	0.1000
MASTER	Yes	No
	0.9000	0.1000
MONITORING_INTEGRITY	Yes	No
	0.8200	0.1800
MONITORING_PRESSURE	Yes	No
	0.8200	0.1800
OPERATIONAL_SAFETY	Yes	No
	0.7528	0.2472
PRESS_SENS	Yes	No
	0.9000	0.1000
RTU	Yes	No
	0.8573	0.1428
SAFETY [CPS]	Yes	No
	0.8264	0.1736
SECURITY [CPS]	Yes	No
	0.7452	0.2548
VALVES	Yes	No
	0.9000	0.1000

Fig. 6: A nominal state of the system in a tabular format (Table 1, case #1, 90% likelihood)

C. Modeling

Several experiments were conducted to assess the impact of specific base elements evidence on the likelihood of the system safety and security represented by top nodes. The results of selected experiments are depicted in Figures 7-9. These three examples show the BBN results when there is failure evidence of elements impacting safety (sensors), impacting security (audit log and encryption), and impacting both (valves and encryption) – all under assumption that likelihood of all other elements being operational is 90%.



Node Name	State 0	State 1
ACCESS_CONTRO	Yes	No
	0.5500	0.4500
FLOWMETERS	Yes	No
	0.0000	1.0000
MASTER	Yes	No
	0.9000	0.1000
PRESS_SENS	Yes	No
	0.0000	1.0000
RTU	Yes	No
	0.2250	0.7750
SAFETY [CPS]	Yes	No
	0.5785	0.4215
SECURITY [CPS]	Yes	No
	0.4575	0.5425
VALVES	Yes	No
	0.9000	0.1000

Fig. 7: An example scenario - safety impact: pressure and flow sensors not working (Table 1, case #3, 90% likelihood).

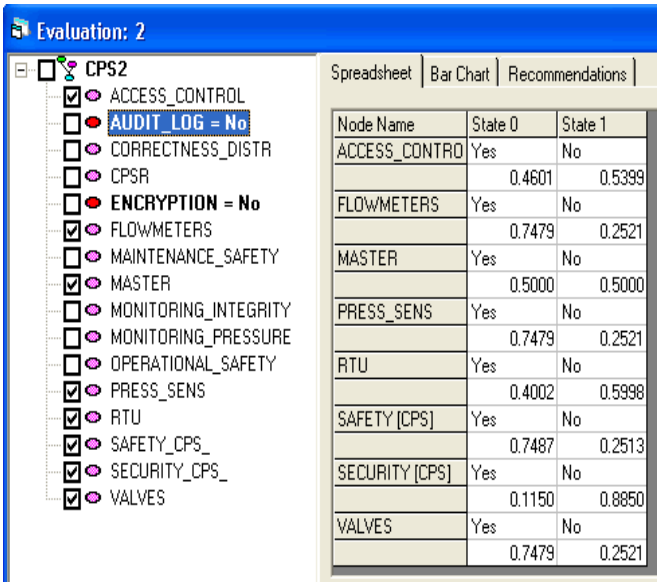


Fig. 8: An example scenario - security impact: audit log and encryption not operational (Table 1, case #8, 90% likelihood)

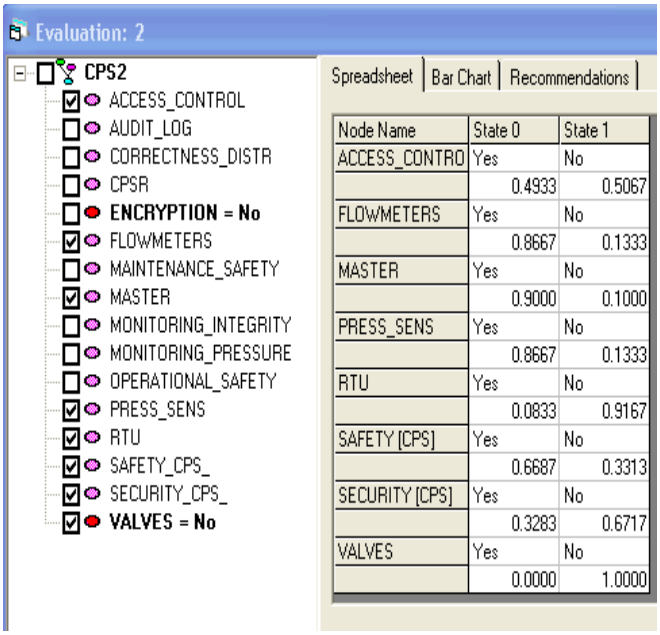


Fig. 9: An example scenario - combined impact valves and encryption not operational (Table 1, case #9, 90% likelihood).

Table 1 illustrates a subset of experiments. After capturing a nominal case (#1), evidence of failing the system components as well as evidence of not operational encryption and audit logs is introduced. Additionally, as presented in the last two rows, the impact of having evidence of a complete failure of safety on security, and vice versa, is analyzed.

Row 1 in Table 1 presents probability levels of safety and security assuming "nominal" values of likelihood of all base events, i.e., when there is no specific evidence and they are defined only by their original probabilities (two scenarios are considered, assuming either 90% or 99% likelihood of correct operations). Consecutive rows present results of deviation from the nominal state and the effect of evidence of these deviations on safety and security. It can be observed that with an evidence of failures or malfunction the probability of a safe/secure operation of the system deteriorates often two-fold.

Using Table 1 one can also compute likelihood of extended scenarios. As an example of a scenario where loss of security negatively impacts safety, consider the case where the RTU in the field was physically tampered which leads to the failure of valve control and thereby permits higher than normal amount of fluid to accumulate in the pipeline. The probability of such compound event can be evaluated from Table 1 as follows:

P(safety violation physical tampering with valve control) = P(safety impact due to security failure)* P(valve failure)

The computation results in probability of safety violation due to a physical tampering destroying the valve control: 0.4668 (0.6907*0.6759) or 0.5811 (0.8357*0.6954), for 90% and 99% scenarios respectively. It is assumed that other evidence is unknown, i.e., the likelihood of correct operation of all remaining components is as specified by the scenario.

TABLE 1: RESULTS OF HYPOTHETICAL SCENARIOS FOR TWO CASES OF THE BASE COMPONENTS OPERATIONAL LIKELIHOOD.

	90% likelihood		*99% likelihood*	
CASE	***Safety***	***Security***	***Safety***	***Security***
#1. nominal	0.8264	0.7452	0.8732	0.8455
#2. valve fails	0.6759	0.5599	0.6954	0.6155
#3. flowmeter & valve fail	0.5785	0.4575	0.5876	0.4993
#4. flowmeter, valve & pressure sensor fail	0.4876	0.3552	0.4876	0.3831
#5. master controller fails	0.8264	0.3552	0.8732	0.3816
#6. encryption fails	0.7487	0.3600	0.8543	0.4350
#7. audit log not operational	0.8264	0.3572	0.8732	0.4372
#8. audit log & encryption fails	0.7487	0.1150	0.8543	0.2047
#9. valve fails & encryption fails	0.6687	0.3283	0.6946	0.3568
Impact				
Safety violation	0	0.6907	0	0.8357
Security violation	0.7893	0	0.8652	0

Additionally, conditional probabilities may also be deduced from Table I. For example, for the scenario where

the valve fails and a security violation occurred, one would wish to deduce the probability that the valve failed due to such security incident. Then:

P(valve failed | security violation) =
P(valve failed and security violation)/P(security violation)

Assuming unknown evidence with 90% scenario, the computation results in the conditional probability 0.7094 (0.5599/0.7893). With 99% scenario, the conditional probability is 0.7114 (0.6155/0.8652).

As shown, BBN's can be used for estimating probabilities of unknown events based on probabilities of known events. As can be expected, the better the data used for modeling BBN, the more trustworthy the computed probabilities. This in turn requires a more accurate modeling of factors affecting both security and safety, which form the basis for BBN's.

The proposed approach allows not only to specify numerical values for safety and security (in terms of their likelihoods), but also allows for quantitative assessment of a relationship between safety and security as well as that of an impact of the status of the system components on safety and security.

An obvious challenge is to identify not only the likelihoods of events at specific nodes representing the system components but also the initial likelihoods of dependency relations between them. These can be derived from failure rates of equipment available, for example, from the military handbook [24] or from industry studies such as [25]. Likelihood estimates can also be obtained from incident rates related to safety and security such as rates of deliberate acts of sabotage and vandalism or the rates of deliberate software attacks [26]. Additionally, the proposed analysis could be conducted, expo facto on a system in which a failure has already occurred to provide some validation evidence and means for calibrating the data.

V. Conclusion

The driving force behind the presented research is that security and safety properties in cyber-physical systems are mutually dependent and influence each other. It is, therefore, natural to seek methods of measuring their mutual impact and assessing their susceptibility to related events and changes in values of basic variables.

In this paper we studied the relationship of safety and security using Bayesian Belief Networks. For a case study of an oil pipeline SCADA based control system, the BBN technique was applied to determine the impact of failures in low-level equipment on the overall security and safety of the entire system and evaluate safety and security in case of equipment or software failures. It turns out that the method proves useful for applied purposes and is comparable to the NFR approach applied previously.

The NFR is a qualitative approach evaluating the safety and security of a cyber-physical system given known factors of a system's configuration (including components and connections). It applies propagation rules to assess NFRs such as safety and security. In contrast, the BBN uses likelihood estimates of a system's configuration to evaluate quantitatively the achievement or denial of safety and security of cyber-physical systems; likelihood estimates can include failure rates of system components and connections or could be likelihood of incidents impacting safety and security. It needs to be noted that [18] describes a qualitative technique to evaluate safety and security. As a result of this evaluation we can conclude to what extent (good or bad) safety and security have simultaneously been achieved in the system. In this paper we have attempted quantitative evaluation of achievement of safety and security in a system using probabilistic computations from BBN.

Therefore, evaluations of safety and security obtained from BBN are rooted in data collected in the field and can be used for both predictive and diagnostic purposes. These data can be used for re-evaluation of contributions in the NFR approach and vice versa, assessments from NFR can be used to re-evaluate critical aspects using the BBN approach. Thus, both these techniques can be used in a complementary manner to iteratively reassess safety and security of cyber-physical systems.

In future work, it would be interesting to include in this study the Safety Case approach [27]. It is based on a graphical Goal Structuring Notation (GSN), similar to NFR, to represent entities and relationships used in the safety argument. Such GSN may be another base to model with BBN's.

References

[1] DO-178C. *Software Considerations in Airborne Systems and Equipment Certification*, RTCA 12-13-11, SC-205, 2011.

[2] DO-326. *Airworthiness Security Process Specification*, RTCA 12-08-10, SC-216, 2010.

[3] N. Mead, J. Allen, M. Ardis, T. Hilburn, A. Kornecki, R. Linger, J. McDonald, *Software Assurance Curriculum Project*, Vol. I, Report CMU/SEI-2010-TR-005, Software Engineering Institute, Pittsburgh, Penn., August 2010

[4] J. Smith, S. Russell, M. Looi. Security as a Safety Issue in Rail Communications, *Proc. SCS 2003, 8th Australian Workshop on Safety Critical Systems and Software*, Canberra, October 9-10, 2003, pp. 79-88.

[5] J. Hahn, D.P. Guillen, T. Anderson. Process Control Systems in the Chemical Industry: Safety vs. Security, *Proc. 20th CCPS, Int'l Conference of the Center of Chemical Process Safety*, Atlanta, Georgia, April 11-13, 2005. Report INL/CON-05-00001.

[6] M.G. Jaatun, T.O. Grotan, M.B. Line. Secure Safety: Secure Remote Access to Critical Safety Systems in Offshore Installations, *Proc. ATC 2008, 5th Intern. Conf. on Autonomic and Trusted Computing*, Oslo, Norway, June 23-25, 2008, pp. 121-133.

[7] T. Novak, A. Treytl. Functional Safety and System Security in Automation. *Proc. ETFA'08, 13th IEEE Conf. on Emerging Technologies and Factory Automation*, Hamburg, Germany, September 15-18, 2008, pp. 311-318.

[8] J. Jalouneix, P. Cousinou, J. Couturier, D. Winter. *A Comparative Approach to Nuclear Safety and Nuclear Security*, Tech. Rep. IRSN 2009/117, Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France, 2009.

[9] A. Kornecki, J. Zalewski. Safety and Security in Industrial Control, *Proc. CSIIRW 2010, 6th Annual Workshop on Cyber Security and Information Intelligence Research*, Oak Ridge, Tenn., April 21-23, 2010.

[10] A. Burns, J. McDermid, J. Dobson. On the Meaning of Safety and Security, *The Computer Journal*, Vol. 35, No. 1, pp. 3-15, 1992.

[11] J. Zalewski, S. Drager, W. McKeever, A. Kornecki. Towards Experimental Assessment of Security Threats in Protecting the Critical Infrastructure. *Proc. ENASE 2012, International Conf. on Evaluation of Novel Software Approaches to Software Engineering*, Wroclaw, Poland, June 29-30, 2012, pp. 207-212.

[12] L. Pièrre-Cambacédès, M. Bouissou. Cross-fertilization between Safety and Security Engineering, *Reliability Engineering and System Safety*, Vol. 110, pp. 110-126, 2013.

[13] W. Boyes, Safety, Security and Complex Systems in Critical Infrastructure Protection, Invited Talk. *SAFECOMP 2009, 28th International Conference on Computer Safety, Reliability and Security*, Hamburg, Germany, September 15-18, 2009. Available at: http://www.controlglobal.com/articles/2009/CriticalInfrastructure0909.html

[14] OCTAVE, Operationally Critical Threat, Asset, and Vulnerability Evaluation, CERT, Software Engineering Institute, Pittsburgh, Penn., 2008. Available from: http://www.cert.org/octave/

[15] N.E. Fenton, M. Neil. Software Metrics: Roadmap. *Proc. ICSE'00, International Conference on the Future of Software Engineering*, Limerick, Ireland, June 4-10, 2000, pp. 357-370.

[16] R. Kazman, M. Klein, P. Clement. *ATAM: Method for Architecture Evaluation*. Technical Report CMU/SEI-2000-TR-004, Software Engineering Institute, Pittsburgh, Penn., August 2000.

[17] L. Chung, B.A. Nixon, E. Yu, J. Mylopoulos. *Non-Functional Requirements in Software Engineering*, Kluwer Academic Publishers, Boston, 2000.

[18] System Architectures for Cyberphysical Systems, *Proc. SysCon 2013, IEEE Intern. Systems Conference*, Orlando, FL, April 15-18, 2013, pp. 634-641.

[19] Center for Petroleum Security Research, University of Texas at Tyler, 2013. Available from: http://www2.uttyler.edu/cpsr/facilities.php

[20] N.E. Fenton, M. Neil. *The Use of Bayes and Causal Modelling in Decision Making, Uncertainty and Risk*. Agena Risk White Paper. Agena, Cambridge, UK, June 2011. Available from: http://www.agenarisk.com/resources/white_papers/fenton_neil_white_paper2011.pdf

[21] F.V. Jensen, T.D. Nielsen. *Bayesian Networks and Decision Graphs*. 2nd Edition, Springer-Verlag, Berlin, 2007.

[22] K. Murphy. *Software Packages for Graphical Models*, University of British Columbia, Vancouver, Canada, February 12, 2013, Available from: http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html

[23] *MSBNx: Bayesian Network Editor and Tool Kit*, Microsoft Research, Redmond, Calif., 2013. URL: http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/

[24] Military Handbook: Reliability Prediction of Electronic Equipment. Notice 2. MIL-HDBK-217F, 1995.

[25] R. Chalupa. *Failure Modes, Effects and Diagnostics Analysis*. Report No. 06-11-25-R001, Rosemount Corp., Eden Prairie, Minn., 2007.

[26] M.E. Whitman, H.J. Mattord. *Management of Information Security*. 3rd Edition, Cengage Learning, Independence, Kentucky, 2010, pp. 288-291.

[27] T.P. Kelly, I J Bate, J A McDermid, A Burns. Building a Preliminary Safety Case: An Example from Aerospace. *Proc. 1997 Australian Workshop on Industrial Experience with Safety Critical Systems and Software*, Sydney, Australia, October 3, 1997.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1389–1392

Object-oriented Approach to Timed Colored Petri Net Simulation

Michał Kowalski and Wojciech Rząsa
Rzeszow University of Technology
Department of Computer and Control Engineering
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
Email: michal.kowalski.87@gmail.com, wrzasa@prz-rzeszow.pl

***Abstract*—This paper presents object-oriented design of library meant for modeling and simulating Timed Colored Petri Net models. The approach is prepared to integrate TCPN models with crucial parts of larger applications implemented in object-oriented languages. The formal models can be tightly joined with applications allowing the latter to interpret states of the formal model in their domain of responsibility. This approach allows less error-prone and more pervasive use of formal methods to improve quality of software created with imperative languages.**

***Index Terms*—Petri nets, simulation, object-oriented, integration.**

I. Introduction

TImed Colored Petri Net (TCPN) is a flavor of Petri Net formalism designed by K. Jensen [6]. It is suitable for modeling and analysis of distributed and concurrent systems. Since in Colored Petri Nets (CPN) tokens can carry values (*colors*) of specified types (*colorsets*) models created using CPN are more compact and thus more clear. Consequently, CPN can be conveniently used to analyze large systems. Time extension of Timed Colored Petri Net enables analysis of time relationships e.g. efficiency.

Petri net based model can be formally analyzed to prove its properties as described e.g. in [9] and for CPN in [6]. The models can also be simulated, to observe behavior of modeled systems. Both approaches are used, to derive conclusions about the analyzed systems [1], [11], [15], [14]. This approach, however, is rarely used in business applications (e.g. in banking, trade, accounting). The first reason, certainly, being complexity and scale of this kind of software. Secondly, for the sake of special competences required for modeling and analysis and thus cost of this process.

Certainly, large applications could benefit from formal-based analysis or from incorporating formally-verified modules. Especially, that most of contemporary systems are not only concurrent, but also distributed and thus significantly complex. However, majority of these systems are implemented using imperative languages that are hardly susceptible to formal analysis, with Java as pervasively used example.

It is however, possible to incorporate formalisms-based approach into larger applications, also the ones implemented using imperative languages [2], [12]. This results in crucial parts of a system that can be formally verified, or directly steered by a formal model, minimizing possibility of errors.

To make this approach more pervasive it is however necessary to prepare libraries designed for programming languages commonly used in business applications and enabling convenient incorporation of formalisms. In this work we present an approach to enable TCPN-based modeling using Java – one of most frequently used object oriented programing languages. We present object-oriented design of library we have implemented to support modeling and simulation of Timed Colored Petri Nets as a part of larger systems. This first approach assumes that TCPN model should be correctly simulated. Formal analysis of TCPN models is considered as possible future work.

II. Related Work

There is great variety of Petri net flavors and there is also great variety of computer tools designed to create and analyze Petri net models. An extensive list is maintained e.g. by University of Hamburg[1]. For Jensen's Timed Colored Petri Nets there are two important software packages: *Design/CPN* [3] and newer, rewritten in Java: *CPN Tools* [7], [10]. Both of them have GUI to create Petri net models and both implement algorithms enabling simulation and formal analysis. Also both tools provide interfaces for external communication, that allow various levels of integration using various programming languages.

Design/CPN has interface called COMMS/CPN [5]. It allows to communicate with simulator while Petri net is being simulated. It is a message passing interface with small set of functions that allow to establish connection, send and receive messages. The functions should be called from Standard ML code being part of Petri net model (e.g. guards). There is Java/CPN interface, that facilitates establishing communication with COMMS/CPN from Java software.

The newer CPN/Tools has two kinds of interfaces forming Access/CPN framework [17]. The first one, available for Standard ML and more tightly connected with the model is *designed with state space analysis in mind*. It is worth mentioning, that this interface is meant to be formalism independent, not strictly connected to TCPN. The second part of Access/CPN is *Java CPN Model Interface*. It allows to

[1]http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/db.html

978-83-60810-53-8/$25.00

create or import model created using CPN Tools. It also supports TCP/IP based communication with running simulator.

The presented TCPN design and analysis tools certainly provide interesting interfaces that can be used in selected applications. The message passing communication allows to steer TCPN simulator and respond to some events. It is however a low level communication solution. It also does not allow for tight coupling of TCPN models with applications: TCPN model should be implemented using Standard ML programing language, and Petri net side of communication should also be programmed in Standard ML. It is not possible to implement guards and inscriptions directly in object-oriented language and pass objects as token values.

Petri Net Kernel [16] is a Java library designed to support implementation of various types of Petri nets. It can also be adapted for Timed Colored Petri Nets and in fact was exploited in our research done so far [12], [13], [14]. This general purpose tool has an important disadvantage: token values in Colored Petri Nets are represented by `String` objects. This approach has certain justification: in Java `String` objects are immutable, as tokens in Petri nets. Consequently, design of the programing language ensures one of important assumptions of the formalism. This has, however, two important disadvantages. First, in applications where tokens should carry values more complex than strings, it is not convenient to develop `String` representation of each possible value to pass it to the model and then parse it to an object when its being collected. Second, the performance loss resulting from necessity of parsing the strings is significant.

III. Challenges of Object-oriented Design for TCPN

Library designed in this work should enable modeling and correct simulation of TCPN models. The simulator should be able to run automatically without requiring user interaction and efficiency of simulation must be sufficient for practical applications.

Tight integration of a model as a part of a larger application should be possible, in order to allow subsequent states of a model to be automatically interpreted by an application in its domain of responsibility. In order to limit the effort required by integration, model of Petri nets should be created using classes and objects of Java. Additionally, the library should put possibly small restrictions on objects that are traversing TCPN models in the form of *tokens* and on implementation of TCPN *guards* and *inscriptions*.

Petri net formalism, however, puts restrictions on behavior of its models. The one that strongly affects the design of the library and presents significant challenge, concerns behavior of tokens. Tokens in TCPN are immutable similarly to symbols in functional languages. Presented solution must correspond to this demand, to ensure correct behavior of TCPN model. This requirement must be reconciled with the need to put almost arbitrary object as token value, as mentioned above.

IV. Object-oriented Representation of TCPN

This section presents object-oriented design of the library together with solutions to major problems.

A. Elements of the TCPN Graph

The bipartite graph of Timed Colored Petri Nets consisting of two types of vertexes: *Places* and *Transitions* and of *Arcs* joining the vertexes is described by the natural entities used in object-oriented design.

Places are represented by objects of class `Place` holding `String` attribute describing `name`, `Class` attribute called `type` and describing type of the place and `marking` attribute described by a class implementing `IMarking` interface.

Transitions are described by two attributes: `name` of class `String` and by *guard* used to determine if the transition can be fired. *Guard* is implemented as a reference to an object implementing `IGuard` interface which defines only one method called `guard`. The method gets binding as an argument and returns a `Boolean` value to indicate the guard result.

In this work we distinguished *input arcs* (from a place to a transition) and *output arcs* (from a transition to a place) and described them separately. For simplicity, we assumed that *input arcs* will be used only to define number of tokens used in firing a transition, more complex operations can be performed on *output arcs* using not only basic expressions but also functions. This assumption does not limit expressiveness power of TCPN and considerably facilitates simulation process.

Input arcs are objects of class `InputArc` and hold references to the places being their sources. The `InputArc` class objects have also `inscription` attribute describing tokens that should flow through this arc while firing transitions. The inscription is a `String` in the format analogous to the one defined by K. Jensen [6] and describing count of tokens and variables.

Output arcs are represented by objects of class `OutputArc` and containing references to their destination places and inscriptions. The inscriptions are represented by objects implementing the `IExpression` interface that defines method called `process`. The method receives binding of the transition being fired as an argument, and returns collection of tokens (class `Token`) that should be put in the output place as a result of transition firing. The `process` method can perform any desired operations on tokens and their contents. `BasicOutputArcExpression` class provides means to define simple expressions analogous to the ones used for input arcs.

The references between subsequent elements of TCPN join transitions with their input and output arcs, while the arcs store references to their source or destination places. This solution allows to conveniently access required information while firing a transition by simulator. Parser for arc inscriptions provided as strings is an implementation of finite automata.

B. Tokens

Tokens indicating state of a Petri net model are represented by generic class `Token`. Type of the class parameter corresponds to the type of the token used in Petri nets. It also determines type for the `value` field that holds the token value. This way any class can be used as token type, according to the requirements. Token timestamp is represented by `timestamp` field of the `Token` class.

Consistent simulation of a Petri net requires token values to be controlled by the net only. In Java this is not easy to achieve, since objects are passed to and from methods by reference, not by value. Consequently, it might happen that one object is referred by more than one token as its value or a user could modify value of an object during simulation interfering with the simulation algorithm.

The TCPN designed by K. Jensen are described using functional programing language. This solves the previously mentioned problems by natural means of this paradigm: the tokens are immutable, as are symbols in functional programming. To emulate this behavior in Java, we decided that the following requirements are crucial: (1) TCPN simulator can be trusted to deal with the tokens respecting Petri net principles. (2) By design we must ensure that the user will not be able to breach the rules of TCPN simulation, despite his or her lack of knowledge concerning these principles. (3) For efficiency reasons objects should be copied as rarely as possible.

In order to meet the above requirements, we clone token values (using deep cloning described in [4]) when they enter simulator structures and when they are about to leave these structures. Thus, the compromise ensures correct simulation, regardless of how the token values are processed outside of the TCPN structures, concurrently preserving reasonable efficiency overhead.

C. Marking

Simulation of TCPN consists on subtracting and adding tokens from and to the markings of Petri net places, therefore proper implementation of marking is crucial for efficiency of the simulator. The structure used for implementation of marking must enable efficient search and verification of existence of tokens of known values. In case of Jensen's timed nets the structure should also enable efficient consideration of token timestamps. As stated in [8] the marking can be a composite structure consisting of more than one data structures holding tokens and of consistent interface allowing to operate on the marking as a whole and designed for Petri net simulator.

The implementation corresponding to the above mentioned requirements is placed in `BasicMarking` class. During simulation of a timed net at the given moment only these tokens from a marking are important, that have timestamps not greater than current simulation time. Therefore, conforming to the solution described by Mortensen et al. in [8], we decided to divide tokens in a marking into two separate structures. The tokens, that can be used in current firing of a transition (i.e. with timestamps not greater than current simulation time) are stored in `activeTokens` field. The tokens with timestamps indicating that they can be used later are placed in `waitingTokens` field. These fields refer to objects of different classes, implementing different data structures. The `activeTokens` data structure is an implementation of hash table with support for storing multiple values for a single key. The hash function in this structure uses token values as keys. Thus tokens with given values can be found efficiently while firing a transition, this being a key of efficient TCPN simulation. When simulation clock is advanced selected tokens from `watingTokens` in each place must be moved to the `activeTokens` structure. To perform this operation efficiently, the `waitingTokens` structure is a priority queue sorted by tokens timestamps (`TokenComparator` class is responsible for proper comparison of the tokens in the queue).

The `BasicMarking` class implements `IMarking` interface and the `BasicMarking` can be easily substituted by different implementation of marking implementing the same interface, which ensures all required methods are provided by the implementation. Each marking must implement `getDistinctTokens` returning list of tokens from the marking used to generate bindings while simulation. Boolean method `containsToken` allows to verify if the given token exists in the marking. Method `getToken` removes from the marking a token holding given value and returns this token. Methods `putToken` and `putTokens` enable adding tokens to the marking. During simulation `getNextTime` method is used to determine the least value of the clock that would release new tokens from `waitingTokens` to `activeTokens` in this marking. Finally, `setTime` method is called while each change of simulation clock.

D. Petri Net Structure

The structure of TCPN model is stored in an object of class `CPNet`. The user is supposed to create TCPN model by creating `Place`, `Transition`, `InputArc` and `OutputArc` objects together with guards and inscriptions implementing `IGuard` and `IExpression` interfaces. For efficiency of simulation it is however vital, to provide various information about the net as quickly as possible, without the need to repeatedly exploit computationally intensive graph algorithms. Therefore, after user provided the model, the `CPNet` class performs additional processing in order to cache data concerning model structure and efficiently provide required information for the simulator.

On the basis of the list of transitions provided by the user the `CPNet` class creates list of all places in the model (as Java `ArrayList`). Additionally it caches Petri net structure in an array analogous to graph incidence matrix. The internal structures are filled by method `init` of class `CPNet` that should be called after the `CPNet` object is provided with the list of objects representing TCPN transition.

The proposed solution allows to reconcile requirements concerning efficient access to different aspects of TCPN model while simulation, with the need to ensure consistency between different representations. Obviously the structure of the Petri net cannot be changed after the call to the `init` method.

V. Simulation Data Structures

Variables are represented in different ways, depending on the context. In the arc structures they are represented as objects of class `Var` containing `String` attribute `name` and `Integer` attribute denoting quantity of tokens to be removed from a place. In a *binding* the variables are represented as their names of class `String`. Similarly to [6] variable's range is limited to arcs connected with one transition and their type must be consistent with type of places their arcs are related to.

Bindings are entities assigning variables to their values during simulation. Subsequent bindings are stored in `HashMap` objects with the key being the variable name and value of class `Token`. Thus, during simulation access to variable values connected with a binding is both convenient and efficient. The `HashMap` is encapsulated in the `Binding` class that exposes two methods: `get` to obtain tokens assigned to a variable and `insert` to set binding between a variable and a token.

VI. Integration Interface

The goal of this work was to enable integration of Petri nets with software implemented using imperative programing language. Designed approach enables this integration in two possible ways.

Firstly, crucial elements that create Petri net model and that are part of its behavior, can be objects used in the other parts of application. Tokens can carry objects of arbitrary classes as their values. Transition guards and arc inscriptions can be implemented to depend on application logic.

Secondly, the presented solution is equipped with event listener interface. The listeners can be registered to be called before and after events concerning changes in markings, transition firing, and changes of simulation clock.

VII. Summary

In this paper we presented object-oriented approach to modeling and simulation using Timed Colored Petri Nets. Petri net model with all its components, including token values, guards and inscriptions, should be implemented in object-oriented programming language. The concept was implemented as library in the pervasively used Java.

The approach allows one to integrate Petri net model in a larger application and let it benefit form the formalism-based simulation. The integration can be done while the model is implemented, by joining its components with components of the larger system. It can also be tightened by the use listener interface that allow to implement code responding to specific events occurring in Petri net model. Consequently, the enclosing software can influence behavior of the formal model. Concurrently, the Petri net can steer behavior of the software that can interpret events from the model and apply them in the software's domain of responsibility.

The design of the approach ensures not only tight and convenient integration with object-oriented software. It also ensures that the rules of TCPN simulation are preserved. Thus, the software can benefit from the formal rules, governing behavior of modeled processes and from their correctness.

The presented solution includes TCPN simulator, that allows execution of created model. Care was taken, to ensure efficiency of TCPN simulation that allows practical applications. Necessary optimizations were designed and implemented and if required, additional solutions can be developed.

References

[1] BinWang, MaodeMa: A Server Independent Authentication Scheme for RFID Systems. IEEE Trans. on Industrial Informatics, vol. 8, no. 3, pp. 689-696, Aug 2012.

[2] Dec G, Jędrzejec B, Rząsa W.: Kolorowana sieć Petriego jako model systemu podejmowania decyzji kredytowej. STUDIA INFORMATICA 2010, Vol. 31, Number 2A (89), 2010.

[3] Christensen S, Jorgensen J. B., Kristensen L., M.: Tools and Algorithms for the Construction and Analysis of Systems Lecture Notes in Computer Science Volume 1217, 1997, pp 209-223 Design/CPN—A computer tool for Coloured Petri Nets

[4] Cooper J. W.: The Design Patterns Java Companion, 1998

[5] Gallasch G., Kristensen L. M.: COMMS/CPN: A communication infrastructure for external communication with Design/CPN. In K. Jensen, editor, Third Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, DAIMI PB-554, pages 75–91. Department of Computer Science, University of Aarhus, Denmark, 2001.

[6] Jensen K.: Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use. Vol. 1, 2, 3. EATCS Monographs on Theoretical Computer Science, Springer-Verlag, 1994.

[7] Jensen K., Kristensen L. M., Wells L.: Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems. International Journal on Software Tools for Technology Transfer (STTT)9(3-4), pp. 213-254, 2007.

[8] Mortensen, K. H. Efficient Data-Structures and Algorithms for a Coloured Petri Nets Simulator. In: Kurt Jensen (Ed.): 3rd Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools (CPN'01), pages 57–74. DAIMI PB-554, Aarhus University, August 2001.

[9] Murata T.: *Petri Nets: Properties, Analysis and Applications.* Proc. of the IEEE, vol. 77, No. 4, April 1989

[10] Ratzer A.V., Wells L., Lassen H. M., Laursen M., Qvortrup J. F., Stissing M. S., Westergaard M., Christensen S., Jensen K.: CPN Tools for Editing, Simulating, and Analysing Coloured Petri Nets. Proc. of 24th International Conference on Applications and Theory of Petri Nets (Petri Nets 2003). Lecture Notes in Computer Science 2679, pp. 450-462, Springer-Verlag Berlin, 2003.

[11] Lv Y., Lee C., Wu Z., Chan H., Ip W.: Priority based Distributed Manufacturing Process Modeling via Hierarchical Timed Colored Petri Net". IEEE Trans. on Industrial Informatics, (to be published).

[12] Rząsa W.: Combining Timed Colored Petri Nets and Real TCP Implementation to Reliably Simulate Distributed Applications. CN 2009, CCIS 39, pp. 79-86, 2009, Eds. A. Kwiecień, P. Gaj, and P. Stera.

[13] Rząsa W.: Timed Colored Petri Net Based Estimation of Efficiency of the Grid Applications. PhD thesis. AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, 2011, Kraków, Poland.

[14] Rząsa W., Bubak M.: Simulation Method Supporting Development of Parallel Applications for Grids. In proc. of CGW'10, pp. 194–201, KrakÃłw 2011, ISBN 978-83-61433-03-3.

[15] Rzońca D., Stec A., Trybus B.: Data Acquisition Server for Mini Distributed Control System, w: KwiecieÅĎ A., Gaj P., Stera P. (Eds.): Computer Networks 2011, Communications in Computer and Information Science 160, Springer-Verlag Berlin Heidelberg 2011, pp. 398-406.

[16] Weber M.:, Kindler E.: The Petri Net Kernel. Petri Net Technology for Communication-Based Systems Lecture Notes in Computer Science Volume 2472, 2003, pp 109–123

[17] Westergaard M., Kristensen L. M.: The Access/CPN Framework: A Tool for Interacting with the CPN Tools Simulator. Applications and Theory of Petri Nets Lecture Notes in Computer Science Volume 5606, 2009, pp 313–322

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1393–1400

Interactive Verification of Cyber-physical Systems: Interfacing Averest and KeYmaera

Xian Li, Kerstin Bauer, Klaus Schneider
Embedded Systems Group
Department of Computer Science
University of Kaiserslautern, Germany
Email: {xian.li,k_bauer,klaus.schneider}@cs.uni-kl.de

***Abstract*—Verification is one of the essential topics in research of cyber-physical systems. Due to the combination of discrete and continuous dynamics, most verification problems are undecidable and need to be dealt with by various kinds abstraction techniques. As systems grow larger and larger, most verification problems are difficult even for purely discrete systems. One way to address this problem is the use of interactive verification. Recently, this approach has also been considered by cyber-physical verification tools like KeYmaera and other classical theorem provers.**

Important requirements for the interactive verification are a precise and readable modeling language as well as the possibility to decompose the system into smaller subsystems. Here, tools like KeYmaera and PVS still need further improvement. On the other hand, these modeling aspects are both addressed within the language Quartz as it provides a complete programming language for cyber-physical systems with standard data types and programming statements as well as a precise compositional semantics that is well-suited for compositional verification.

In this paper, we take the advantages of two different tools, the Averest system and KeYmaera, for the interactive verification of cyber-physical systems. This way, we combine modeling and verification capabilities of Averest and the verification capability of KeYmaera, in order to provide a basis for powerful tool set for the interactive verification of cyber-physical systems.

I. Motivation

CYBER-physical systems are systems that combine discrete and continuous dynamics. The environment of embedded reactive systems often consist of continuous behaviors that are determined by the laws of physics. Formal verification is already hard for discrete systems because of the size of the transition systems, and most verification problems for cyber-physical system are even undecidable, due to the combination of discrete and continuous dynamics.

Inspired by the success of model checking [1] in hardware verification and protocol analysis, there has been increasing research on developing techniques for the automated verification of cyber-physical systems. The main line of research concentrates on model checking of finite abstractions of restricted subclasses of the general model. Most techniques proposed so far in this area either rely on bounded state reachability or on abstraction refinement techniques [2–4]. While the first approach suffers inherently of incompleteness, the latter approach often introduces unrealistic behaviour that may yield spurious errors being reported within the analysis.

Despite the theoretic achievements in research, only a few tools are available to verify non-trivial cyber-physical systems. Tools like e.g. PHAVer [5], HyTech [6], Charon [7, 8] focus on the continuous dynamics. They lack typical program statements and data types.

Furthermore, current tools often require an explicit enumeration of the discrete state space. Although the discrete state space typically consists of only finitely many states, the number of these discrete states can become too large to be handled properly by current computers [9].

HySAT [10] and MathSAT [11, 12] are built based upon a SAT-solver that calls a linear program solver for conjunctions of the linear continuous-part constraints. As this technique requires to encode the whole problem space first, the size of the handleable problems is quite small. BACH [13, 14] provides a convenient GUI to construct rectangular hybrid automata with linear location invariants together with a powerful bounded reachability checker for these systems. Another tool, HybridSAL relation abstracter [15, 16] abstracts the discrete and continuous dynamics of the hybrid system automatically to infinite state discrete transition systems that can be model checked by SAL tools [17].

An alternative approach to verification is based on interactive theorem provers. An approach based on higher-order logic [18] for specification and verification of hybrid control system is described in [19]. A verification framework is presented in [20] to strike the balance between the expressiveness of theorem proving and the efficiency and automation of the state exploration techniques. In order to assist in the deductive verification of hybrid systems, [21] presents a tool implemented as a part of STeP [22]. Deductive methods are used in [23] to deal with the parallel composition of hybrid systems, the operational step semantics and a number of proof-rules within PVS [24] have been formalized as well. KeYmaera [25, 26] is an automated and interactive theorem prover for specification and verification logics for differential dynamic logics for hybrid system. It integrates numerous techniques for automated theorem proving, combining deductive, real algebraic, and computer algebraic prover technologies.

Recently, a new language for modeling, simulation, and verification of cyber-physical systems has been developed in our research group [27]. This language is an extension of the synchronous language Quartz that is derived from the Esterel language. Originating from a programming language for discrete systems, there is a rich set of data types, and many

978-83-60810-53-8/$25.00

statements for expressing discrete behaviors in a convenient way. In particular, generic statements and module hierarchies allow one to describe large parametric systems in a concise way. Thus, the modeling capabilities of Quartz are in many cases better than in comparable languages. Like Quartz, also the extension to cyber-physical systems has a precise formal semantics that defines unique behaviors for given input traces. For this reason, the language lends itself well for formal verification. In particular, Quartz programs can be translated to equivalent symbolic transition relations, and thus provide a sound basis for formal verification. The determinism of the language is also very important for simulation, since it allows one to reproduce once observed behaviors. Based on the programing language Quartz, the *Averest* toolset has been developed. Besides of transformations, hardware/software synthesis, a symbolic model checker and other tools, the Averest toolset has recently been extended by a technique for interactive verification [28]. However, up to now these interactive verification techniques are restricted to the discrete component of Quartz, models with hybrid components cannot yet be dealt with.

In this paper, we therefore propose a new approach for the interactive verification of cyber-physical systems by interfacing the Averest toolset and KeYmaera. While the overall verification task remains within the interactive Averest prover, assertions for the continuous components can be verified with the help of KeYmaera. Thus, the advantages of both systems – the modelling and verification capabilities of Quartz especially w.r.t. the discrete component and the verification capabilities of continuous components within KeYmaera – can be combined in order to result in a powerful tool for the interactive verification of cyber-physical systems. Due to the underlying synchronous language this approach will be very well suited for compositional verification that is a great challenge in the context of cyber-physical systems.

The outline of the paper is as follows. In Section II the synchronous language Quartz and the hybrid language used by KeYmaera are briefly introduced. Then, Section III gives a detailed overview of the interfacing of Averest and KeYmaera for applying interactive verification. The paper will be concluded with the application of the interactive verification to a widely known example in section IV.

II. Preliminaries

In the following, we give a brief overview over the synchronous language Quartz, its hybrid extension for modeling cyber-physical systems and the KeYmaera language.

A. The Synchronous Language Quartz

Quartz is a synchronous language that is derived from the Esterel language. The execution of a Quartz program is defined by so-called *micro and macro steps*, where a macro step consists of finitely micro steps whose maximal number is known at compile time. Macro steps correspond to reaction steps of reactive systems, and micro steps correspond with atomic actions like assignments of the program that implement these reactions. Variables of a synchronous program are *synchronously updated* between macro steps so that the execution of the micro steps within a macro steps is done in the same variable environment of their macro step. This synchronous update is important for avoiding data races, and therefore to ensure determinism.

The language offers many data types like booleans, bit-vectors, signed and unsigned integers that may be bounded or unbounded, real numbers, as well as compound data types like arrays and tuples. Modules are declared with an interface that determines inputs and outputs, and a body statement that may use additional local variables. In the following, we list some of the possible statements to describe the examples given in this paper. A complete definition of the language is found in [29] for the discrete case, and in [27] for the hybrid extension.

Provided that S, S_1, and S_2 are statements, ℓ is a location variable, x is a variable, σ is a boolean expression, and α is a type, then the following are statements (parts given in square brackets are optional):

- $x = \tau$ and $\texttt{next}(x) = \tau$ (assignments)
- $\texttt{assume}(\varphi)$, $\texttt{assert}(\varphi)$ (assumptions and assertions)
- $\ell : \texttt{pause}$ (start/end of macro step)
- $S_1; S_2$ (sequences)
- $S_1 \parallel S_2$ (synchronous concurrency)
- $\texttt{if } (\sigma) \ S_1 \ \texttt{else} \ S_2$ (conditional)
- $\texttt{do} \ S \ \texttt{while}(\sigma)$ (loops)
- $\{\alpha \ S\}$ (local variable)

the `pause` statement defines a control flow location ℓ – a boolean variable being true iff the control flow is currently at ℓ : `pause`. Since all other statements are executed in zero time, the control flow only rests at these positions in the program, and thus the possible (discrete) control flow states are the subsets of these locations.

There are two variants of assignments that both evaluate the right-hand side τ in the current macro step: Immediate assignments $x = \tau$ transfer the value of τ to the left-hand side x directly, while delayed assignments $\texttt{next}(x) = \tau$ assign the value in the next macro step.

If the value of a variable is not determined by assignments of the current of previous macro step, a default value is used according to the declaration of the variable. To this end, declarations of variables consist of a *storage class in addition to their type*. There are two storage classes, namely `mem` and `event` that choose the previous value (`mem` variables) or a default value (`event` variables) in case no assignment determines the value of a variable.

In addition to the statements known from other imperative languages (conditionals, sequences and loops), Quartz offers synchronous concurrency $S_1 \parallel S_2$ and sophisticated preemption and suspension statements (not shown in the above list), as well as many more statements for the comfortable descriptions of reactive systems. There is also the possibility to call once implemented modules and to store modules in packages to support the re-use in the form known from software libraries.

Our Averest system provides algorithms that translate a synchronous program to a set of guarded actions [29], i.e.,

pairs (γ, α) consisting of a trigger condition γ and an action α. Actions are thereby assignments $x = \tau$ and $\texttt{next}(x) = \tau$, assumptions $\texttt{assume}(\varphi)$, or assertions $\texttt{assert}(\varphi)$. The meaning of a guarded action is obvious: in every macro step, all actions are executed whose guards are true. Thus, it is straightforward to construct a symbolic representation or extended finite state machine (EFSM) of the transition relation in terms of the guarded actions (see [29]).

While time in synchronous languages is given in the abstract form of macro steps, cyber-physical systems require the consideration of physical time. In order to combine these inherently different concepts of time, the computational model of macro steps is endowed by a continuous transition that takes place between the immediate and delayed assignments of the macro step. During the continuous transition, which consumes physical time, variables of the new storage class `hybrid` change their values according to the new *flow assignments* `x` $\leftarrow \tau$ or `drv(x)` $\leftarrow \tau$ (that equate variable `x` or its derivation on time `drv(x)` with the expression τ).

The continuous transition of the macro step starts with the variable environment determined by the immediate assignments as initial values. To distinguish between the 'discrete' value and the changing value during the continuous transitions, a new operator `cont(x)` is introduced: `x` always refers to the discrete value of a variable, whereas `cont(x)` refers to the (changing) value during the continuous evolution. For memorized and event variables `x` and `cont(x)` always coincide as these variables do not change during continuous evolutions.

The continuous actions may only occur in special statements of the form `flow` S `until(`σ`)` where S is a list of flow assignments and σ is a so-called *release condition* that terminates the continuous phase defined by the flow statement. Figure 1 depicts a program fragment together with the corresponding EFSM. Starting from location ℓ_1, the immediate assignment `x = 0.0` is executed so that the continuous transition starts with initial value `x = 0.0`. The derivation of `x` is then 1 during the continuous transition, and the continuous transition terminates as soon as the continuous value of `x` is 1. Then, the control flow will move to ℓ_2. However, it may be the case that another flow-statement runs in parallel, and that its continuous transition terminates before `x = 1.0` holds. In this case, the control flow moves to ℓ_2', and it will be restarted from there in the next macro step.

B. KeYmaera: Hybrid Programs

KeYmaera is a verification tool for hybrid systems that combines deductive, real algebraic, and computer algebraic prover technologies [26]. It is an automated and interactive theorem prover for a natural specification and verification logic for hybrid systems. In this section, we give an incomplete overview of the syntax and semantics of KeYmaera programs. Statements not needed in the remainder of the paper will be omitted here, for more detailed information consider [25, 26].

The relevant program statements of KeYmaera are summarized in Table I. During a discrete transition, all right hand sides of the actions $x_i := \tau_i$ are computed in parallel and

Program Statement

ℓ_1`: pause`
`x = 0.0;`
ℓ_2, ℓ_2'`:flow{drv(x)<-1.0} until (cont(x)>=1.0)`

Extended Finite State Machine

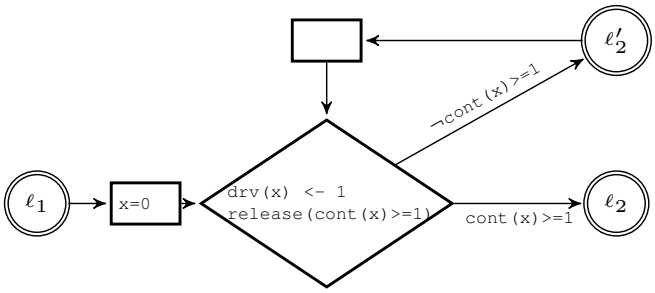


Fig. 1. The Flow Statement of Quartz Programs

TABLE I
SYNTAX OF KEYMAERA (INCOMPLETE)

1	$x_1 := \tau_1, \ldots, x_n := \tau_n$	Discrete jump set
2	$\{x_1' = \tau_1, \ldots, x_n' = \tau_n, H\}$	Continuous evolution
3	$\alpha \,;\, \beta$	Sequential composition
4	if(ϕ) then α else β fi	Deterministic choice
5	$< \alpha > \phi$	Existential operator
6	$[\alpha]\phi$	Universal operator
7	$[[\alpha]]\phi$	Universal Path operator

assigned in a second step, which essentially corresponds to the delayed actions of synchronous languages. Continuous transitions are defined by the differential equation systems $x_i' = \tau_i$ of the variables and the evolution domain H, which is defined by a set of location invariants. Continuous evolutions *may* be terminated at any point of time, they *must* be terminated at the latest, when location invariants would be violated. Thus, continuous transitions are always non-deterministic. As stated in line 3, KeYmaera provides sequential composition. The statement if-then-else in line 4 provides a deterministic choice, that depending on the condition ϕ either executes α or β.

Further statements such as loops and non-deterministic choice will not be considered here. Thus, the only non-determinism provided in the presented fragment of KeYmaera lies within the continuous transition, as all other statements in Table I are deterministic.

The formulas from line $5-7$ are used for verification. $< \alpha > \phi$ is an existential operator that evaluates to `true` iff ϕ holds after at least one valid run of the hybrid program α. Analogously, $[\alpha]\phi$ is an universal operator that evaluates to `true` iff ϕ holds after all valid runs of the hybrid program α. $[[\alpha]]\phi$ is a universal path operator that holds true iff during all runs of the hybrid program α the condition ϕ is satisfied.

III. INTERFACING AVEREST AND KEYMAERA

Due to the combination of discrete and continuous dynamics, most verification problems are undecidable for cyber-physical systems. Even the simple reachability problem is undecidable for most families of cyber-physical systems and the few

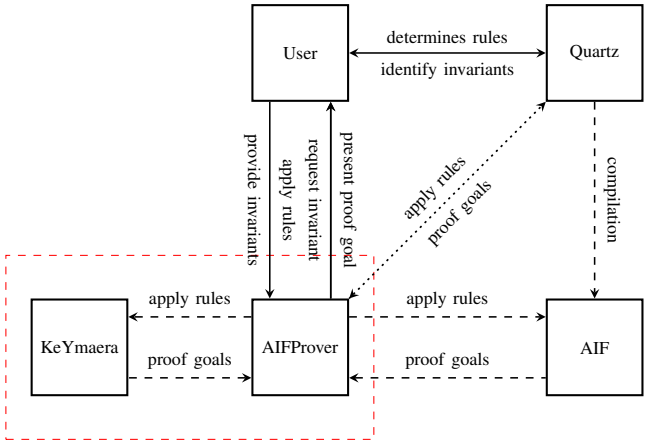


Fig. 2. Idea of our Approach

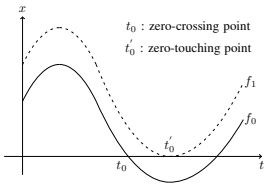


Fig. 3. Zero-Crossing and Zero-Touching Function

l, l': **flow**{
drv$(x_1) \leftarrow \tau_1$; ...; **drv**$(x_n) \leftarrow \tau_n$;
}**until**$(\phi \leq 0)$

Fig. 4. Continuous Transition within Flow Statement

decidability results depend on strong restrictions of either the discrete or continuous component.

An interactive verification approach is pursued by the tool KeYmaera that is especially suitable for verifying parametric hybrid systems [26, 30]. Unfortunately, like most tools for cyber-physical systems, KeYmaera focuses on the continuous component, e.g. only real-valued variables can be modelled. Furthermore, the tool still lacks good capabilities for modelling the parallel composition of systems.

Recently, our research group proposed a new verification approach for discrete Quartz programs (see Figure 2), which is based on the Averest system. Assuming that the system is given as a Quartz program, the user determines rules based on the program structure which are then verified on the intermediate code format AIF, that essentially is a set of guarded actions. By rule applications, the guarded actions are decomposed into smaller AIF files. The proof goals can also be decomposed flexibly, so that the compositional reasoning could be used for verification by the AIFProver, meaning that already proved goals/assertions of some program fragments can be used as assumptions for the remaining program fragments. At the moment, this approach only supports discrete Quartz programs.

In this section, we propose a way to integrate the continuous component of Quartz programs into that framework by interfacing the tool KeYmaera with AIFProver as depicted in Figure 2. As already mentioned in the preliminaries, Quartz and the underlying language of KeYmaera differ in major points (especially the discrete semantics) which makes it difficult to translate complete Quartz programs to KeYmaera. Thus, we will interface Averest and KeYmaera such that proof rules that depend on continuous transitions will be proved by KeYmaera while the overall verification remains within the Averest.

A. General Idea

Recall that continuous transitions of Quartz programs and KeYmaera programs are based on quite different semantics.

In Quartz programs, continuous evolutions *must* terminate at exactly the first point of time, where an active release condition evaluates to true. The continuous assertions given in the form of constraints (which essentially correspond to location invariants) do not influence the control flow of the Quartz program and may be used only for verification purposes. Thus, continuous transitions in Quartz programs are completely deterministic. Contrary to that, continuous transitions of KeYmaera programs are non-deterministic. They *must* be terminated before active location invariants would be violated, they *may* be terminated at any point of time before that. There is no equivalent to our release condition.

In the following, we explain the general idea of how to adapt continuous transitions in Quartz to KeYmaera and still provide the required semantics. Assume for simplicity to have a continuous transition (compare Figure 4) with only one active release condition in the form of $\phi(t) \leq 0$ together with the proof goal σ that shall hold true at the end of the continuous transition, i.e. when the transition terminates according to its release condition $\phi \leq 0$. Assume furthermore, that the release function ϕ satisfies the condition, that it will have a zero-crossing in finite time and that the first zero-crossing point is a 'real' crossing point instead of only a 'zero-touching' point as depicted by the straight line in Figure 3. Then it holds, that for $0 \leq t \leq t_0$ the invariant $\phi \geq 0$ holds, while the same invariant will be violated for $t > t_0$. Thus, adding this invariant to the KeYmaera program as the evolution domain of the continuous transition enforces the transition to terminate at t_0 at the latest while not changing the transition otherwise.

Now, in order to enforce KeYmaera only to consider *one* path, where the continuous transition terminates at time t_0, the proof goal σ at the end of the KeYmaera program must be changed to $\phi \leq 0 \rightarrow \sigma$. According to our assumption w.r.t. the release function ϕ we know that at least one path exists where the continuous transition terminates at t_0. Thus, $\phi \leq 0 \rightarrow \sigma$ evaluates to true iff σ holds on thepath we are interested in.

B. Transformation

Figure 5 shows the transformation of the continuous transition of a macro step to an equivalent KeYmaera program according to the ideas in the previous subsection. The left-hand side of the figure depicts the continuous transition of the macro step together with the given assumptions and proof goals (assertions). The corresponding KeYmaera program (for ease of notation without variable initializations) on the right-hand-side is described in detail below. The transformation of the continuous transition is divided into four parts:

- *Initialization*
  Already obtained assumptions by the interactive verification together with known initial values of the variables are gathered in the assumption ψ_{assume}. Furthermore, a new location label $s := -1$ is introduced that is used for bookkeeping whether the continuous transition has been terminated because of the release-conditions or prematurely.
- *Continuous transition*
  The differential equations of the hybrid variables can be translated one-to-one. $\Psi := \bigvee_i \sigma_i$ is the disjunction of all active release conditions and states the termination criterion for the continuous transition. As already sketched previously, a set of location invariants $\hat{\Psi}$ needs to be added, that is determined by the active release conditions Ψ:
  For ease of notation assume that each single release condition σ is a conjunction of basic expressions of the form $f(x_0,..,x_n) \leq 0$ or $f(x_0,..,x_n) = 0$. If disjunctions occur, these can be dealt with in the same way as several release conditions in parallel. Define now for the first case
  $$\hat{f}_i := f(x_0,..,x_n) \geq 0$$
  and for the latter case
  $$\hat{f}_i := f(x_0,..,x_n) \begin{cases} \leq 0: & f(x_0,..,x_n) < 0 \textit{ initially} \\ \geq 0: & f(x_0,..,x_n) > 0 \textit{ initially} \end{cases}$$
  Then, the release condition σ is replaced by $\hat{\sigma} := \bigvee \hat{f}_i$, as the continuous transition must be terminated only, when all of the basic expressions evaluate to true.
  Several release conditions in parallel correspond to a disjunction of these conditions, as it is only necessary that at least one release triggers. Thus, the location invariants determined by each single release condition must hold true in parallel, i.e. $\hat{\Psi} := \bigwedge_i \hat{\sigma}_i$
- *Termination of the continuous transition*

  Lines $7-10$ define the control flow of the program for the next macro step. Depending on the fulfillment of the active release conditions Ψ, new values of the location variables are determined.
  By definition, the continuous transition terminates due to the release conditions iff $\Psi := \bigvee_i \sigma_i$ evaluates to true. Thus, the arc from line $7-9$ will be executed iff the continuous transition terminates because of the active release conditions. For ease of notation, this information is stored in the location label s, that is either set to 1 or 0, regarding to the termination condition.
- *Proof goal assertion*
  The proof goal ψ_{assert} of the Quartz program must evaluate to true iff the considered path terminated the continuous transition according to the active release conditions. Thus, it must be proved that $s = 1$ holds which is enforced by the KeYmaera proof goal $s = 1 \rightarrow \psi_{assert}$

C. Correctness

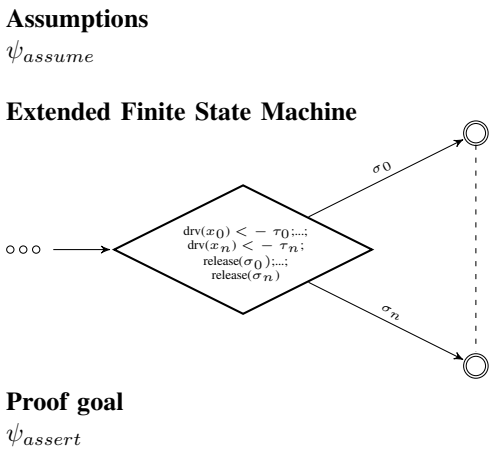
In this section, we present how to use the transformation given in the previous subsection for the interactive verification. The main difficulty lies within the different semantics of KeYmaera and Quartz, namely the difference between determinism (Quartz) and non-determinism (KeYmaera). These difficulties have been dealt with by the introduction of the location invariants $\hat{\Psi}$ together with the additional location label s. The following theorem now states the correctness of transfering the proof results within KeYmaera back to the Quartz language.

Theorem. *Consider a continuous transition in Quartz as given in Figure 5 together with assumptions and proof goals (assertions). Assume furthermore, that the release functions of the continuous transition provide a real zero-crossing (see Figure 3) and that the continuous transition will be terminated in finite time t_0. Then, the following holds:*

1) *The continuous evolution of the variables given by the differential equations are analogous to the ones in Quartz.*
2) *The continuous transition of the KeYmaera program must be terminated at the time $0 \leq t \leq t_0$ and after executing the KeYmaera program, $s = 1$ holds iff $t = t_0$, otherwise $s = 0$.*
3) *If the goal $s = 1 \rightarrow \psi_{assert}$ is proven in KeYmaera, then ψ_{assert} holds after the continuous transition in Quartz.*

Proof.

1) Obvious.
2) The continuous evolutions of the hybrid variables are equivalent within Quartz and KeYmaera. By assumption t_0 is the first point of time where an active release condition evaluates to true. Thus, the location invariants as defined in the previous section are satisfied for all times $0 \leq t \leq t_0$. Furthermore, according to the assumption that the release conditions satisfy the condition as depicted in Figure 3, at least one of the location invariants will be directly violated for $t > t_0$.
   Thus, the duration of the continuous transition of the KeYmaera program can be any time t with $0 \leq t \leq t_0$. The arc given in lines $7-9$ will be executed iff at least one release condition evaluates to true, i.e. if the continuous transition of KeYmaera terminates at time $t = t_0$. This yields that after executing the program $s = 1$ holds iff $t = t_0$, otherwise $s = 0$.




```
1.  \problem {
        /* Initialization */
2.      \ [ ψ_assume
3.      (       s = −1
4.  → \ ⟨
        /* Continuous transition */
5.      {   x_0' = τ_0, ..., x_n' = τ_n,    Ψ̂   } ;
        /* Control flow assignment*/
6.      if( Ψ )    then {
7.          if(σ_0) then l_0 := 1 fi ; ... ;
8.          if(σ_n) then l_n := 1 else ℓ_n = 0 fi ;
9.          s := 1
10.     } else {   s := 0   }    fi
11. \ ⟩
        /* Proof goal assertion*/
12.     ) ( (s = 1)→ ψ_assert )
13. }
```

Fig. 5. Continuous transition in Quartz together with the corresponding KeYmaera program

3) Since by assumption the continuous transition in Quartz terminates at finite time t_0, according to 2), there exists exactly one path satisfying $s = 1$ after executing the KeYmaera program, which corresponds to the continuous transition in Quartz. Thus, if the condition $s = 1 \rightarrow \psi_{assert}$ can be proven by KeYmaera, then ψ_{assert} holds after the continuous transition within Quartz.

■

The first assumption, that the release functions have the form as given in Figure 3 is not a severe restriction, as this is the realistic behaviour in most cases and can be checked in advance. The second assumption, that the continuous transition terminates in finite time can directly be checked by KeYmaera, by simply exchanging the proof goal to the existence of a path satisfying $s = 1$ at the end (compare line 5 in Table I).

If one is interested in proving the continuous assertions of Quartz programs, this can be achieved by replacing the proof goal with the universal path operator as depicted in line 7 in Table I.

IV. Interactive Verification

In this section, we apply the framework as depicted in Figure 2 to an adaptation of the well-known bouncing ball example. While the bouncing ball itself is a standard example with relatively simple continuous dynamics, the parallel composition of a number of balls is difficult to model for most tools as without a suitable compositional semantics the model suffers severely from state space explosion.

Figures 6 and 7 depict the Quartz code of N parallel balls, where N is an arbitrary parameter and the balls start from arbitrary heights. In the main module the observer `n_sum` counts the overall number of bounces of *all* balls, whereas the module `Ball` models an individual ball.

The correctness of the implementation of this observer is proved in two steps: In the first step, the correctness of the counters of each single ball is proven. As the module `Ball` works independently of the rest of the program and is stable under parallel composition, this goal can be achieved by only considering that submodule, disregarding of the number of parallel balls (for more information on the stability of Quartz programs under parallel composition compare [27]). In the second step, the only proof goal is the correctness of the overall observer, i.e. that for all times `n_sum` is defined as the sum of the single observers.

A. Verification of a Single Counter

In Quartz programs, it is possible to have several variable values for the same physical point of time, as the semantics of the language does not only consist of physical but also of logical time. Therefore, the event of the ball hitting the floor cannot simply be described by $h \leq 0$. A good alternative is given by 'the ball reaches the floor during a continuous transition, the starting point may be in the air or directly after a bounce'. The latter part is defined by $\sigma :=$ `h=0 and v>0 or h>0`, whereas reaching the floor during the continuous transition is defined by `cont(h)<=0 and cont(v)<0`. To prove the correctness of the event description, it suffices to show, that during any continuous transition satisfying σ, this condition holds as an invariant for all points except the termination point of the continuous transition. Thus, the event of hitting the floor can only be triggered at the end of but not during the continuous transition. In terms of Quartz, the first property is expressed by σ`-> constrainSM(cont(`σ`))`. The corresponding KeYmaera proof for the only flow statement is depicted in Figure 8.

Define now `NEXTSTEP(`ϕ`)` as a macro for the macro step following one, where a given condition ϕ holds. Then, in a second step the global invariant `NEXTSTEP(h>0 and cont(h)<=0)->next(n)= n+1` is proven. Here, KeYmaera must prove that the release condition `cont(h)<=0 and cont(v)<=0` implies the condition `h>0 and cont(h)<=0`, which is obviously the case and omitted here. The rest of the proof rules will be done by the AIFProver.

```
module Ball(real ? init_h, ? init_v ,int n,) {
   hybrid real h, v;
   h = init_h ; v = init_v ;
   loop{
      flow {
         drv(h) <- cont(v);
         drv(v) <- -9.81;
      } until(cont(h) <= 0 and cont(v) <= 0);
       next(v) = -v/2.0;
       next(n) = n + 1 ;
       flow {} until(true);
   }
}
```

Fig. 6. Hybrid Quartz Module One Ball

```
import BounceBall.*;
macro N = ?;
module NBalls([N]real ? InitH){
    hybrid [N]real h, v;
    [N]int n; int n_sum;
    for(i=0..N-1) {h[i] = InitH(i); v[i] = 0.0; }
    /* n parallel balls */
        {for(i=0..N-1) do || Ball(h[i],v[i],n[i]);}
    ||
        loop{
            n_sum = sum(i =0..N-1 ) n[i];
            pause;
        }
}
```

Fig. 7. Hybrid Quartz Module N Balls

1.	\programVariables {
2.	R h, v ;
3.	R s, t ;
4.	R l_1 ,l_2 ;
5.	}
6.	\problem {
7.	$(\sigma) \wedge s = -1$
8.	$\rightarrow \setminus [$
9.	$\{h' = v, v' = -9.8, t' = 1, h \geq 0 \vee v \geq 0\};$
10.	if$(h = 0 \wedge v \leq 0)$
11.	then $l_1 := 1$; $l_2 := 0$; $s := 1$
12.	else $s := 0$
13.	fi
14.	$\setminus]$ (($s = 0$) $\rightarrow$ ((σ))
15.	}

Fig. 8. KeYmaera Program for only Flow Statement

B. Verification of the Overall Counter

The verification of the overall counter is very simple, as the single counters work correctly. As the counters are discrete variables, it suffices to show that it holds globally `n_sum = n[1] + ... + n[N]`, which is easily done by the AIFProver.

V. Conclusion

Quartz is a powerful synchronous language for the modelling of cyber-physical systems with non-trivial discrete dynamics. This language provides possibilities for the interactive verification of purely discrete programs based on the program structure. In this paper, we presented how to combine the modelling and discrete verification capabilities of Averest with the verification capabilities of KeYmaera. To that end, we interfaced KeYmaera and Averest and showed the capabilities of this tool combination by interactively verifying a non-trivial example.

References

[1] O. Grumberg and H. Veith, Eds., *25 Years of Model Checking – History, Achievements, Perspectives*, ser. LNCS, vol. 5000. Springer, 2008.

[2] A. Fehnker, E. Clarke, S. Kumar Jha, and B. Krogh, "Refining abstractions of hybrid systems using counterexample fragments," in *Hybrid Systems: Computation and Control (HSCC)*, ser. LNCS, M. Morari and L. Thiele, Eds., vol. 3414. Zurich, Switzerland: Springer, 2005, pp. 242–257.

[3] T. Dzetkulic and S. Ratschan, "Incremental computation of succinct abstractions for hybrid systems," in *Formal Modeling and Analysis of Timed Systems (FORMATS)*, ser. LNCS, U. Fahrenberg and S. Tripakis, Eds., vol. 6919. Aalborg, Denmark: Springer, 2011, pp. 271–285.

[4] K. Bauer, R. Gentilini, and K. Schneider, "A uniform approach to three-valued semantics for mu-calculus on abstractions of hybrid automata," in *Haifa Verification Conference (HVC)*, ser. LNCS, H. Chockler and A. Hu, Eds., vol. 5394. Haifa, Israel: Springer, 2009, pp. 38–52.

[5] G. Frehse, "PHAVer: Algorithmic verification of hybrid systems past HyTech," in *Hybrid Systems: Computation and Control (HSCC)*, ser. LNCS, M. Morari and L. Thiele, Eds., vol. 3414. Zurich, Switzerland: Springer, 2005, pp. 258–273.

[6] T. Henzinger, P.-H. Ho, and H. Wong-Toi, "HYTECH: a model checker for hybrid systems," *Software Tools for Technology Transfer (STTT)*, vol. 1, no. 1-2, pp. 110–122, December 1997.

[7] F. Kratz, O. Sokolsky, G. Pappas, and I. Lee, "R-Charon, a modeling language for reconfigurable hybrid systems," in *Hybrid Systems: Computation and Control (HSCC)*, ser. LNCS, J. Hespanha and A. Tiwari, Eds., vol. 3927. Santa Barbara, California, USA: Springer, 2006, pp. 392–406.

[8] R. Alur, R. Grosu, Y. Hur, V. Kumar, and I. Lee, "Modular specification of hybrid systems in Charon," in *Hybrid Systems: Computation and Control (HSCC)*, ser. LNCS, N. Lynch and B. Krogh, Eds., vol. 1790. Pittsburgh, Pennsylvania, USA: Springer, 2000, pp. 6–19.

[9] X. Briand and B. Jeannet, "Combining control and data abstraction in the verification of hybrid systems," in *Formal Methods and Models for Codesign (MEMOCODE)*, R. Bloem and P. Schaumont, Eds. Cambridge, Massachusetts, USA: IEEE Computer Society, 2009, pp. 141–150.

[10] M. Fränzle and C. Herde, "HySAT: An efficient proof engine for bounded model checking of hybrid systems," *Formal Methods in System Design (FMSD)*, vol. 30, pp. 179–198, 2007.

[11] G. Audemard, M. Bozzano, A. Cimatti, and R. Sebastiani, "Verifying industrial hybrid systems with MathSAT," *Electronic Notes in Theoretical Computer Science (ENTCS)*, vol. 119, pp. 17–32, 2005.

[12] R. Bruttomesso, A. Cimatti, A. Franzén, A. Griggio, and R. Sebastiani, "The MathSAT 4SMT solver," in *Computer Aided Verification (CAV)*, ser. LNCS, A. Gupta and S. Malik, Eds., vol. 5123. Princeton, New Jersey, USA: Springer, 2008, pp. 299–303.

[13] L. Bu, Y. Li, L. Wang, and X. Li, "BACH: Bounded reachability checker for linear hybrid automata," in *Formal Methods in Computer-Aided Design (FMCAD)*. Portland, Oregon, USA: IEEE Computer Society, 2008, pp. 1–4.

[14] L. Bu, Y. Li, L. Wang, X. Chen, and X. Li, "BACH 2: Bounded ReachAbility CHecker for compositional linear hybrid systems," in *Design, Automation and Test in Europe (DATE)*. Dresden, Germany: EDA Consortium, 2010, pp. 1512–1517.

[15] M. Sabry, A. Sridhar, D. Atienza, Y. Temiz, Y. Leblebici, S. Szczukiewicz, N. Borhani, J. Thome, T. Brunschwiler, and B. Michel, "Towards thermally-aware design of 3D MPSoCs with inter-tier cooling," in *Design, Automation and Test in Europe (DATE)*. Grenoble, France: IEEE Computer Society, 2011, pp. 1466–1471.

[16] A. Tiwari, "HybridSAL relational abstracter," in *Computer Aided Verification (CAV)*, ser. LNCS, P. Madhusudan and S. Seshia, Eds., vol. 7358. Berkeley, California, USA: Springer, 2012, pp. 725–731.

[17] L. de Moura, S. Owre, H. Rueß, J. Rushby, N. Shankar, M. Sorea, and A. Tiwari, "SAL 2," in *Computer Aided Verification (CAV)*, ser. LNCS, R. Alur and D. Peled, Eds., vol. 3114. Boston, Massachusetts, USA: Springer, 2004, pp. 496–500.

[18] M. Gordon, "HOL: A machine oriented formulation of higher order logic," Computer Laboratory, University of Cambridge, Tech. Rep. 68, May 1985.

[19] N. Völker, "Towards a HOL framework for the deductive analysis of hybrid control systems," in *Automation of Mixed Processes: Hybrid Dynamic Systems (ADPM)*, S. Engell, S. Kowalewski, and J. Zaytoon, Eds. Shaker, 2000, pp. 243–250.

[20] T. Mhamdi and S. Tahar, "Providing automated verification in HOL using MDGs," in *Automated Technology for Verification and Analysis (ATVA)*, ser. LNCS, F. Wang, Ed., vol. 3299. Taipei, Taiwan: Springer, 2004, pp. 278–293.

[21] Z. Manna and H. Sipma, "Deductive verification of hybrid systems using STeP," in *Hybrid Systems: Computation and Control (HSCC)*, ser. LNCS, T. Henzinger and S. Sastry, Eds., vol. 1386. Berkeley, California, USA: Springer, 1998, pp. 305–318.

[22] N. Bjørner, A. Browne, E. Chang, M. Colon, A. Kapur, Z. Manna, H. Sipma, and T. Uribe, "STeP: deductive-algorithmic verification of reactive and real-time systems," in *Computer Aided Verification (CAV)*, ser. LNCS, R. Alur and T. Henzinger, Eds., vol. 1102. New Brunswick, New Jersey, USA: Springer, 1996, pp. 415–418.

[23] E. Abraham-Mumm, U. Hannemann, and M. Steffen, "Verification of hybrid systems: formalization and proof rules in PVS," in *International Conference on Engineering of Complex Computer Systems*. Skovde, Sweden: IEEE Computer Society, 2001, pp. 48–57.

[24] S. Owre, J. Rushby, and N. Shankar, "PVS: A prototype verification system," in *Conference on Automated Deduction (CADE)*, ser. LNCS, D. Kapur, Ed., vol. 607. Saratoga Springs, New York, USA: Springer, 1992, pp. 748–752.

[25] A. Platzer and J.-D. Quesel, "KeYmaera: A hybrid theorem prover for hybrid systems (system description)," in *International Joint Conference on Automated Reasoning (IJCAR)*, ser. LNCS, A. Armando, P. Baumgartner, and G. Dowek, Eds., vol. 5195. Sydney, New South Wales, Australia: Springer, 2008, pp. 171–178.

[26] A. Platzer, *Logical Analysis of Hybrid Systems – Proving Theorems for Complex Dynamics*. Springer, 2010.

[27] K. Bauer, "A new modelling language for cyber-physical systems," Ph.D. dissertation, Department of Computer Science, University of Kaiserslautern, Germany, Kaiserslautern, Germany, January 2012.

[28] M. Gesell and K. Schneider, "Interactive verification of synchronous systems," in *Formal Methods and Models for Codesign (MEMOCODE)*, S. Shukla, L. Carloni, D. Kroening, and J. Brandt, Eds. Arlington, Virginia, USA: ACM, 2012, pp. 75–84.

[29] K. Schneider, "The synchronous programming language Quartz," Department of Computer Science, University of Kaiserslautern, Kaiserslautern, Germany, Internal Report 375, December 2009.

[30] A. Platzer and J. Quesel, "European train control system: A case study in formal verification," in *International Conference on Formal Engineering Methods (ICFEM)*, ser. LNCS, K. Breitman and A. Cavalcanti, Eds., vol. 5885. Rio de Janeiro, Brazil: Springer, 2009, pp. 246–265.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1401–1406

Inter-Domain Requirements and their Future Realisability: The ARAMiS Cyber-Physical Systems Scenario

Birgit Penzenstadler, Jonas Eckhardt, Wolfgang Schwitzer
Technische Universität München
Munich, Germany
{penzenst|eckharjo|schwitze}@in.tum.de

María Victoria Cengarle, Sebastian Voss
Fortiss GmbH
Munich, Germany
{cengarle|voss}@fortiss.org

***Abstract*—Systems whose functionality and services span over multiple, interconnected application domains have become known as cyber-physical system (CPS) and currently receive much attention in research and practice. So far, CPS still come with a variety of development-process-related and technical challenges. These challenges include the interaction between the different domain-specific systems and possible conflicts between their requirements, as well as the choice of appropriate modelling concepts.**

This paper makes two main contributions: First, we show how such an inter-domain development-process can be structured, beginning with a a model-based requirements engineering approach. In order to illustrate the concepts, this paper provides a continuous example scenario, developed within a group of the respective domain experts, that outlines the future of mobility using technologies currently under development in the ARAMiS project. The intention is to allow for an analysis of interaction and possible interference between domain-specific scenarios as well as the analysis of the relation between derived domain-specific scenarios and the global, cross-domain scenario. Second, we provide an analysis of the realisability of the scenario steps according to a set of quality criteria and estimate the respective time horizon, derived from interviews with experts from different domains.

The described scenario allows the reification of goals and requirements of CPS for the mobility domain. Moreover, it makes apparent the need for connecting CPS of different domains. Our validation research provides an accompanying resource for future analysis of the interaction between domains and the relation between their requirements as well as teaching requirements engineering in the domain of CPS.

I. Introduction

ADVANCED features in the mobility domains of automotive, avionics and railway require high-performance computing technologies for complex processing or increased networking, as current technologies used in control devices run up against their performance limit. Future control units will have to perform a greater number of more elaborate functions simultaneously.

One class of such systems with the challenge of integrating different system types are cyber-physical systems (CPS). CPS are integrations of computation and physical processes. Lee [6] defines CPS' as embedded computers and networks that monitor and control physical processes, usually with feedback loops where physical processes affect computations and vice versa. The functionality provided by CPS enables us to realize complex business processes, or complex logistic services as in world-wide travelling.

a) Problem: Today there are typically no or only few shared domain-spanning development artefacts. As a result, domain-spanning RE artefacts are not or only in few cases documented. This results in system functionality which may be adequate for the individual domains, yet cross-domain topics of interests and analyses of interaction and mutual interference between multiple domains are neglected. Furthermore, as domain-specific requirements are located in each individual domain, there is no possibility to link requirements between different domains or associate a common rationale on the CPS level.

b) Contribution: Our comprehensive CPS scenario spans over the relevant mobility domains (automotive, railway, avionics) such that cross-domain topics of interest permit the analysis of interaction and mutual interference possibly arising between domain-specific scenarios as well as the analysis of the interplay of each domain-specific scenario with the global, cross-domain scenario. Furthermore, we provide an analysis of the realisability of the scenario steps according to a set of quality criteria and estimate the respective time horizon.

II. Background & Related Work

A. Systems of Systems (SoS) & Cyber-Physical Systems

The software engineering community has developed methodologies to cope with the engineering process of large systems. The term **S**ystem **o**f **S**ystems (SoS) was introduced to characterize such large systems. Shenhar[10] defines SoS as "a large widespread collection or network of systems functioning together to achieve a common purpose". SoS thus stand out because of their composed nature, their large scale, their decentralized control mechanism, their evolving environments, and their large number of stakeholders.

One class of systems of systems with the additional challenge of integrating different system types are cyber-physical systems (CPS). CPS are integrations of computation and physical processes. Lee [6] defines CPS' as embedded computers and networks that monitor and control physical processes,

978-83-60810-53-8/$25.00

usually with feedback loops where physical processes affect computations and vice versa. This leads to complex functionality that spans a variety of application domains. A helpful overview with a body of knowledge and links to further reading is provided on http://cyberphysicalsystems.org/.

From a more technical point of view, [3] characterises a CPS as system with embedded systems, which may directly record physical data using sensors and affect physical processes, evaluate and save recorded data, is connected with one another and in global networks via digital communication facilities and uses globally available data and services.

Vincentelli et al. [9] and Gezgin et al. [4] discuss the challenges of designing CPS and propose to use contract-based design. Dillon et al. [2] present a case study that presents a framework to link a CPS to the web of things. Lin et al. [7] offer a case study on intelligent water distribution by the integrated simulation of CPS. Huang et al. [5] perform a case study on CPS for real-time hybrid structural testing.

All of these works focus on design and/or implementation instead of requirements and do not provide a case study that reflects and describes the complexity of a large CPS. This is the gap the paper at hand intends to fill.

B. The ARAMiS project

The German academy of technical sciences (acatech) has recently completed a study on the perspectives in CPS research, development, and application [3]. This study serves as scientific basis for the publicly funded research project ARAMiS: Automotive, Railway, and Avionics in Multicore Systems (see http://www.projekt-aramis.de/). The main goal of ARAMIS is to provide for the technological basis for improving safety, efficiency, and comfort in the mobility domains of automotive, railway, and avionics by using multicore technology. The insights gained in the project build the indispensable foundation for the successful integration of embedded systems to cyber-physical systems. The structure and decomposition of the ARAMiS systems of systems, the CPS', is depicted in Fig. 1.

III. The ARAMiS CPS Case Study

A. Methodical Approach

The CPS scenario was developed in a combination of top-down and bottom-up approaches in the following phases: We sketched the scenario in a creative workshop and described the initial storyline (top-down). Then the scenario was reviewed in various workshops with domain experts and the storyline was extended with domain-specific contents (bottom-up). In the next phase, the scenario was specified according to the project-wide reference artefact model [8]. In another series of workshops, an assessment scheme was defined to evaluate the realizability of the individual scenario steps and the assessment was performed by a group of domain experts. Finally, concrete requirements were derived from the scenario steps and the assessment results to provide a rationale and an explicit relation to the domain-specific case studies.

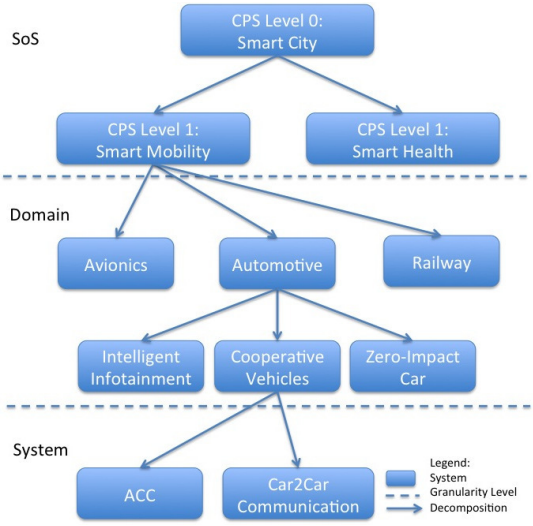


Fig. 1. The ARAMiS structuring and decomposition of SoS

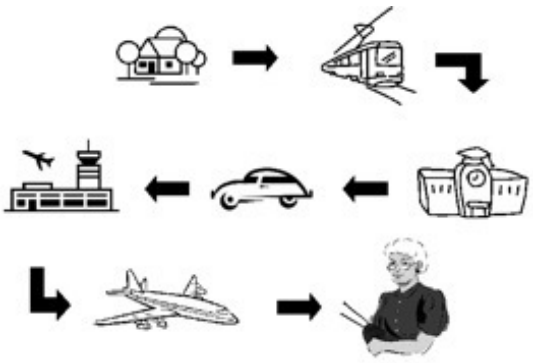


Fig. 2. Overview of the journey

B. Scenario Description

The scenario's starting situation is as follows: Ms Rosemarie Weber plans to spend the next Christmas break with her two children at her mother's, Ms Pauline Mayer. The Weber family lives in Munich, Ms Mayer lives in Sandvika near Oslo. Ms Weber's intention is to pick up her children from school and from there to travel directly to her mother.

Ms Weber enters departure time as well as from and to locations, a maximum cost amount for the entire route as well as passengers' names in the Travel Management Service (TMS) of her smart device. The mobile device is connected to various providers and to Ms Weber's private cloud, and makes suggestions for the trip. In the following, the individual steps of the envisioned scenario are described.

1) Leaving Home: Ms Weber accepts the TMS's suggestion with the proviso that the car be hybrid and capable of autonomous driving. The TMS issues a ticket for Ms Weber's ride in the urban railway, a car reservation according to her preferences, and three flight tickets from Munich to Oslo; name and age of the passengers as well as Ms Weber's possibly further preferences are stored in the cloud.

2) Local transportation (from home to school): Shortly before departure Ms Weber gets a notice on her handheld device about the current status of the local transportation train. As the train is delayed, she takes the opportunity to call and have a little chat with her mother. Afterwards she leaves her home; as she and her children will be away longer, the home is automatically locked, energy saving mechanisms of all devices are enabled, lights are switched off and the home security is activated. Finally she reaches her train on time.

3) At School: Due to the cancellation of the day's last lesson, Ms Weber's children are allowed to go earlier to their day-care centre nearby. This occurred once Ms Weber already set off to school, and she is informed via her smart device of the new location where to pick up her children. At the day-care centre, the children join their respective project teams, organized to collaboratively do their homework. The younger child's group is not yet done with the homework by the time Ms Weber arrives at the day-care centre. Spontaneously, the group members decide to stay a little longer and complete their task. Given that Ms Weber and her children have a plane to catch, the homework group resolves to keep in touch with the leaving child by means of the videoconferencing support put at disposal by the infrastructure.

4) Car-Sharing (from school to airport): For the route connecting the school with the airport the TMS booked an e-mobility car of a car-sharing provider. Ms Weber picks up her dedicated car in front of the school. The car has her driver profile already preloaded, so that the seat and entertainment system is automatically adjusted to her preferences. In addition, the discussion of her child's homework group is streamed to the in-vehicle infotainment (IVI) system and distributed to the corresponding rear-seat screen. As Ms Weber and the two children enter the car, the navigation system starts and suggests the most efficient route to the airport. The IVI offers Ms Weber to book the "premium lane" on the autobahn, which includes a guaranteed arrival time at the airport as an option. The car leaves the parking slot automatically and integrates itself in the traffic flow. The traffic lights are taken into account in two ways. On one hand, there is a coarse-grained traffic dynamics reduction that is triggered by the (smart city) backend in communication with all connected cars. On the other hand, there is direct communication between traffic lights and cars that provides fine-tuning with more precise local information, including an analysis of movements in front of the car. During the drive on the autobahn, the car is being automatically alerted by car-to-car communication about an approaching rescue vehicle on the "premium lane". The car informs Ms Weber, immediately changes lanes, and reduces its speed. In this car-to-car communication, the rescue coordination center informs the rescue vehicle about the accident location to ensure quick appearance with the most up-to-date traffic information.

Back on the "premium lane" the car's speed is controlled by the supervisory TMS. The TMS detects unconnected cars and monitors them using cameras. The performance of unconnected cars is taken into particular consideration during traffic control and planning. Suddenly the car in front brakes, but the collision can be avoided as the optimum evasive maneuver is initiated by the TMS and applied to all connected cars. Unconnected cars are considered accordingly in the scenario; see Figure 3. Thereby, the emergency brake application is calculated within the vehicle and the information is forwarded to the backend. By Car-to-X communication, the braking maneuver is also broadcasted to other vehicles in the direct neighbourhood. This information about braking maneuvers is collected in the backend, to issue a general warning to the traffic section in case the traffic is prone to producing a traffic jam or an accident.

Fig. 3. Avoidance of collision

Close to the airport an Unmanned Aerial Vehicle (UAV) registers heavy rain at position "A (48.456303,12.148819)" with wind direction SE. This information is sent to the TMS, which communicates the upcoming weather situation to every intelligent Road-Side Unit (RSU) within a suitable radius. These RSUs collate the information received with the data they locally sense (wind, rain). The TMS analyses if there is a risk of aquaplaning, in which case a number of actions are taken: unconnected cars are warned using traffic signs, cars equipped with advanced navigation systems are informed in real-time through the system, cars with car-to-x (C2X) close-range communication capabilities receive the warning through nearby RSUs (802.11p) and adapt to the slippery road, and autonomous driving convoys lower speed automatically. Ms Weber arrives at the airport; the car stops and parks in front of the departure entrance according to the flight details, which are sent to the car through the TMS and frequently updated. Ms Weber and her children get out of the car, and label and dispatch their luggage using the automatic check-in counter at the entrance. The car drives autonomously to the parking deck for e-mobility cars of the respective car-sharing provider.

5) Flight (Munich towards Oslo): At the gate, the flight is announced and Ms Weber and her children embark the plane and take their reserved seats. After the boarding is completed, the aircraft takes off in the direction of Oslo.

When the plane has reached a certain height and changes to cruise flight mode, the passengers are allowed to use their personal electronic devices to connect to the wireless passenger network on-board. After reading the digital version of the on-board magazine and ordering drinks and duty free perfumes, Ms Weber starts to watch a TV series and the younger child connects to the school working group via a video conference tool using his tablet device and re-joins the video session that was already joined in the car to the airport; the security of the data exchanged is guaranteed. After twenty minutes into the flight, Ms Weber is informed on her personal

smart device that someone rang the bell at her home. She tabs on the notification on the screen and the video and audio signal from her home's door camera is transmitted to her smart device. She informs the calling neighbour that they will be in Norway for the next few days and wishes him a nice holiday season. Meanwhile the pilot notices a warning on the weather RADAR and is informed by air traffic control that there was a major incident on an oil platform with a huge fire and catastrophic leaking into the North Sea. To ensure the safety of the passengers, the flight is dynamically re-routed by SESAR, but can still reach Oslo with the available fuel. The system organizes the new flight routes of all the planes in the airspace and schedules them to safely reach their airports. To warn the passengers of the expected turbulences, the pilot switches on the seatbelt signs in the passenger service unit and makes afterwards an announcement to all cabin loudspeakers in order to inform the passengers of the redirection. Ms Weber's TMS is informed of the redirection and the plane's estimated arrival time in Oslo. Right after the customer service system prompts her if she would like to notify anybody of the incident, Ms Weber receives a call from her mother, Ms Mayer. The old lady has a headache and would prefer not to drive to the airport to pick them up. So instead of using the customer service system, Ms Weber uses the TMS to change her final destination to Sandvika and the system automatically chooses the next available train to Sandvika and reserves seats in the family wagon in order to ensure that Ms Weber and her children are able to reach their final destination.

6) Railway (Oslo to Sandvika): Once Ms Weber and her children occupy their train seats, she discretely discusses via chat with her mother about Christmas presents for the children. The children notice an attendant talking to a senior passenger. The man's pacemaker detected an irregularity in his heart rhythm; therefore, the Telemedicine System (TS) automatically intervenes: It notifies the train personnel as well as the man's cardiologist. The attendant is guided through the immediate actions to be taken by his smart device. An ambulance with the adequate equipment and remedies is sent to the next train stop, which is to be reached within 20 minutes. The man's health is thus properly nursed.

Ms Weber and her children finally arrive at their destination, where the grandmother cheerfully welcomes them.

C. Model of the Scenario

The scenario was modelled according to the ARAMiS artefact model [8] in the tool Enterprise Architect, which is used project-wide for requirements and system modelling. For this purpose, we developed a profile that provides the modeling elements for the defined content items, such that all requirements documents across the project follow the same template structure and use the same elements.

Figures 4 and 5 depict two exemplary illustrations from the model, namely excerpts of the usage model and the functional hierarchy. There are 23 use cases in Figure 4 which describe the journey in detail and about the same number of overall system user functions in Figure 5 which represent the system-sided realization of these use cases. The use cases were also the basis for the realisability assessment in Section III-D. The models were realized in collaboration among the partners from the various domains and detailed further on the respective domain-specific system levels. Further details of the model can be found in [1].

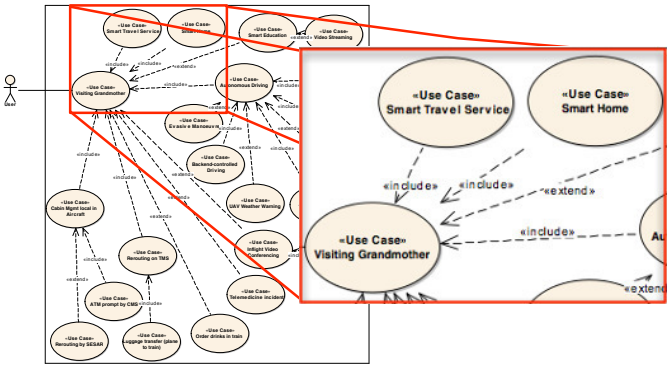


Fig. 4. Use Case Model of the CPS Scenario

D. Technology Realisability Assessments

When describing a future scenario like the CPS scenario, the probably most interesting aspect to assess is the time horizon of the technical realisability of the different parts of the scenario. We performed such an assessment in a number of workshops and in iterations with domain experts. An overview of the results is depicted in Fig. 6 and 7. The domain experts agreed on a list of quality characteristics, for example infrastructure criteria and quality of service criteria, that were relevant to be assessed for judging the realisability of the CPS scenario. We distinguished 3 time horizons (colour-coded in the figures): available today (green), realisable within 5 years (orange), and realisable within 20 years (red). The assessment was performed for all 23 scenario steps in the top row of each table, and for each of the 14 quality characteristics in the first column of the tables. The rationale for each estimation is provided in additional documentation [1]. The time horizon resulting from the justification is coded by colour in the figures for an easy overview of the results. For example, the transmission of a driver profile to a rental car (2nd step "Driver Profil" in the table) and its necessary backend communication (car2backed) is already technically available. This may be implemented in rental cars within the next 5 years (resulting in colour orange), but to actually implement the service, a common data format for driver profiles would have to be standardised among the car manufacturers, which will presumably take considerably longer and is therefore estimated with 10-20 years (resulting in colour red).

IV. Conclusion & Future Work

This paper presented the ARAMiS cyber-physical systems scneario, including the methodical approach for developing the scenario, a storyline description, illustrative excerpts from

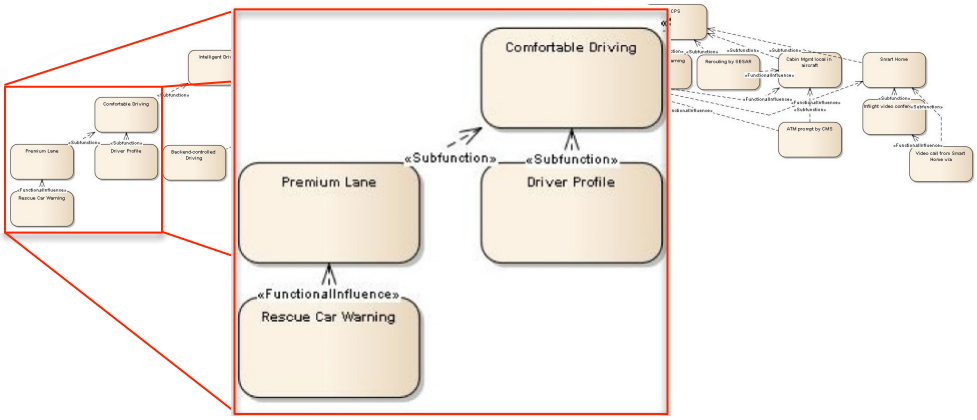


Fig. 5. Functional Hierarchy of the CPS Scenario

Criteria for Realisability		E-mobility										
available today / realisable within 5 years / realisable within 20 years / assumed to be possible		Booking	Driver Profile	Video streaming	Navi-gation	Premium Lane	Autono mous Driving	Backend-controlled Driving	Rescue Car Warning	Evasive Manoeuvre	UAV Weather Warning	Autonomous Car Check-in
Infrastructure	Connectivity											
	Interoperability											
	Portability											
	Performance											
Quality of Service	Security						n.a.			n.a.		n.a.
	Automation											
	Safety		n.a.	n.a.		n.a.						
Reliability												
Availability												
Scalability												
Robustness												
Efficiency												
Adaptivity												
Certifiability												

Fig. 6. Overview of the realisability assessment

Criteria for Realisability					Flight								
		Smart Travel Service	Smart Home	Smart Education	Cabin Mgmt local in aircraft	Inflight video conferencing	Video call from Smart Home	Rerouting by SESAR	ATM prompt by CMS	Rerouting on TMS	Luggage transfer (plane to train)	Order drinks in train	Tele-medicine incident
Infrastructure	Connectivity												
	Interoperability												
	Portability	n.a.			n.a.								
	Performance												
Quality of Service	Security												
	Automation												
	Safety	n.a.											
Reliability													
Availability													
Scalability					n.a.								
Robustness													
Efficiency													
Adaptivity													
Certifiability													

Fig. 7. Overview of the realisability assessment (cont.)

the requirements models, and an overview of the conducted technology assessment. It provides the rationale and the basis for the domain-specific developments in ARAMiS. For the research community, it offers a first available CPS case study of a fictitious system based on real facts targeted for the mobility domain. Therefore, it might serve as input for further research and as a resource for teaching, especially as it might be considered more on the Systems of Systems level, which is a good starting point to educate about CPS. Furthermore, as our paper also provides a preliminary assumption on the scenario parts' technical feasibility in the future, it provides the adequate basis to then go into details with further design-oriented case studies in the respective application domains.

Future Work: The next step is the explicit linking from the domain-specific scenario models back to the overall CPS scenario model in the project-spanning Enterprise Architect repository to allow forward and backward tracing and to provide the traceability for explicit rationale for every requirement in the domain-specific scenarios. The targeted outcome of the project is to provide a showcase that starts with the system of systems scenario at hand and details down to two or three specific use cases in the repsective mobility domains including the demonstrators that show the realization of the prototyped technologies.

Apart from the traceability analysis, we plan evaluation of the quality of the complete model repository to assess the advantages and drawbacks of a cross-domain reference model.

Acknowledgements: This work was funded within the project ARAMiS by the German Federal Ministry for Education and Research with the funding IDs 01IS11035. The responsibility for the content remains with the authors. We would like to thank Lisa Hüfner, Astrid Steingrüber and Jürgen Hairbucher, Oliver Hanka, Stefan Kuntz, and Oliver Sander for helpful feedback.

REFERENCES

[1] María Victoria Cengarle, Jonas Eckhardt, Jürgen Hairbucher, Oliver Hanka, Lisa Hüfner, Stefan Kuntz, Birgit Penzenstadler, Oliver Sander, Wolfgang Schwitzer, and Astrid Steingrüber. The ARAMiS Cyber-Physical Systems Scenario. Technical report, Technische Universität München, 2013. to be published, will be made available to reviewers upon request.

[2] Tharam S. Dillon, Hai Zhuge, Chen Wu, Jaipal Singh, and Elizabeth Chang. Web-of-things framework for cyber-physical systems. *Concurrency and Computation: Practice and Experience*, 23(9):905–923, 2011.

[3] Eva Geisberger, Manfred Broy, María Victoria Cengarle, Patrick Keil, Jürgen Niehaus, Christian Thiel, and Hans-Jürgen Thönnißen-Fries. agendaCPS — Integrierte Forschungsagenda Cyber-Physical Systems. Technical report, acatech — Deutsche Akademie der Technikwissenschaften, 2012.

[4] Tayfun Gezgin, Etzien Christoph, Stefan Henkler, and Achim Rettberg. Towards a rigorous modeling formalism for systems of systems. In *2012 IEEE 15th International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops*, 04 2012.

[5] Huang-Ming Huang, Terry Tidwell, Christopher Gill, Chenyang Lu, Xiuyu Gao, and Shirley Dyke. Cyber-physical systems for real-time hybrid structural testing: a case study. In *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems*, ICCPS '10, pages 69–78, New York, NY, USA, 2010. ACM.

[6] Edward A. Lee. Cyber Physical Systems: Design Challenges. Technical Report UCB/EECS-2008-8, Univ. of California, Berkeley, Jan 2008.

[7] Jing Lin, Sahra Sedigh, and Ann Miller. Towards integrated simulation of cyber-physical systems: A case study on intelligent water distribution. In *Proceedings of the 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, DASC '09, pages 690–695, Washington, DC, USA, 2009. IEEE Computer Society.

[8] Birgit Penzenstadler and Jonas Eckhardt. A Requirements Engineering Content Model for Cyber-physical Systems. *Workshop on Requirements Engineering for Systems, Services and Systems-of-Systems (RESS)*, 2012.

[9] Alberto Sangiovanni-Vincentelli, Werner Damm, and Roberto Passerone. Taming dr. frankenstein: Contract-based design for cyber-physical systems. *European Journal on Control*, May 2012.

[10] Aaron J. Shenhar. A new systems engineering taxonomy. In *Symp. Nat. Council Syst. Eng.*, pages 261–276, 1994.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1407–1413

Safety Analysis of Autonomous Ground Vehicle Optical Systems: Bayesian Belief Networks Approach

Daniel Reyes Duran,
Elliot Robinson,
Andrew J. Kornecki
Embry Riddle Aeronautical University.
Electrical, Computer, Systems and Software Engineering Dept.
Daytona Beach, Florida, USA
Email: {reyes73a, robinse3, kornecka}@erau.edu

Janusz Zalewski
Florida Gulf Coast University
College of Engineering
Dept. of Software Engineering
Ft. Myers, Florida, USA
Email: zalewski@fgcu.edu

***Abstract*—Autonomous Ground Vehicles (AGV) require diverse sensor systems to support the navigation and sense-and-avoid tasks. Two of these systems are discussed in the paper: dual camera-based computer vision (CV) and laser-based detection and ranging (LIDAR). Reliable operation of these optical systems is critical to safety since potential faults or failures could result in mishaps leading to loss of life and property. The paper identifies basic hazards and, using fault tree analysis, the causes and effects of these hazards as related to LIDAR and CV systems. A Bayesian Belief Network approach (BN) supported by automated tool is subsequently used to obtain quantitative probabilistic estimation of system safety.**

I. Introduction

Light Detection and Ranging (LIDAR) combined with dual-camera computer vision (CV) are used as a primary technology for navigation representing a typical optical sensor systems for autonomous ground vehicles (AGV). Researchers at the National Institute for Standards and Technology [1], [2], [3], the U.S. Army [4], and Carnegie-Mellon University [5] have been using such systems to detect obstacles and navigate at ever increasing speeds. Obviously, AGVs combing physical and computing components are typical cyber-physical systems.

AGVs may also be equipped with other navigation technologies such as Inertial Measurement Units (IMU) or Global Positioning System (GPS) receivers; however, the accuracy provided by optically-based navigation controls is absolutely necessary for a safe and precise vehicle operation. Since GPS position errors may be in the range of several meters, GPS alone is not sufficient to safely control a vehicle in urban environment without endangering street signs, pedestrians, and other vehicles. Additionally, GPS reception is obstructed by tall buildings, making GPS unsuitable as a primary navigation tool in an urban environment. IMUs may be used to supplement GPS in low signal areas. However, as kinematic devices, IMUs quickly build up internal error, making them essentially useless for prolonged autonomous navigation. Due to these issues, it is necessary to navigate with a combined system which relies heavily upon optically-based sensor devices.

The paper specific focus is on dependability of the LIDAR and CV sensor systems. It is critical to analyze how the systems work jointly during normal operations and how they work separately under exceptional conditions. To encourage simplicity and maintain focus on the technology, the LIDAR and CV systems are viewed in a generic manner with no effort to model a specific brand of sensor platform. The increasing importance of these issues is realized as fully autonomous vehicles begin to find their way onto roads. Environmental sensing i.e., the capability of the AGV to recognize its location with respect to the environmental obstacles, is the major reason why LIDAR and CV systems are required. Since any unsafe AGV operation may result in violation of safety (property loss or harm to people), there is an evident need for safety analysis. The AGV is safety-critical, software intensive system and potential faults or failures could result in mishaps leading to loss of life and property. By analyzing the hazards posed by the system, the chance of mishap may be reduced or, in some cases, entirely eliminated.

The paper classifies the risk related to AGV operations and describes hazard analyses focusing on impact of LIDAR and CV systems on these operations. Fault Tree Analysis (FT) is used to identify undesirable events and sequences of events leading to top level mishaps such as pedestrian injury, vehicle damage, and external property damage. The Bayesian Belief Network (BN) was modeled based on the FT diagrams along with estimations of likelihood of the events and decision nodes. The presented model can be a good estimator of AGV optical navigation systems as a whole.

II. System Description

A. The AGV Sensor Systems

LIDAR and CV systems are optical sensor systems typically installed on AGV. Together with other sensors they are capable of providing kinematic information about a vehicle (position, velocity, acceleration) and physical information about surroundings (obstacles, road signs, pedestrians, etc). The information from the sensors feeds into a sensor integrator subsystem which filters and integrates data from all vehicle sensors. Detectable anomalies and erroneous data are typically filtered at this stage.

978-83-60810-53-8/$25.00

Having been filtered, the input data are packaged and sent to a state estimator which performs additional filtering and estimates the current state of the vehicle. This state data are then sent to the navigation module which acts as a high-level controller for the individual control algorithms related to degrees of freedom of the vehicle (velocity, heading, etc).

The navigation module (i.e., waypoint manager) handles high level control of the AGV's navigation. In the event of an optical sensor failure, the data passed on to the state estimator will be either corrupt or missing, generating a biased position estimate for the navigation module. The navigation module relies on this data to know where in the world the AGV is with respect to the waypoints, so a simple LIDAR failure could result in the navigation module thinking that the AGV is only 10 meters away from the target when in reality it is 100 meters away.

B. LIDAR

Regardless of measurement technique, all LIDAR units include the following (often redundant) components: laser, lens filter, receiver, power regulator, rotating mirror, position encoder, and onboard processors. As an example, Fig. 1 shows a LIDAR unit (by SICK) with panoramic scanning using rotating a mirror, allowing the laser diode to remain stationary. Detection is accomplished through a complicated combination of synchronizing hardware (including precision motors, and position encoders), and onboard processing capabilities. LIDAR systems typically use a lens filter to block wavelengths of light not identical to that emitted by the laser diode, thus passively reducing interference in the receiver and avoiding the additional complexity, software, circuitry, and cost associated with active filtering. Received laser signals are processed based on this synchronization data to produce a two or three dimensional point cloud. Any error in the system can obviously lead to incorrect depth or position calculations.

Despite being designed for an outdoor use, the high-precision moving parts and optics in modern LIDAR systems are very sensitive to shock. It is important to always place the device at safe, strategic locations around the vehicle. LIDAR should be placed at high clearance locations from the ground, minimizing the amount of vehicle parts obstructing the field of view. Precautions should be taken to protect the device. Foreign-object impact, shock, and vibrations resulting from crashes or rough terrain navigation could cause device failure.

The LIDAR optical filter is one of the most important components of the device. Any damage to the filter will adversely affect measuring accuracy. The LIDAR filter should be protected with a shroud to prevent or reduce impacts and scratches due to vegetation.

C. Computer Vision

Similar to the LIDAR, the CV camera system can produce a two dimensional image using a single camera (or a three dimensional image using dual camera system with two cameras arranged stereoscopically). Using two cameras also allows the failure of a single camera to degrade but not completely void the CV system functionality.

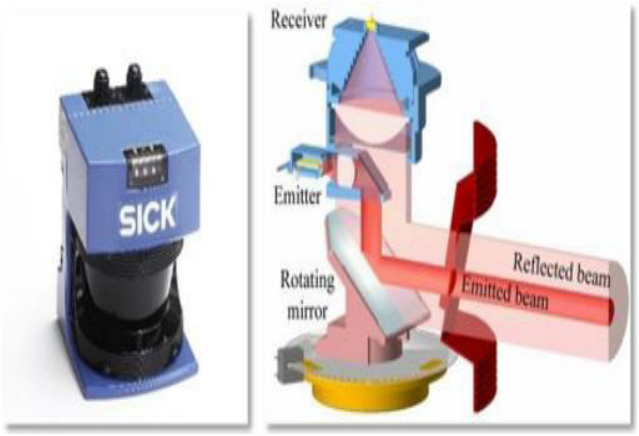


Fig. 1 Example LIDAR

Computer vision software is fundamentally bounded by image quality which is often related to the number of pixels. As each pixel must be processed at least once, the quantities of data and necessary memory may be overwhelming for a system with limited resources. With sufficient processing hardware it is possible to extract quantitative information from scenes, detect obstacles, or track targets using nothing but CV software.

Computer vision algorithms depend upon the video signal received from the camera device. Almost every camera generates some form of distortion which may degrade or even prevent a CV algorithm from operating properly. For this reason, it is necessary to properly calibrate camera and correct image distortion prior to using the image as a source for CV. Improper lighting typically affects both cameras at the same time due to overcast or night condition etc. High-intensity headlights and ambient light sensors on the cameras would be the mitigation technique.

LIDAR failure alone should not significantly degrade system performance. Cameras misalignment would occur if they are displaced by an impact or vibrate free (which can be mitigated with appropriate hardware, e.g., lock washers) and periodic maintenance. Optical receiver misalignment should be extremely rare and can only be caused by manufacturing defect or by physical stress on the device over time (i.e., vibrations from road).

III. Safety Assessment

A. Risks and Hazards

Incorrect operation of AGV may result in mishaps of various severity levels (Table I). One may identify risk as a measure of potential consequence of a hazard representing both the likelihood and the severity of something bad or undesired happening. During the hazard identification stage, hazards are classified according to their risks. A Preliminary Hazard Analysis (PHA) is the starting point to classify these hazards. As with most safety critical systems, the AGV system hazards can be classified in a qualitative manner, using pre-defined arbitrary categories known as risk classes computed as a product of severity and the likelihood of occurrence. For the AGV system, these levels are: negligible

(RV < 1), marginal (1 < RV < 10), critical (10 < RV < 100) and catastrophic (RV > 100).

TABLE I
MISHAP SEVERITY LEVELS

Severity Level	Description
1	No loss of any kind
2	Minor property loss (low cost hardware parts)
3	Major property loss, damage to the environment
4	Loss of critical hardware, human injuries, major damage to the environment
5	Catastrophic loss of life, loss of the entire AGV system, serious environment damage

From a safety standpoint, hazards become the source for safety requirements. Typically, loss of any system functionality may lead to a hazard (e.g., if the laser head of the LIDAR system stops rotating due to mechanical failure). The loss of functionality usually allows identifying a hazard. In turn, the hazard identification allows determining a control measure to be established to prevent or control this hazard. Finally, this control measure can be then converted into a safety requirement for the system and thus be considered in the system development lifecycle.

For the LIDAR example, there is a known hazard of losing the mechanical functionality of the laser rotor head due to wear or manufacturing defects. Typically, engineering department would design and test systems well enough to provide recommendations of conditions for safe operation of their product. Furthermore, manufacturers will typically add recommended maintenance checkups to prevent hazards from transforming into accidents and mishaps. Hazards are always dormant, that is, they exist harmlessly unless certain conditions and/or set of events occur, transforming the hazard into an accident or mishap. Hazards by themselves are not doing any harm unless some transformation takes place. For instance, an energy build-up (e.g., stress due to shaft miss-alignment) would be required for the electric motor that rotates the laser head to eventually give up and stop working. This process takes time. As time passes, wear and stress builds up on the weakest motor parts (with more critical defects). Eventually there will be enough wear and stress accumulated that a point of no return will be reached and a triggering event occurs. That is, at this point there is nothing else that can be done to prevent this hazard from transforming into some accident or mishap. As all this is happening, the system safety levels are gradually degrading from an initial "safe and controlled state" to a final "unsafe and uncontrolled state" leading eventually to a mishap. In all cases a hazard is a prerequisite to the final accident or mishap.

Since an AGV system is dependent on both LIDAR and CV subsystems to assure proper navigation and avoidance of obstacles, any hazards of either subsystem constitutes also a hazard for the AGV system. From this perspective, both LIDAR and CV are safety-critical. Any failure of these subsystems could propagate and lead to a disaster with severe consequences for the AGV. Preliminary list of hazards for LIDAR and CV subsystem, including also hazards and their severity levels is shown in Table II.

TABLE II
HAZARDS IDENTIFICATION

Item	Sub-item	Fault Condition	Hazard	Severity Level
LIDAR	Position Encoder	Fails to read position data	Mirror Motor Malfunction	4
	Electrical	Short circuit	Electrical Failure	4
	Electrical	Overvoltage	Electrical Failure	4
	Optical Receiver	Misalignment	Optical Receiver Error	3
	Optical Filter	Damaged	Optical Receiver Error	3
	Mirror Motor	Malfunction	LIDAR Failure	4
			laser Radiation	4
CV	Camera	Misalignment	CV Failure	3
	IR Filter	Missing	CV Failure	3
	Lens	Damaged	CV Failure	3
	Camera	Improper Lighting	CV Failure	3
Navigation Module	N/A	LIDAR Failure	Navigational Failure	4
		CV Failure	Navigational Failure	4
		Navigational Failure	Property Damage	5
			Vehicle Damage	5
			Pedestrian Injury	5

The most critical sequence of events is one that eventually leads to top level mishaps in the AGV system. Top level mishaps typically relate to loss of life, property or severe damage to the environment. The main goal of Safety Engineering is to prevent these mishaps from happening.

Generally, mishaps are not caused by single events. The accidents are almost always caused by a sequence of events that eventually take the system to an unstable and unsafe state causing the mishap. In the case of the AGV system there are three major top level mishaps: Critical Vehicle Damage, Pedestrian Injury, and Other Property Damage. Critical Vehicle Damage refers to a sequence of events leading to an accident in which the AGV is lost. This type of mishaps results from uncontrolled travel that leads to a physical contact involving substantial volume of kinetic energy between the AGV system and the physical world (i.e., crash or collision) which results in property loss. Pedestrian injury or fatality is by far the most undesirable top level mishap that may be caused by an uncontrolled travel of the AGV system. Other property damage refers to loss of property caused to third parties that are not a part of the AGV system. For instance, during collision with another vehicle the AGV system may cause damage to the other vehicle. All three above mentioned mishaps can be categorized at the highest severity level.

B. Fault-Tree Analysis

Based on the preliminary hazard identification presented in Table II, we may use Fault Tree Analysis (FTA) to show the chain of events that may lead to mishaps. FTA allows us not only to identify the set of events leading to top level mishaps but also determine the intermediate events that constitute a cause-effect chain [6].

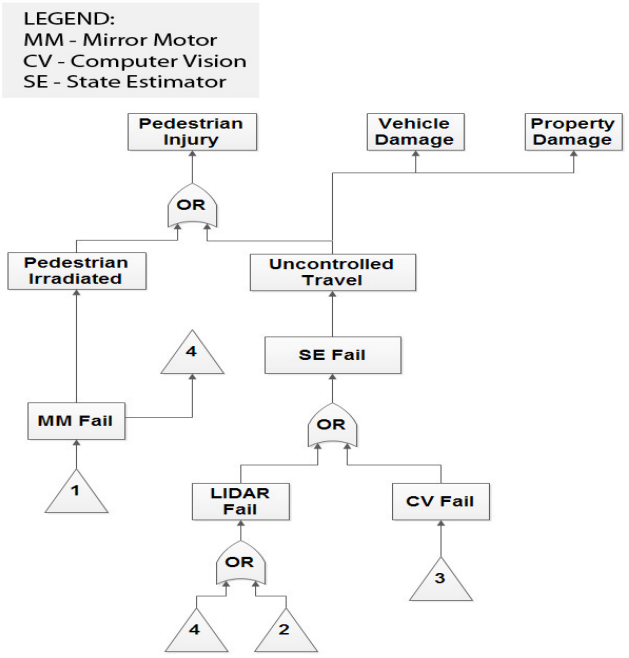


Figure 2 Top-level FTA for UGV

Fig. 2 presents top-level fault tree identifying three top-level mishaps and the LIDAR/CV contribution to these mishaps. LIDAR and CV subsystems fault trees are presented in Fig. 3 and 4. As Portinale [7] observed: "*Any FT can be transformed into a corresponding BN, by creating a binary BN node for each event in the FT, and by setting the probability of BN root nodes (corresponding to basic events in the FT).*" Thus, the cause-effect relations between the events are the basis of subsequent quantitative probabilistic analysis using a Bayesian Belief Network.

The FTA analyses show that any major mishap will involve the failure of one or more subsystem components. In the case of both subsystem components failing (CV and LIDAR) the resulting behavior of the AGV system will always reach a top level mishap scenario.

IV. Bayesian Belief Networks

A. Background

Bayesian Belief Networks have been widely used in Industrial Information Systems for solving variety of computational problems with insufficient information and excessive uncertainty [6, 8, 9]. Since the 18th century mathematician Rev. Thomas Bayes introduced the concept of updating probabilities based on new information, the method has been widely applied in probability and statistics. The basis for the

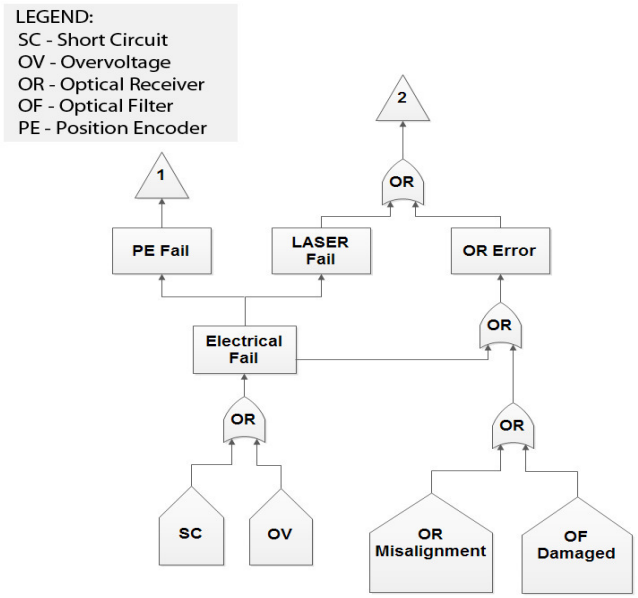


Figure 3 Low-level FTA for LIDAR

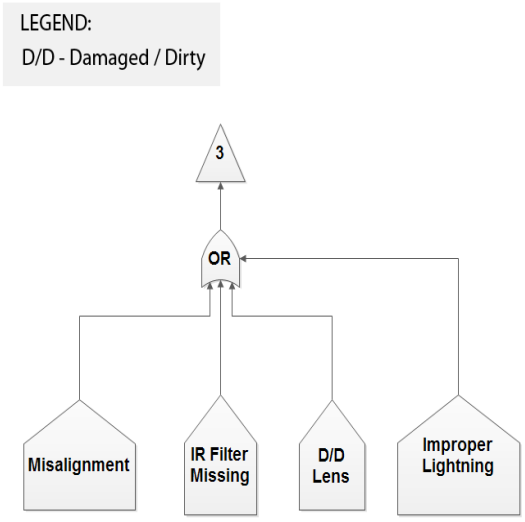


Figure 4 Low-level FTA for CV

method is the inversion formula for belief updating from evidence (E) about a hypothesis (H) using probability measurements of the prior truth of the statement enhanced by posterior evidence:

$$P(H|E) = (P(E|H)*P(H)) / P(E) \qquad (1)$$

A Bayesian belief network is a probabilistic graphical model. The belief network represents the joint probability distribution of a set of random variables with explicit interdependence assumptions. In this research a Bayesian network is defined by a directed acyclic graph of nodes representing variables and arcs representing probabilistic dependency relations among the variables [9].

An arc from node A to another node B indicates that variable B depends directly on variable A. If the variable represented by a node has a known value then the node is said to be observed as an evidence node. A node can represent any kind of variable, be it a measured parameter, a latent variable, or a hypothesis. Nodes are not restricted to representing random variables: this is what "Bayesian" is about a belief network.

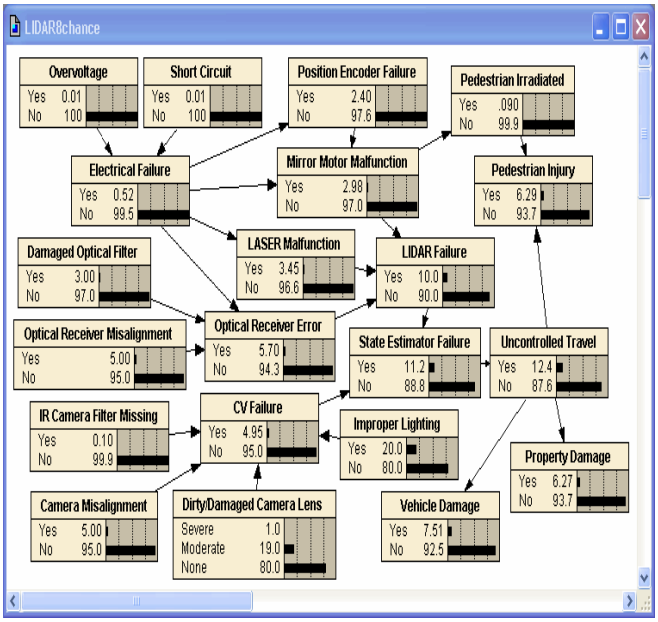


Figure 5 BN of the AGV System – a nominal scenario

The approach supports three types of reasoning. Predictive reasoning observes the causal evidence and updates the middle and upper layer nodes reasoning from a cause to the effects. Diagnostic reasoning observes the evidence of effects and updates the middle and the bottom layer causal variables, reasoning from an effect to the cause. The BN's also allow explanatory (inter-causal) reasoning, in which middle layer reasoning evidence is used to update both the causal and the effect variables.

B. Preliminary Modeling

There is a variety of tools supporting BN modeling: www.dsl-lab.org/ml_tutorial/software_bayesian_networks.html. The computations were done using Bayesian Networks generated by the tool Netica [10]. Based on the fault tree diagrams, along with assumption of base events likelihood (leaf nodes) and the conditional probabilities, it was possible to create a model which represents a good estimator of AGV optical navigation systems dependability. Using nominal likelihood values based on the system analyses and the available data, i.e., assuming no deterministic evidence about the status of the system components, we were able to assess the likelihood of top level mishaps and thus identify their criticality. Fig. 5 presents a screenshot of the tool in such nominal scenario.

From this nominal scenario the BN allows to introduce evidence of selected events and analyze the impact of this evidence on other events. The predictive reasoning property of BN allows us to introduce the evidence of base events (as an example: a camera misalignment) and observe the impact on intermediate events and ultimately on the top level mishaps (Fig. 6).

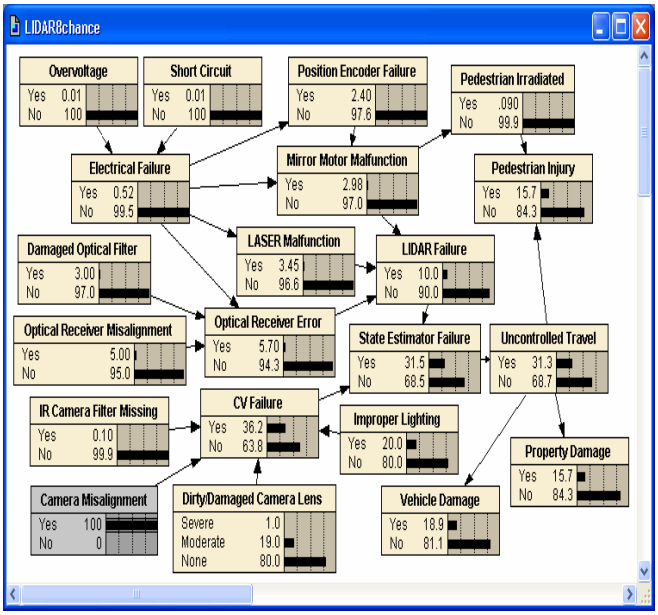


Figure 6 BN predictive reasoning – evidence of camera misalignment

Another scenario allows the presentation of the inter-causal reasoning property of the BN, i.e., analyzing impact of known evidence of intermediate events (e.g. malfunction of laser, mirror motor, optical receiver) up and down the causal chain. As an example, we show how introducing evidence of LIDAR failure results in over fivefold increase of pedestrian injury and property/vehicle damage probabilities (Fig. 7).

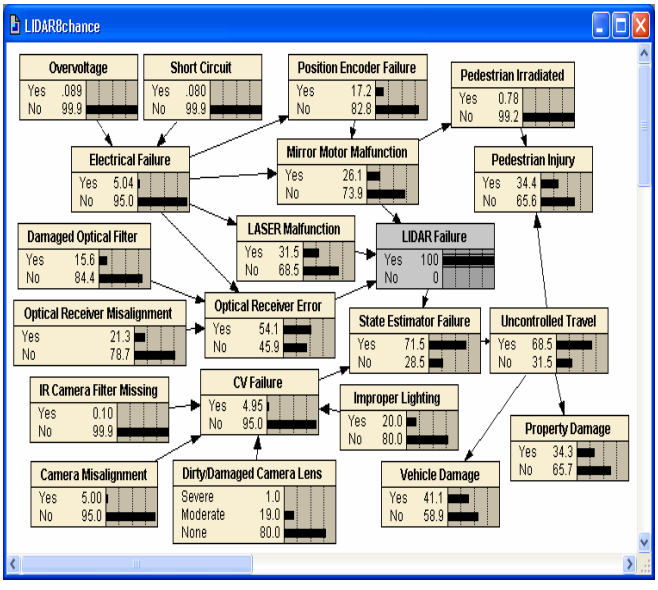


Figure 7 BN inter-causal reasoning – effect of the LIDAR failure

Similarly, the evidence of CV failure results in significant increase to the likelihood of top level mishaps (Fig. 8). Using the inter-causal reasoning one can also assess the poten-

tial causes of the malfunctions observing increased likelihood of causal events such as electrical failure, optical filter damage, or misalignment of the receiver.

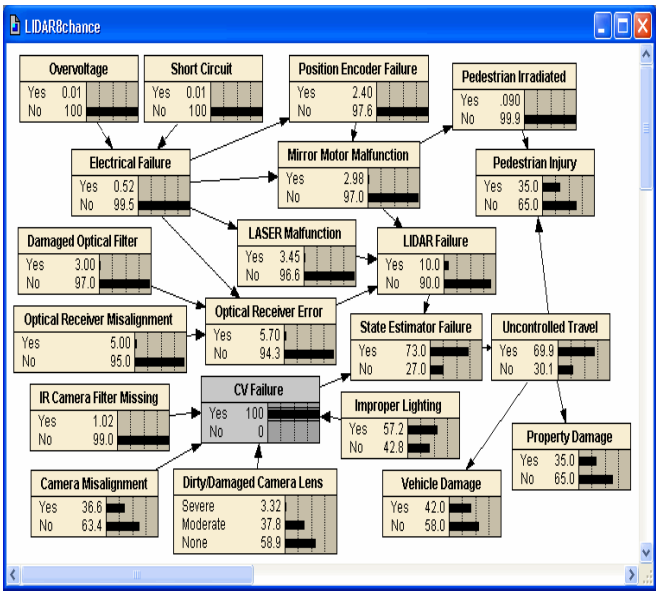


Figure 8 BN inter-causal reasoning – effects of CV failure

B. *Detailed Analysis*

Subsequently, a variety of scenarios were attempted to identify the impact of specific events and the criticality of top level mishaps. The model base probabilities are the best reasonable estimate numbers. Due to uncertainty built into the model, the top-level mishaps show relatively high likelihood of occurrence even with the evidence of correct operation and lack of any problems. The model thus implicitly accounts for unidentified hardware failures and other potential system defects that may cause uncontrolled travel. Table III presents partial results of the predictive and inter-causal reasoning modeling. In a nominal scenario (when all base nodes probabilities are "unknown" i.e. set to the assumed values based on the system analyses and available data), the probability of pedestrian injury is 6.29%. Given evidence of events such as overvoltage or damaged optical filter allows one to predict 200-400% increase in the likelihood of mishap. However, improper lighting or damaged camera lens increases the probability by less than twofold.

TABLE III
PREDICTIVE AND INTER-CAUSAL REASONING – IMPACT OF THE EVIDENCE ON PEDESTRIAN INJURY

Evidence	Pedestrian Injury Likelihood %	Impact (in relation to evidence unknown)
All base nodes "perfect"	***3.82***	***-39%***
All base nodes "unknown"	***6.29***	***0%***
All base nodes "bad"	***48.20***	***666%***
Improper lighting	9.07	44%
Severe damaged camera lens	9.76	55%
Camera misalignment	15.70	150%
Optical receiver misalignment	16.50	162%
Damaged optical filter	19.50	210%
Mirror motor malfunction	31.00	393%
Overvoltage	31.40	399%
LIDAR failure	34.40	447%
CV failure	35.00	456%
CV and LIDAR failures	47.70	658%

Using the predictive and inter-causal reasoning capabilities of Bayesian networks, it is possible to gain additional insight into improvements. For example, simply reducing the chance of inferior lighting is sufficient to remove catastrophic risks and substantially reduce the number of critical risks.

The BN also allows for a diagnostic reasoning (i.e., from effects to cause). For example, having evidence of pedestrian injury the BN estimates that the probability of mirror motor malfunction grows to over six times, CV failure five times, and state estimator failure nearly eight times its original value (Fig. 9).

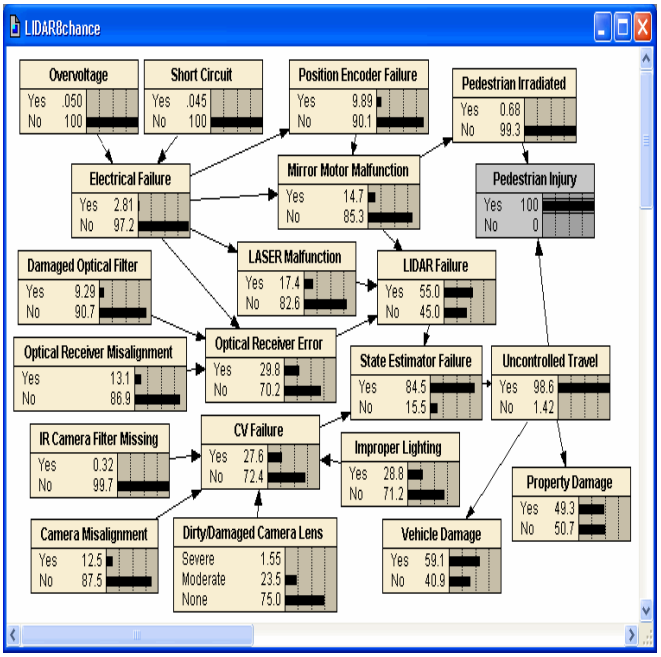


Figure 9 BN diagnostic reasoning – causes of pedestrian injury

Diagnostic reasoning is the most desirable in this research, since it allows making predictions on potential causes of mishaps, including quantitative assessment of risk. This, in turn, makes it possible to prepare for catastrophic events by minimizing their consequences or avoid them by paying closer attention to potential causes.

Using diagnostic reasoning it is possible to derive interesting statistics about the system, such as the rate of property and pedestrian damage in incidents of uncontrolled travel involving vehicle damage. Using available evidence it has been determined that vehicular damage will result in nearly

50% likelihood of property damage and pedestrian injury. It is also possible to estimate that, given the evidence of pedestrian injury, there is a 98.6% chance that it is caused by uncontrolled travel. Having evidence of vehicle damage, the reasoning allows us to estimate the likelihood of LIDAR failure to be 55% and CV failure 27.6%. However, with the evidence of no CV failure the likelihood of LIDAR failure increases to 70.8%. The proposed approach allows thus to analyze impact of given evidence on system in a variety of scenarios.

Table IV presents another partial result of the modeling. The columns present likelihood of the model events in two scenarios: when there is no evidence and when there is evidence of pedestrian injury. The LIDAR and CV failure show as the leading causes of potential pedestrian injury.

TABLE IV
DIAGNOSTIC REASONING – POTENTIAL CAUSES OF PEDESTRIAN INJURY

	No evidence of pedestrian injury	Evidence of pedestrian injury
Electrical failure	0.52	2.81
Damaged optical filter	3.00	9.29
Position encoder failure	2.40	9.89
Camera misalignment	5.00	12.50
Optical receiver misalignment	5.00	13.10
Mirror motor malfunction	2.98	14.70
Laser malfunction	3.45	17.40
CV failure	4.95	27.60
Improper lighting	20.00	28.80
LIDAR failure	10.00	55.00

Interestingly, and in accordance with the hazard table, the BN analysis shows that in the case of total state estimator and thus navigation failure, improper lighting bears a significant probability of being the reason, with CV and LIDAR being evidently on the top. The risk value corresponds equally well to LIDAR failures on the BN, with the laser malfunction as the primary cause. This correlation between the hazard table and the BN implies that the proposed approach provides reasonable base for quantitative assessment of system dependability.

V. CONCLUSIONS

This paper describes the analysis of autonomous ground vehicle system optical navigation components to identify hazards leading to potential safety violations and top level mishaps. We used safety analytical modeling techniques including Fault Tree analysis and Bayesian Belief Networks to better understand the sequence of events that could lead to a major accident or mishap. The quantitative analysis helps to determine the most important hazards that need to be mitigated or controlled. Analysis results confirmed the importance of the reliability and availability of the AGV sensor LIDAR and CV subsystems. Based on the analysis, specific mitigation measures can be recommended in order to reduce the risk of loss of life and/or property. These risk mitigations would lead to reducing the probability for subsystem and system malfunction.

By utilizing both Fault Tree Analysis and Bayesian Belief Networks it is possible to better determine what the sequences of events and their impact on the top level mishaps. From the FTA results it is clear that any major mishap will always involve the failure of one or more subsystem components. In the case of both subsystem components failing (CV and LIDAR) the resulting behavior of the AGV system will always reach a top level mishap scenario. Using predictive reasoning capabilities of Bayesian Networks, it was possible to gain additional insight into the system operation and identify the potential mitigation sources.

Future work would need to assure that the numerical values for the likelihood of events as well as the dependency relations between the nodes closely represent reality. A good source for these values would be published equipment failure rates (e.g. based on military handbook MIL-HDBK-217F 1995) or collected from industry studies related to safety incident rates [11].

REFERENCES

[1] D. Coombs, K. Murphy, A. Lacaze, A. and S. Legowik, "Driving autonomously offroad up to 35 km/h". *Proceedings of the IEEE Intelligent Vehicles Symposium*, Dearborn, Michigan. 2000.

[2] T.H. Hong, M.O. Shneier, C. Rasmussen and T. Chang, "Road detection and tracking for autonomous mobile robots". *SPIE 16th Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, Orlando, Florida. 2002.

[3] C. Rasmussen, "Combining laser range, color, and texture cues for autonomous road following", *IEEE International Conference on Robotics and Automation*, Washington, DC. 2002.

[4] J.A. Bornstein and C.M. Shoemaker, "Army ground robotics research program", *Unmanned Ground Vehicle Technology V*, Orlando, Florida. SPIE Proceedings Series, Volume 5083, pp. 303–310, 2003.

[5] C. Urmson, *Navigation regimes for off-road driving*, Technical Report No. CMU-RI-TR-05-23. Pittsburgh, PA: Carnegie Mellon University, Robotics Institute, 2005.

[6] W. Vesely et al., *Fault Tree Handbook with Aerospace Applications*, NASA Office of Safety and Mission Assurance, August 2002.

[7] L. Portinale, "Bayesian Belief Networks in Reliability", Tutorial Notes, 2012 Annual Reliability and Maintainability Symposium, Reno, NV, URL: http://www.xcdsystem.com/rams2012/cdrom/tutorials/09a.pdf

[8] N.E. Fenton and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*, CRC Press, ISBN: 9781439809105, 2012.

[9] F.V. Jensen and T.D. Nielsen, *Bayesian Networks and Decision Graphs*. Second Edition, Springer-Verlag, 2007.

[10] *Netica Software Package*. Norsys Software Corp., Vancouver, BC. URL: http://www.norsys.com/netica.html.

[11] R. Chalupa, *Failure Modes, Effects and Diagnostics Analysis*. Report No. 06-11-25-R001, Rosemount Corp., Eden Prairie, Minn. 2007.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1415–1422

Towards the Applicability of Alf to Model Cyber-Physical Systems

Alessandro Gerlinger Romero
Brazilian National Institute for Space Research, Avenida dos Astrounautas, 1758, 12227-010, São José dos Campos, São Paulo, Brazil.
Email: romgerale@yahoo.com.br

Klaus Schneider
University of Kaiserslautern Computer Science Department, Po box 3049, 67653, Kaiserslautern, Germany.
Email: klaus.schneider@cs.uni-kl.de

Maurício Gonçalves Vieira Ferreira
Brazilian National Institute for Space Research, Avenida dos Astrounautas, 1758, 12227-010, São José dos Campos, São Paulo, Brazil.
Email: mauricio@ccs.inpe.br

Abstract— Systems engineers use SysML as a vendor-independent language to model Cyber-Physical Systems. However, SysML does not provide an executable form to define behavior but this is needed to detect critical issues as soon as possible. Action Language for Foundational UML (Alf) integrated with SysML can offer some degree of precision. In this paper, we present an Alf specialization that introduces the synchronous-reactive model of computation to SysML, through definition of not explicitly constrained semantics: timing, concurrency, and inter-object communication. The proposed specialization is well-suited for safety-critical systems because it is deterministic. We study one example already modeled in the literature, to compare these approaches with our one. The initial results show that the proposed specialization helps to couple complexity, provides better composition, and enables deterministic behavior definition.

I. Introduction

CYBER-Physical Systems (CPSs) are an integration of computational and physical processes [14]. The difficulty in modeling cyber-physical systems comes from the diversity of these systems. The most promising approach to mitigate this problem is developing expressive and precise modeling languages [8].

Accordingly, the Object Management Group (OMG) and the International Council on Systems Engineering (INCOSE) developed Systems Modeling Language (SysML) [20]; a general-purpose modeling language for systems engineering applications. SysML has demonstrated a capability for top-down design refinement for large-scale systems [11]; therefore, SysML is expressive, but the lack of formal foundations in the SysML results in imprecise models.

A major current focus in systems engineering is how to introduce precision in the approaches based on SysML through formal methods. This introduction can be a legal requirement when dealing with safety-critical systems; e.g., the IEC 61508 (Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems) defines formal methods as highly recommended technique for the highest safety integrity level; moreover, DO-178C (Software Considerations in Airborne Systems and Equipment Certification) addresses formal methods as a complement to testing. There are languages with a formal semantics such as Esterel [5] or the B-language [7]; nonetheless, there are no modeling languages with widespread use in systems engineering community that have the attraction of SysML [10].

This paper focuses on the evaluation of a formal foundation in SysML engineering approaches concerning behavioural definitions. Behavior is defined using SysML, and also using Unified Modeling Language (UML) [18], mainly by Activity Diagrams, Sequence Diagrams, and State Machine Diagrams, which do not have precise semantics given by OMG; and, in general, are not executable.

Behavioural definition could evolve with the Semantics of a Foundational Subset for Executable UML Models (fUML) [19]; this specification defines a formal semantics for an executable subset of UML. Moreover, OMG Action Language for Foundational UML (Alf) is the textual language for fUML [21].

On the contrary, there are research papers [4][22] stating that fUML and Alf are not suitable for behavioural modeling the safety-critical systems yet. The reasons can be classified as follows: (1) nondeterminism in the execution model [4]; and, (2) current tools do not allow the use of model-checking or theorem proving [22]. Hereafter, we will explore the reason (1) in detail.

fUML standard execution model is based on a model of computation (MoC), which is nondeterministic (we consider this in Section III.A). On the other hand, there is one MoC that can provide determinism, and can simplify the modeling and verification tasks; it is called synchronous-reactive [14].

The synchronous-reactive MoC can provide determinism using the fundamental model of time as a sequence of discrete instants and parallel composition as a conjunction of behaviors [3]. This MoC has been established as a technology of choice for specifying, modeling, and verifying real-time embedded applications [3], e.g. Esterel [5], Lustre (as well as, Lustre-based commercial Scade tool) [3], Signal [3], and Quartz [26] are languages also based on this MoC.

This work was supported by the Brazilian Coordination for Enhancement of Higher Education Personnel (CAPES) and German Academic Exchange Service (DAAD).

The synchronous-reactive MoC means that most of the statements are executed in zero time (at least in the idealized model). Synchronous computations consist of a possibly infinite sequence of atomic reactions that are triggered by a global logical clock. In each reaction, all inputs are read and all outputs are computed by all components in parallel. In the synchronous-reactive MoC, the communication and computation of values is done in zero time. Consumption of time must be explicitly defined with special statements, as e.g. the pause statement in Esterel [5] and Quartz [26].

Comparing a system described in the synchronous-reactive MoC against a system described following an asynchronous MoC for dual redundant flight guidance system, Miller et. al. [15] made the following observation: "the properties themselves are more difficult to state, were weaker than could be achieved in the synchronous case, and required considerable complexity to be added to the model to ensure that even the weakened properties were true". Furthermore, systems described by a synchronous-reactive MoC can be desynchronized [3] in a concrete solution that is then asynchronous, e.g. to generate Globally Asynchronous Locally Synchronous architectures (GALS) [15].

In this paper, we explore the causes of nondeterminism in fUML and Alf, and, present a deterministic specialization of Alf for CPSs modeling based on the synchronous-reactive MoC. This specialization removes deficiencies found by [2] [4] in fUML and Alf, and can be an alternative to define deterministic behaviors in SysML. The initial results show that the proposed specialization does not add complexity to the task of modeling CPSs using SysML, and enables a deterministic definition of the behavior.

The remainder of this paper is organized as follows: in Section II, related works are explored; in Section III, the relationships between Alf and other OMG specifications are explored; in Section IV, we present the initial approach; in Section V, we discuss the initial approach; finally, conclusions are shared in the last section.

II. Related Works

There is a large number of research papers about semantics for models defined using UML, and consequently, SysML. Hußmann [12] proposed the following classification for approaches concerning structural semantics: (1) naive set-theory, (2) meta-modeling, and (3) translation. This classification can also be used for the works focused on behavioural semantics.

Extending naive set-theory, Graves and Bijan [11] proposed one approach where behavior defined using SysML State Machine Diagrams is represented as a set of axioms in type theory. Graves [10] stated that SysML uses diagrams to model structure, and these diagrams can be encoded as axiom sets in OWL (Web Ontology language). The last work did not cover behavioural modeling, but it suggested that behavioural modeling should follow the same path of the structural modeling, i.e. behavior should be encoded as sets of axioms.

Alf [21], and the foundational subset for executable UML models (fUML) [19], combines the meta-modeling and an extension of set-theory, because the semantics of behavior is described operationally by fUML itself, and by a set of axioms (we consider this in Section III).

A broad set of researches adheres to translation through definition of a mapping between SysML and a formal language. Bousse et. al. [7] proposed a transformation from a subset of SysML into a subset of the B method; the selected subset of SysML covers behavioural definitions expressed by Alf. Afterwards, the resulting B method representation is proved by a specialized tool. Pétin et. al. [23] defined transformation from SysML requirements and SysML behavior (defined by State Machine Diagrams and Activity Diagrams, without use of fUML) into temporal logic and timed automata, respectively. Henceforth, the UPPAAL model checker is used to check safety requirements. Abdelhalim et. al. [1] defined a method that receiving State Machine Diagrams and Activity Diagrams (according to fUML) applies a transformation to Communicating Sequential Processes (CSP). Later, the method uses a model checker to verify the resulting CSP representation. This work focuses on maintaining the behavioural consistency between State Machine Diagrams and Activity Diagrams. Abdelhalim et. al. [2] refined their initial approach defining a subset of CSP to be used because difficulties emerge when non-trivial fUML inter-object communication mechanism is formalized. This work identifies patterns that are correct from the modeller's point of view and the system representation; however, when model checking the CSP representation of this model is performed, a state space explosion problem may occur. Perseil [22] suggested that a subset of Alf should be translated to PlusCal, which has precise semantics defined by a translation to TLA (Temporal Logic of Actions); later, the model checker from TLA would be used.

Some degree of semantics for models is a prerequisite for verification. Taking into account verification, there are a large number of research papers about the verification of UML, and consequently SysML, behavioural models, focusing on State Machine Diagrams, Sequence Diagrams and Activity Diagrams; nonetheless, a way to check the correctness of behavioural representations is still not agreed [24]. Planas et. al. [24] presented a method to verify correctness of behaviors defined using Alf through analysis of all possible execution paths. This method uses as input an UML model, and performs its checks directly on this model. This work states that translating UML behavioural models into other formalisms or languages could compromise scalability of these proposed methods.

However, few researches addressed the problem of nondeterminism, and its roots, in behavioural representations using fUML, and Alf.

Benyahia et. al. [4] showed that fUML, and also Alf, is not directly feasible to safety-critical systems because the MoC defined in the fUML execution model is nondeterministic. In spite of variation points provided by fUML, this work recognized that they are not powerful enough to change the MoC, and an alternative extension of the core execution model was presented to accommodate different MoCs.

III. OMG Specifications and MoCs

Execution and verification of models is the cornerstone of any Model-Driven Development (MDD). One prominent alternative for MDD is Model-Driven Architecture (MDA) [17] established by OMG. MDA defines three levels of abstraction: (1) Computational Independent Model (CIM) – focuses on the environment of the mission and mission's requirements; (2) Platform Independent Model (PIM) – defines requirements, structure and behavior for candidate abstract solutions; (3) PSM (Platform Specification Model) – describes concrete solutions.

An important OMG specification for PIM is Alf [21]. Alf is a textual surface representation for UML modeling elements. It is an action language that includes primitive types (including real numbers), primitive actions (e.g. assignments), and control flow mechanisms, among others. It is object-oriented, and it is an imperative language (like C and Java). Further, Alf has the expressivity of OCL (Object Constraint Language) in the use and manipulation of sequences of values, enabling an OCL-like syntax.

The execution semantics of Alf is given by mapping the Alf concrete syntax to the abstract syntax of fUML [19]. fUML abstract syntax is a subset of UML with additional constraints, so a well-formed model is one that meets all constraints imposed on its syntactic elements by the UML abstract syntax as well as all additional constraints imposed on those elements by the fUML abstract syntax.

Moreover, the execution semantics of fUML is an executable model written in fUML. However, instead of using Activity Diagrams, activities are written as equivalent code in Java; to support that, a mapping from Java to Activity is defined for core elements of activities (Base UML - bUML). The circularity is broken by the base semantics for bUML, which is specified in first order logic based on Process Specification Language (PSL). PSL (ISO 18629) provides a way to disambiguate common flow modeling constructs in terms of constraints on runtime sequences of behavior execution; desired behavior is specified by constraining which of the possible executions is allowed [6]. Fig. 1 shows relationships between these OMG specifications. In the following, "fUML execution model" refers to fUML and Alf.

SysML reuses a subset of UML 2 and provides additional extensions to satisfy the necessities of systems engineering, e.g. Requirement Diagrams [20]. SysML and Alf integrate seamlessly because Alf can be used in context of models not limited to the fUML subset [19].

Concerning the MoC provided by UML, one basic premise from this modeling language is that all behaviors are ultimately caused by actions executed by active objects [18], which is an instance of an active class (executed concurrently).

This establishes concurrent processes (active objects) but does not define a specific MoC because all BehavioralFeatures (e.g., Operations and Receptions) in UML allow three types of concurrency: sequential, guarded, and concurrent. Therefore, the semantics is unconstrained, which supports heterogeneous MoCs; in fact, it is one of the goals of the specification.

fUML constrains the concurrency for all BehavioralFeatures to the sequential type; as a result, the sole mechanism for asynchronous invocation in fUML is sending signals (SendSignalAction) to other active objects [19]. Further, the sending action is not blocking, i.e., an object sends a signal and continues its execution; it does not wait for a response, or an acknowledgment (nonblocking write). In contrast, the reception action is blocking, i.e., one computation running is blocked when it expects to receive a determined signal (blocking read). Moreover, the received signals are stored in an unbounded event pool for each active object, which is a FIFO (first-in first-out) in the fUML standard execution model (this is a variation point [19]). Consequently, the fUML standard execution model is characterized by concurrent processes (active objects) communicating with each other through unidirectional unbounded FIFO event pools, where writes to the event pool are nonblocking, and reads are blocking.

These fUML's characteristics are what the Kahn process networks have [13]. However, fUML standard execution model defines that signals coming from different active objects should be stored in the same target event pool. Allowing more than one process to write to an event pool (channel), the resulting process network is neither deterministic [13] nor a Kahn Process Network (in the strict sense). Consequently, the resulting process network can be described by active objects that receive (input) and emit

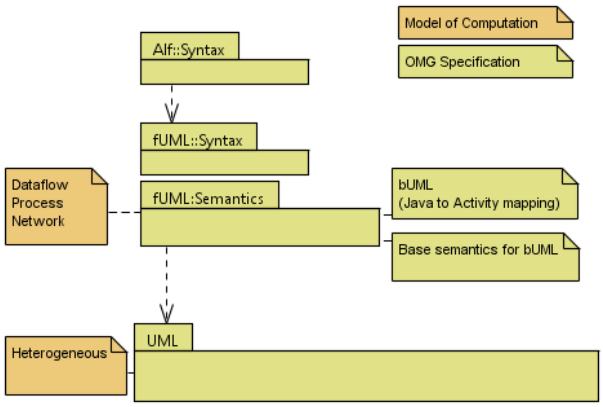


Fig. 1. Relationships between OMG specifications and MoCs.

(output) signals, and a set of firing rules (encoded in the behavior) defining when an active object should be fired; these characteristics are what the dataflow process networks have [13].

Nondeterminism can be a powerful modeling tool, but it should be used only when necessary [13]. Consequently, deterministic languages that allow nondeterminism remove it using precise techniques, e.g. the Quartz [26] compiler adds new control events to remove nondeterminism allowed by some statements.

Despite the nondeterminism of fUML MoC, it is designed to support a variety of different MoCs. This is pursued using two techniques: (1) defining explicit variation points, which are: event dispatching scheduling (used in the inter-object communication), and polymorphic operation dispatching; (2) leaving some semantics elements unconstrained that are: timing, concurrency, and inter-object communication.

IV. The Initial Approach

CPSs are often safety-critical systems [14]; hence executable models describing them must be deterministic: given a state $x(t_i)$ and an input $w(t_i)$ the system must generate the same output $u(t_i)$ for each reaction in state $x(t_n)=x(t_i)$ and input $w(t_n)=w(t_i)$.

The fUML specification states that there are a number of cases in which the UML indicates that the execution semantics in a certain area are nondeterministic [19]. In order to understand these nondeterministic areas, the next subsection discusses the roots of nondeterminism in the fUML execution model.

A. fUML and Nondeterminism

In order to analyze fUML's nondeterminism, behavior should be classified, which is done by UML [18] as: (1) intra-object behavior addresses the behavior occurring within classes; (2) inter-object behavior, which deals with how active classes communicate with each other.

The roots of nondeterminism in the fUML specification can be grouped as follows: (1) structural features manipulation – e.g. set one value to a property of an object; (2) conditions – fUML conditional clauses, e.g. defined using if or switch Alf statements; (3) token flow semantics – defines intra-object behavior semantics, e.g. how are tokens offered, and, consequently, in which sequence are nodes executed; (4) ObjectActivation – a key class responsible to bind inter-object behavior with intra-object behavior.

1) Structural feature manipulation

A property in a class, defined by a modeller, is a StructuralFeature in the meta-model of UML. Actions that write or remove values in a StructuralFeature can be nondeterministic. The nondeterminism occurs when a target property has multiplicity greater than 1, it is not ordered, and it does not have the uniqueness property; i.e., the property is a bag.

This nondeterminism can be a challenge for verification but it compromises neither the given definition of deterministic models nor the fUML MoC.

2) Conditions

Conditions are modeled in fUML using ConditionalNodes. ConditionalNode has an association with Clauses; each Clause can have an association with predecessor Clauses. The fUML execution model states that sequential evaluation is performed when the predecessor chain is defined.

Two statements in Alf map to ConditionalNodes in fUML: if, and switch. The statement if is mapped using predecessor clauses in fUML when the modeller uses the construct "if (*condition) else …"*, so the sequence of evaluation of clauses is deterministic; on the other hand, when the modeller uses the construct "if (condition) or …" the sequence of evaluation of clauses is nondeterministic. Finally, the statement switch is mapped without use of predecessor clauses in fUML so the evaluation of clauses is not deterministic.

As a result, the modeller has two options to produce a deterministic model, concerning conditions using Alf: (1) define conditions that are mutually-exclusive (assured by the modeller, or by an automated assistant); or, (2) use the Alf construct "*if (condition) else …*". A nondeterministic model is defined otherwise.

This nondeterminism compromises the given definition of a deterministic model, but it does not affect the fUML MoC.

3) Token flow semantics

fUML states that different execution traces for the same inputs in an identical environment (including same state) are allowed to be different [19].

For example, given two actions that are not directly or indirectly ordered by their relationships, the order of execution is determined neither by UML semantics nor by fUML execution model, as recognized by [4]. Other example, a ForkNode enables race conditions. Therefore, nondeterminism is established in the intra-object behaviors.

Some basic nondeterminism (coming from UML), in the token flow semantics, are removed by semantic mapping from Alf to fUML, e.g., a naive modeller can, using fUML, connect an OutputPin at two InputPins without using a ForkNode (it copies tokens). However, that construction is not possible in Alf, which generates a ForkNode for each local name [21].

This nondeterminism (if these different traces lead to different outputs or signals sent to other active objects) compromises the given definition of deterministic models, and can contribute to the nondeterminism in the fUML MoC.

4) ObjectActivation

ObjectActivation is the class defined in the execution model to handle the active behavior of an active object. It is responsible to bind inter-object behavior with intra-object

behavior because it, together with EventAccepters, offers the blocking read feature for fUML MoC.

Two associations of this class are important for analysis of nondeterminism: (1) eventPool - the list, without upper bound, of pending signals sent to the object handled by this object activation; (2) waitingEventAccepters - the set of event accepters waiting for signals to be received by the object handled by this object activation.

For example, an execution sequence (ES) for two active objects communication can be explained as follows: (1) an active object (A) reaches an AcceptEventAction (statement accept defined by Alf), this is a blocking read for a signal; (2) the corresponding ObjectActivation object registers an EventAccepter in the waitingEventAccepter; (3) another active object sends a signal, that matches (A) receptions, and the registered accept statement; (4) the ObjectActivation object inserts this new signal at the end of eventPool; (5) considering that eventPool had no previous signals, this signal is removed from eventPool, dispatched to respective accept statement, and EventAccepter is unregistered.

The step (5) is one of two explicit variation points from fUML, called event dispatching scheduling. The standard execution model provides the implementation described above, where events are dispatched from the pool using a first-in first-out (FIFO) rule.

The ObjectActivation is the key to understand how nondeterminism in the fUML MoC and in the token flow semantics is combined. Exploring the execution model of fUML, Fig. 2 shows an Activity Diagram for an active class. Further, Fig. 3 shows an Alf representation for the Activity Diagram presented in Fig 2.

In Fig. 2, there are two concurrent AcceptEventAction waiting for the same type of signal; they are designed to execute two different tasks using received signals. The ForkNode, together with the fact that the next two actions wait for the same signal, defines a race condition, where the output depends on the sequence of tokens offered. Considering that a signal sent by another active object arrived after the two EventAccepters were registered, and the execution sequence (ES) presented above; during the event dispatching phase (5), there are two registered EventAccepters. In this case, the execution model chooses nondeterministically one of these [19], dispatches the event to it, and unregisters it.

This nondeterminism compromises the given definition of a deterministic model, and contributes to the nondeterminism in the fUML together with fUML MoC.

B. Proposed specialization of Alf and fUML

The initial approach is described as follows: given the semantics defined by fUML, we specialize the explicitly unconstrained elements with the purpose of deterministic behavioural definitions using SysML and Alf. We chose to discuss the semantics in an informal way, and to present concrete additional Alf constructs for the specialization.

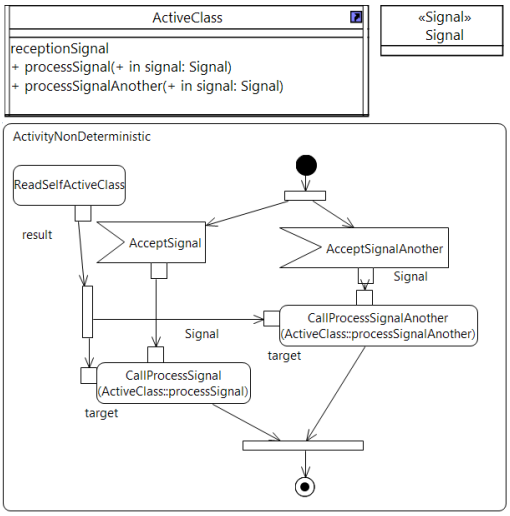


Fig. 2. fUML Activity diagram – nondeterministic.

These additional language constructs are defined using Annotation. According to Alf abstract syntax [26], annotation is a way to identify a modification to the behavior of an annotated statement. The applied approach allows us an early evaluation of the proposed specialization.

Therefore, a first concern is to introduce a synchronous-reactive MoC on fUML and Alf. A second concern is to specialize fUML and Alf, which means: do not change syntax parsing of Alf, but change its semantics.

The next three subsections explore the introduction of the synchronous-reactive MoC in Alf using unconstrained elements, and the variation points. The fourth subsection summarizes the proposal.

1) Timing

The behavioral semantics of UML only deals with discrete behaviors [18]. Accordingly, the timing semantics proposed divides the time scale in a discrete sequence of instants, each instant corresponds to one macro step as defined in the next subsection.

The annotation @delayed was introduced; it is the only way to assign new values to an already assigned variable in the current macro step. This annotation can be used in the assignments and in the SendSignalActions.

2) Concurrency

Concurrency can be achieved in Alf using two complementary techniques: (A) multiple active objects that, in general, imply the necessity of inter-object communication; or, (B) inside a given definition by the use of the annotation @parallel [21].

The alternative proposed is a combination of concurrency and synchrony (where computation and communication are instantaneous) through introducing the synchronous-reactive MoC to fUML and Alf. According to this MoC, a program can be defined by so-called micro and macro steps. Each macro step is divided into finitely many micro steps, which are all executed in zero time and within the same variable

environment. As a consequence, the values of the variables are uniquely defined for each macro step. Macro steps correspond to reactions of reactive systems, while micro steps correspond to atomic actions [26], e.g., assignments using Alf.

The demarcation of macro steps was introduced by the annotation @pausable; it is one of two ways to define demarcation between two macro steps. The second way is the use of the accept statement. This annotation is designed to be used with loop constructs (while, for, do while) but it can be also used with an empty statement of Alf. The semantics is: after each execution of the loop body, it waits for the next macro step. It follows that all concurrent behaviors run in lockstep: they execute the actions inside the loop in zero time, and synchronize before the next iteration.

The annotation @parallel can be used to define that all the statements in the block are executed concurrently. The block does not complete execution until all statements complete their execution; i.e., there is an implicit join of the concurrent executions of the statements [21].

Alf provides also an annotation called @isolated, it is defined in [21]: no object accessed as part of the execution of the statement or as the result of a synchronous invocation from the statement may be modified by any action that is not executed as part of the statement. Similar to this annotation, Alf provides the isolation expression through character $. Both options are not compatible with the synchronous-reactive MoC, where variables are uniquely defined for each macro step.

3) Inter-Object Communication and Event-Dispatching

Inter-object communication in Alf is performed by sending signals to other active objects. A signal is a specification of what can be carried; furthermore, a signal event represents the receipt of a signal instance in an active object [21]. A signal instance is identified by its contents.

Signals are based on the paradigm of message passing; furthermore, fUML provides a point-to-point (also known as unicast) message pattern. A signal is sent to a receiver (an active object) using a reference to it. In contrast, multicasting is required in many safety-critical systems, e.g., fault-tolerance by active redundancy [16]. Multicasting also supports the non-intrusive observation of component interactions by an independent object, and enables better composition [16].

```
                                                    //ACC0
//@parallel
{
    {
        accept( sig:Signal );                       //ACC1
        this.processSignal( sig );
    }
    {
        accept( sigAnother:Signal );                //ACC2
        this.processSignalAnother( sigAnother );
    }
}
```

Fig. 3. Alf representation for fUML Activity diagram – nondeterminitisc.

Multicasting is provided by the introduction of an active class called MessageDispatcher; it provides a service for multicast message exchange. Instances of this class work as bus transferring instances of signals between previously registered active objects, which generate events in the target active object. Every signal handled by MessageDispatcher has a specific identifiable sender, and zero or more receivers.

The set of receivers (active objects) is defined by existence of the reception for the sent signal. All signals generated in the current macro step are available instantaneously in the synchronous-reactive MoC. Further, signals not consumed during a macro step are lost. Delayed SendSignalActions are available in the next macro step.

It is possible to receive signals individually or as a set. Receiving a set of signals is important for those active objects that need to process all signals sent in the current macro step. However, individual signal receiving is fundamental for those active objects that should only process one signal sent to them. For this case (individual signal), the annotation @nonblocking was introduced; it is the only way to receive signals without blocking (nonblocking read).

In a macro step just one signal value (a signal is identified by its contents) is allowed for a given signal type, and MessageDispatcher; therefore, values of the signals for a given MessageDispatcher are uniquely defined for each macro step.

4) Summary

Table I summarizes the annotations available in the specialization of Alf. All other annotations available in Alf now are just comments, as well as, isolation expressions.

Considering that execution model of fUML has changed to accommodate proposed specialization, the semantics of Alf representation in Fig. 3 changes. As just one signal value in a macro step is allowed for a given signal type, and MessageDispatcher; the same signal instance is dispatched for those two parallel accepts, and computation follows in the same macro step concurrently.

The specialized semantics removes the nondeterminism indicated in section “IV.A.4 Object Activation” as described earlier. Also, it removes the nondeterminism indicated in section “IV.A.3 Token flow semantics” because the ordering of micro steps does not influence the semantics of a model. However, the new semantics does not remove the nondeterminism indicated in section “IV.A.2 Conditions”,

TABLE I.
ANNOTATIONS IN THE SPECIALIZED ALF

Annotation	Informal semantics
@delayed	Delayed assignment or SendSignalAction
@pausable	Macro step demarcation
@parallel	Computations on each block are carried out concurrently
@nonblocking	AcceptEventAction read nonblocking, makes optional signals available

which should be rejected by an interpreter for proposed semantics (when conditions are not mutually-exclusive).

With the proposed specialization, Fig. 3 can be changed without modification of the semantics: the two accepts (ACC1 and ACC2) could be removed, and a new one (ACC0) could be inserted before the concurrent block. This is referential transparency, which means syntactically identical expressions have the same semantics regardless of their lexical position [13].

5) Example

We evaluate the example from [4] but a case study with well-known CPS is [25]. Fig. 4 shows the Block Definition Diagram (BDD) for it. A PingPongSystem is composed by one Player1 and one Player2; both players are active classes. These two active classes communicate by exchanging signals Ping and Pong. The respective Alf representation for the behavior of each player is presented using comments.

In a given macro step, Player1 sends a delayed Pong (P11), and awaits for Ping (P12). In the next macro step, Pong is received by Player2 (P21), who sends a delayed Ping (P22). The game continues forever as showed in Fig. 5.

Fig. 5 shows the Internal Block Diagram (IBD) for the system, and the Alf representation for the main behavior. Player1 (S3) and Player2 (S2) are created passing an object of MessageDispatcher (S3); later, an infinite loop annotated with @pausable (S5), containing an empty statement, is used to define the evolution of time.

In contrast to [4], which uses static Association between the players, it is used Connectors that specify links between instances playing the connected parts only [18] (decoupling Player1 from Player2). The communication is provided by the instance of MessageDispatcher. The Alf specialization makes the example different concerning evolution of time, signal events, and communication. Therefore, this model is deterministic while [4] is nondeterministic.

V. DISCUSSION

Activity Diagrams are used frequently [1][4][2][23][24]; however, for significant activities, these diagrams quickly become large, intractable to draw and hard to comprehend [19].

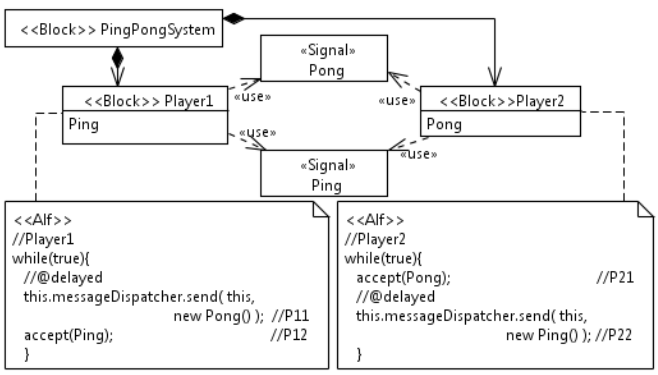


Fig. 4. BDD for the PingPongSystem.

State Machine Diagrams are another commonly used form of diagrams, especially suited for modeling state-based behavior [1][2][4][11][23][24]. However, UML, fUML, SysML, and Alf do not define precise semantics for state machines [9]. This is ratified by the Alf specification itself, which states that a normative semantic integration of state machines with Alf will be formalized later [21]. Indeed, environments of synchronous languages offer tools to visualize the resulting automata [3], e.g. Fig. 3 can be automatically transformed in a State Machine Diagram.

Transformation from SysML to other languages or formalism could bring some serious problems [12], and could compromise scalability [24]. However, we consider the certification process [22] more challenging because it is needed to assess the original model, and the translated model (or even the transformation itself). Nevertheless, these transformations are powerful, and can provide feedback for the fUML specification MoC. For example, [2] defines a pattern suitable of optimization called "fUML-Opti-Rule(2): Detecting unacknowledged signals" - an unacknowledged signal is one that has been sent from an active object to another active object, and then it (source object) continues sending further signals without waiting for an acknowledgment signal. This pattern is detected through model checking executed over a CSP representation, which is the result of a transformation of a fUML model [2]. Based on this feedback, the modeller should evaluate acknowledging those signals to reduce the state space of the corresponding CSP model. Although, this is a rendezvous that is common in CSP MoC; considering this case, fUML MoC needs more design effort than CSP MoC.

Concerning [10][11] which propose to encode SysML structure as a set of axioms, fUML and PSL [6] are well suited, hence axioms about structure and behavior can be combined and evaluated together.

The evaluation presented corroborates [4] concerning two points about fUML (and also Alf) as it is: (1) the execution model is nondeterministic; (2) it is not suitable for safety-critical systems. Nonetheless, Alf should be specialized to allow safety-critical systems modeling [22].

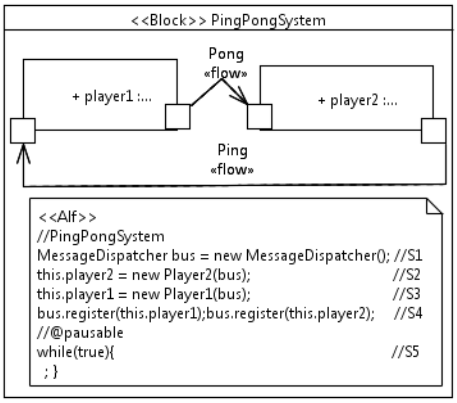


Fig. 5. IBD for the PingPongSystem (graphical view of flows).

The proposed specialization of Alf adheres the idea of introducing synchronous-reactive MoC during early stages of a system development [3]. The major drawback of this MoC is that the computer interpretation of the models is difficult [3]; further, polymorphism, reclassification, and dynamicity (actions: create, and destroy) can be even more challenging [3].

fUML states that every specialization must be defined using bUML; in fact, the initial approach presented here provides a complete description from the viewpoint of the modeller. It defines the semantics for three additional constructs for Alf that together with MessageDispatcher can transform Alf in a synchronous action language; however, the changes needed in the fUML execution model to support it must be defined.

VI. Conclusion

This paper shows the results of the proposed specialization of Alf, according to the synchronous-reactive MoC. It helps to couple complexity, provides better composition, and enables deterministic behavior definitions.

CPS is about the intersection of the computation, control and communication [14]. The initial approach focuses on the discrete computational and communicational aspects of CPSs. It can be composed with discrete control. A case study with a well-known CPS [25] shows that the initial approach can bring solid mathematical foundation from synchronous-reactive MoCs [3] to SysML executable models. We consider this as an intermediary step, located before a formal verification of executable discrete SysML models.

In summary, we believe that specializing well-known vendor-independent specifications (Alf and SysML) can provide an understandable and compact set of languages for modeling, analyzing and verifying of CPSs. Moreover, such a set of languages can enable formal verification for discrete parts of CPSs.

References

[1] Abdelhalim, I.; Schneider, S.; Treharne, H. (2011). Towards a practical approach to check UML/fUML models consistency using CSP. In Proc. ICFEM 2011 Proceeding of the 13th International Conference on Formal methods and software engineering, 2011, pg. 33-48.

[2] Abdelhalim, I.; Schneider, S.; Treharne, H. (2012). An Optimization Approach for Effective Formalized fUML Model Checking. In Proc. SEFM2012 Proceeding of the 10th International Conference on Software Engineering and Formal methods, 2012, pg. 248-262.

[3] Benveniste, A. ; Caspi, P.; Edwards, S.; Halbwachs, N.; Guernic, P.; Simone, R. (2003). The synchronous languages twelve years later. Proceedings of the IEEE, 2003, pg. 64–83.

[4] Benyahia, A.;Cuccuru, A.; Taha, S.; Terrier, F.; Boulanger, F.; Gérard, S. (2010). Extending the Standard Execution Model of UML for Real-Time Systems. In Proc. DIPES/BICC, 2010, pg. 43-54.

[5] Berry, G. (2000). The Esterel v5 Language Primer: version:5.91. France. Available at: <http://francois.touchard.perso.esil.univmed.fr/3/esterel/primer.pdf>. Access date: 14.Apr.2013.

[6] Bock, C.; Gruninger, M. (2005). PSL: A semantic domain for flow models. In Software & Systems Modeling, May 2005, Volume 4, Issue 2, pp 209-231, 2005. Springer.

[7] Bousse, E.; Mentré, D. Combemale, B.; Baudry, B.; Katsuragi, T. (2012). Aligning SysML with the B Method to Provide V&V for Systems Engineering. Proc. Of 12th Model-Driven Engineering, Verification, and Validation 2012.

[8] Cartwright, R.; Kelly, K.; Koushanfar, F.; Taha, W. (2006). Model-Centric Cyber-Physical Computing. In proceedings … NSF Workshop on Cyber-Physical Systems, 2006, Austin, Texas: USA.

[9] Fecher, H.; Schönborn, J.; Kyas, M.; Roever, W. (2005). 29 New Unclarities in the Semantics of UML 2.0 State Machines. In Proceedings of the Int. Conf. on Formal Engineering Methods, LNCS 3785, Berlin/Heidelberg, Germany, Springer-Verlag, 2005, pg. 52-65.

[10] Graves, H. (2012). Integrating Reasoning with SysML. In Proc. of 22th Annual INCOSE International Symposium. Rome, Italy, July, 2012.

[11] Graves, H.; Bijan, Y. (2011). Using formal methods with SysML in aerospace design and engineering. Journal Annals of Mathematics and Artifical Intelligence. Volume 63, Issue1, September, 2011. pg 53-102.

[12] Hußmann, H. (2002). Loose semantics for UML, OCL, in: Proceedings 6th World Conference on Integrated Design and Process Technology, IDPT 2002, June, Society for Design and Process Science, 2002.

[13] Lee, E.; Parks, T. (1995). Dataflow process networks. Proceedings of the IEEE, vol. 83, no. 5, May, 1995. pg. 773-801.

[14] Lee, E.; Seshia, S. (2011). Introduction to Embedded Systems - A Cyber-Physical Systems Approach. http://leeseshia.org/, 2011. ISBN 978-0-557-70857-4.

[15] Miller, P.; Whalen, M.; Obrien, D.; Heimdahl, M.; Joshi, A. (2005). A methodology for the design and verification of globally asynchronous/locally synchronous architectures. NASA Contractor Report NASA/CR-2005-213912.

[16] Obermaisser, R.; Kopetz, H. (2009). Genesys – A candidate for an ARTEMIS Cross-Domain Reference Architecture for Embedded Systems. 2009. Available at: <http://www.genesys-platform.eu/genesys_book.pdf> Access date: 17.May.2011.

[17] Object Management Group (OMG). (2003). Model-Driven Architecture. USA: OMG, 2003. Available at: <http://www.omg.org/mda>. Acesso em: 17 may. 2009.

[18] Object Management Group (OMG). (2011). Unified Modeling Language Superstructure: Version: 2.4.1. USA: OMG, 2011. Available at: <http://www.omg.org/spec/UML/2.4.1/>. Access date: 14.Apr.2013.

[19] Object Management Group (OMG). (2012). Semantics of a Foundational Subset for Executable UML Models: Version 1.1 RTF Beta. USA: OMG, 2012. Available at: <http://www.omg.org/spec/FUML/>. Access date: 24.Apr.2013.

[20] Object Management Group (OMG). (2012). Systems Modeling Language: Version: 1.3. USA: OMG, 2012. Available at: < http://www.omgsysml.org/>. Access date: 24.Apr.2013.

[21] Object Management Group (OMG). (2013). Concrete Syntax for UML Action Language (Action Language for Foundational UML - ALF): Version: 1.0.1 - Beta. USA: OMG, 2013. Available at: <http://www.omg.org/spec/ALF/>. Access date: 27.Apr.2013.

[22] Perseil, I. (2011). ALF Formal. Journal Innovations in Systems and Software Engineering, Volume 7, Issue4, December, 2011. pg. 325-326.

[23] Pétin, J.; Evrot, D.; Morel, G.; Lamy, P. (2010). Combining SysML and formal models for safety requirements verification. In 22nd International Conference on Software & Systems Engineering and their Applications, France, 2010.

[24] Planas, E.; Cabot, J.; Gomez, C. (2011). Lightweight Verification of Executable Models. In Proc. ER 2011 Proceedings of the 30th International Conference on Conceptual Modeling, 2011. pg. 467–475.

[25] Romero, A. G.; Schneider, K.; Ferreira, M. G. V. (2013). Synchronous Specialization of Alf for Cyber-Physical Systems. In First Open EIT ICT Labs Workshop on Cyber-Physical Systems Engineering, 2013, Trento, Italy.

[26] Schneider, K. (2009). The synchronous programming language Quartz. Internal Report 375, Department of Computer Science, University of Kaiserslautern, Kaiserslautern, Germany, December 2009.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1423–1428

Improving security in SCADA systems through firewall policy analysis

Ondrej Rysavy Jaroslav Rab Miroslav Sveda

Faculty of Information Technology
Brno University of Technology,
612 66 Brno, Czech Republic
e-mail:{rysavy, rabj, sveda}@fit.vutbr.cz

***Abstract*—Modern SCADA networks are connected to both the companys enterprise network and the Internet. Because these industrial systems often control critical processes the cyber-security requirements become a priority for their design.**

This paper deals with the network security in SCADA environment implemented by firewall devices. We proposed a method for verification of firewall configurations against a security policy to detect and reveal potential holes in implemented rule sets. We present a straightforward verification method based on representation of a firewall configuration as a set of logical formulas suitable for automated analysis using SAT/SMT tools. We demonstrate how such configuration can be analyzed for security policy violation that can be inferred from a security policy specification of an industrial automation system.

I. Introduction

SCADA (Supervisory Control and Data Acquisition) systems are commonly deployed to continuously monitor and control industrial processes to assure proper functioning, by automating telemetry and data acquisition. Historically, SCADA systems were believed to be secure because they were isolated networks: an operator console, or human-machine interface (HMI), connected to remote terminal units (RTUs) and programmable logic controllers (PLCs) through a proprietary purpose-specific protocol. Yielding to market pressure, that demands industries to operate with low costs and high efficiency, these systems are becoming increasingly more interconnected. Many of modern SCADA networks are connected to both the companys enterprise network and the Internet. Furthermore, it is common that the HMI is a commodity PC, which is connected to RTUs and PLCs using standard technologies, such as Ethernet and WLAN (see Fig. 1). Such configuration has exposed these networks to a wide range of security problems. The access to individual subnetworks are secured by firewalls that implement basic network security policy.

Securing networks properly by configuring firewall rules is difficult, time consuming and error-prone task. Wool has analyzed possible threats of incorrectly configured firewalls in [1] and called for methods that would help to improve the quality of firewall rules. The stated observation considers the complexity and the size of firewall rule sets as the main source of errors. He identified major source of difficulties in creating complex firewall configurations. Although Wool considered only a small set of relatively obvious errors, his survey demonstrated that a rule set having 1000 items includes more than 8 errors on average.

The approach described in this paper is close to the work done by Guttman [2], Bera, Ghosh and Dasgupta [3], and Al-Shaer et el [4]. Similarly we develop the method that is able to verify correctness and consistency of firewall configurations against network security policy given a set of simple policy rules. We show a simple translation of policies and firewall rules into logical formulas and describe the Satifiability Modulo Theory (SMT) verification method. The SMT tools employ algorithms for solving logical formulas with respect to combinations of background theories expressed in classical first-order logic with equality. In the present work we use Microsoft's Z3 tool that implements an efficient SMT decisions procedures.

Packet filters implement the basic level of security policies in the network. By restricting the accessibility of certain services, computers or subnetworks, we deploy rough but efficient security measures. Our network model deals only with IP addresses and services or ports. Therefore, the analysis does not reflect hardware or Operating Systems (OS) attacks. The contents of TCP/UDP packets are not examined, but it is possible to extend the description to support this. Our primary goal is to verify safety or resistance of the network with respect to the effect of dynamic routing. Therefore, this classification includes only basic categories of network security properties. Since it can utilize typical fields from IP, TCP, or UDP headers, namely source/destination IP address and service/port allows us to specify wide range of different communications to be analyzed in the network.

This paper is structured as follows: Section II discusses various packet filter representations. Section III presents representation of filtering rules in form of SMT formulas. In Section IV we define a verification method for a single firewall configuration. This is extended to the cascade of firewalls in Section V, thus providing a method for system-wide security policy verification. In Section VII we present a preliminary experimental results showing performance of the presented method. The paper concludes in Section VII by comparing presented method to related work and suggesting further development.

978-83-60810-53-8/$25.00

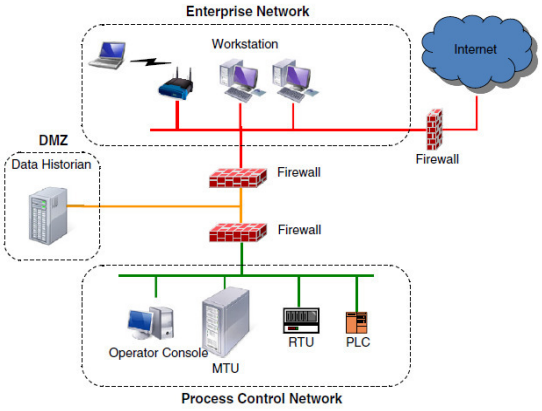


Fig. 1: An example of modern SCADA network.

II. Representation of Packet Filters

Firewall configurations are usually written in form of access control lists (ACL). An ACL format is easy to understand for network administrators and it is also suitable for machine processing. Nevertheless, for an efficient formal analysis this format may represent a problem because it admits conflicting rules. Conflicting rules are pairs of rules that match the same set of packets. These conflicts are solved at runtime by implementing first match semantics. However, certain classes of conflicts can signalize a configuration error, for instance, a rule that completely hides some other rules. Several methods to check conflicts in ACLs and constructing a non-conflicting *rule sets* were proposed, e.g. [5], [6], [7], [8].

Rules have multidimensional structure. Dimensions correspond to fields in a packet header, in particular, source and destination addresses, port numbers and a protocol type. Formally, we define a rule as a tuple $\langle src, dst, srv, act\rangle$, where *src* and *dst* are set of addresses, *srv* is a set of services, and *act* is an action.

A logical formula that is a translation of a simple rule $r = \langle s, d, v, a\rangle$ consists of a conjunction of all selectors. A selector is represented by a predicate that extracts required header field from packet p. Thus, for rule r the formula is written as follows:

$$src\_adr(p) \in s \wedge dst\_adr(p) \in d \wedge service(p) \in v.$$

A list of all possible selectors is shown in Table I A network-mask convention is adapted for representing a sequence of continuous addresses. For instance, address prefix 147.229.12.0/24 is a set of addresses ranging from 147.229.12.0 to 147.229.12.255. We can use the standard set operations, e.g., $src\_adr(p) \in 147.229.12.0/24$ or $dst\_adr(p) \in 147.12.28.0/24 \cup 147.12.30.0/24$. The latter can be expanded to $dst\_adr(p) \in 147.12.28.0/24 \vee dst\_adr(p) \in 147.12.30.0/24$, which allows us to use network-mask format for the canonical address representation.

Often, rule sets implicitly assume the existence of a default rule, which has the lowest priority and matches all packets not matched by any of the previous rules. For a single rule set of an ACL configuration we compute two logical filter representations. A *positive* filter represents all packets permitted by the ACL configuration. A *negative* filter represents all packets denied by the ACL configuration.

TABLE I: Network field selectors

Function	Description
$dst\_adr(p)$	Destination address of a packet p.
$src\_adr(p)$	Source address of a packet p.
$dst\_port(p)$	Destination port of udp or tcp datagram carried in packet p.
$src\_port(p)$	Source port of udp or tcp datagram carried in packet p.
$service(p)$	Service of a packet p.

III. Filter Representation

As proposed in [9] the output of reachability analysis and the input for consecutive security property analysis consist of a collection of reachability sets for forwarding paths in an analyzed network. There are various methods to calculate reachability sets. In this section, we discuss several issues related to these calculations. We overview the problem of efficient address encoding and rule set representation.

Guttman has described an approach to deal with abstract address scheme [2]. The abstract address is a symbolic name of a host or a subnetwork. This address scheme avoids dealing with huge IP address space, which consists of 2^{32}addresses. An abstract packet consists of an abstract source address, an abstract destination address, service identification, and a flow orientation. The flow direction represents the communication direction that is either client to server, or server to client. This approach leads to very reasonable complexity which is dependent on the size of the network and mainly on the number of interesting destinations and services. For an example, considering a network with N different distinguished addresses, S different distinguished services, then the abstract packet space of size will be $N^2 \cdot 2S$.

Different approach was proposed by Bera, Ghosh and Dasgupta in [3]. In their work, the IP address space is explicitly represented by bit variables. The bit variables $s_1, \ldots, s_{32}$ represents a source address, bit variables $d_1, \ldots d_{32}$ represents a destination address, and a vector of bit variables $v_1, \ldots, v_n$ of the appropriate length n, represents a service. A flow direction may be modeled separately by a single bit variable or encoded in the service vector. In this way, there is an explicit representation not only for each packet but also for each network represented in network-mask format.

Independently on whether we use abstract address representation or explicit representation, we construct logical formula for each rule in a filter. These are used in composition of formulas for positive and negative filters. Such formula can be encoded as a SAT instance using the Boolean reduction approach, which is defined in detail for explicit address scheme in [3]. If the abstract address scheme is used each abstract address has to be represented by a single Boolean variable.

These two approaches differ from the number of Boolean variables in generated SAT instances. While explicit represen-

tation requires the fixed number of variables, the number of variables used by abstract approach depends on the number of abstract addresses. On the other hand, the former may generate a large number of clauses while the latter tends to keep number of clauses smaller. It remains for future work to analyze and compare both approaches from the practical perspective on real data.

IV. SMT-BASED VERIFICATION METHOD

In this section, we describe an SMT-based verification method for validation of a network security policy. Given requirements on a packet flow and a filter specification in form of a rule set, we compute a subset of rules that violates these requirements. If the subset is empty than all requirements are satisfied.

First, we present the method to verify a single filter against a security policy. Later this method will be extended for verifying a cascade of filters. Checking if the specified packet flow p is permitted by a filter f it is enough to show that formula $\bar{f} \wedge p$ cannot be satisfied. For instance, assuming that s_0, s_1, s_2 are atomic propositions capturing abstract packet properties. Then filter $\bar{f}$ and a policy p are expressed as follows:

$$\bar{f} = \bigvee \begin{array}{l} s_0 \wedge s_1 \\ s_0 \wedge s_2 \\ s_1 \wedge s_2 \end{array}, \quad p = s_0 \wedge \bar{s_1}$$

In thi case, it is possible to find an assignment $s_0 = 1, s_1 = 0, s_2 = 1$ that satisfies $\bar{f} \wedge p$. While this gives us the required answer we would like to obtain more information to track the problem. To do so, we enrich the filter representation with information that refers to corresponding filtering rules.

$$\bar{f} = \bigvee \begin{array}{l} r = 0 \wedge s_0 \wedge s_1 \\ r = 1 \wedge s_0 \wedge s_2 \\ r = 2 \wedge s_1 \wedge s_2 \end{array}, \quad p = s_0 \wedge \bar{s_1}$$

where r is a bit vector that encodes a rule number. Using this annotation the answer contains information on deny rule that denied the analyzed packet flow, which is, $r = 1$.

To capture network security policy we employ Security Policy Specification Language (SPSL) as defined in [10]. This simple language allows us to express services available between different network zones. For network presented in Fig.1, such policy specification can be as follows:

```
zone ENTP [10.10.100.0/24];
zone DMZ [10.10.10.0/24];
zone PCN [10.10.200.0/24];
zone Inernet [*];

service HTTP = TCP [port = 80];
service SSH = TCP [port=22];
service TELNET = TCP [port=23];

policy p1 = deny [telnet,http]([ENTP],[PCN]);
policy p2 = deny [*]([Internet],[PCN]);
policy p3 = permit [http]([Internet],[DMZ]);
```

For instance, a specification of policy $p1$ can be converted to the following SMT representation:

```
01  (define-fun p_1 () Bool
02  ; deny [telnet,http]([ENTP],[PCN])
03     (and
04         (= (bvand dst_ip PCN_MASK) PCN)
05         (= (bvand src_ip ENTP_MASK) ENTP)
06         (or (and (= pt TCP)(= dst_pn HTTP))
07             (and (= pt TCP)(= dst_pn TELNET))
08         )
09      )
10  )
```

This policy denies telnet and http traffic to the Process Control Network. This is encoded by specifying source (line 5) and destination (line 4) address ranges of the packets that should be denied. Lines 6 and 7 describe protocol type and destination port numbers that correspond to telnet and http traffic, respectively. Addresses are encoded as bit vectors of size 32. Encoding constraints on addresses follows the general pattern:

```
(= (bvand x net_mask) net_addr)
```

Here, `bvand` is a standard bit wise AND operation on bit vectors. Port numbers are encoded as bit vectors of size 16. Using this direct encoding it is possible to directly express policy rules using a standard bit vector theory available in SMT tools.

We demonstrate the translation of ACL configuration to positive and negative filters using the following ACL snippet:

```
R  ip access-list extended paper-example
1   permit icmp any any echo-reply
2   permit icmp any any echo
3   deny   ip any 10.10.10.0 0.0.0.255
4   deny   ip any 10.10.11.0 0.0.0.255
5   permit  ip any any
```

These five rules permit any icmp echo and echo-reply traffic and forbid other traffic to target network. The translation to SMT yields four definitions of functions. Note that default permit rule is not translated.

```
(define-fun f1_r1 () Bool
; permit icmp any any echo-reply
    (and
       (= pt ICMP)
       (= dst_pn ECHO_REPLY)
    )
)
(define-fun f1_r2 () Bool
; permit icmp any any echo
    (and
       (= pt ICMP)
       (= dst_pn ECHO)
    )
)
(define-fun f1_r3 () Bool
; deny   ip any 10.10.10.0 0.0.0.255
    (and
       (= (bvor dst_ip #x000000ff) #x0a0a0aff)
    )
)
(define-fun f1_r4 () Bool
; deny   ip any 10.10.11.0 0.0.0.255
    (and
       (= (bvor dst_ip #x000000ff) #x0a0a0bff)
    )
)
```

Rules constraint only properties explicitly defined. Argument `any` is not represented as it expresses that the variable

is constrained by the valid range of the corresponding type, which is implicitly enforced by the type system of SMT. The translation of addresses and wild cards are according to the following pattern:

```
(= (bvor x wildcard) (bvor address wildcard))
```

To verify that ACL obeys a network security policy we need to obtain a representation in form of two partial filters. The negative filter, denoted as `f1_deny`, is a boolean formula that is satisfied for all denied abstract packets. Likewise, the positive filter, denoted as `f1_permit`, is a boolean formula that is satisfied for all permitted packets. We use this splitting to simplify the process of verification and finding counter-examples. The general method for computation of permit and deny filters is presented as Algorithm 1. We will explain the idea of this algorithm on an example of a deny filter. A list of ACL rules is processed in a reverse order. The deny filter formula is constructed in several steps. The immediate result of each step is denoted as f_d^i. Initially, f_d^0 is empty. The formula f_d^{i+1} is constructed as follows:

- If rule r is deny than its logical representation ϕ_r is added to formula $f_d^{i+1} = f_d^i \vee \phi_r$.
- If rule r is permit than its logical representation ϕ_r is combined with filter as $f_d^{i+1} = f_d^i \wedge (\neg\phi_r)$.

Note that in the algorithm the construction of a formula is slightly modified to improve compactness of the resulting formula. All consecutive rules sharing the same action is threated in a single step. Thus, in case of deny rule, we have $f_d^{i+1} = \bigvee f_d^i, \phi_{r_1}, \ldots, \phi_{r_n}$. The deny filter for ACL from the previous example is generated as follows:

```
01  (define-fun f1_deny () Bool
02       (and
03           (not f1_r1)
04           (not f1_r2)
05           (or
06               (and f1_r4 (= deny 4))
07               (and f1_r3 (= deny 3)))))
```

It can be seen that with deny rules there are annotations referring to ACL rules. The annotations allow us to infer information for counter-examples. The permit rule is computed in similar way. Line 8 contains a representation of permit all rule. Permit/Deny all rules match all abstract packets, thus logical representation is constant `true`.

```
01  (define-fun f1_permit () Bool
02       (or
03           (and f1_r1 (= permit 1))
04           (and f1_r2 (= permit 2))
05           (and
06               (not f1_r3)
07               (not f1_r4)
08               (and true (= permit 5)))
```

Policy verification is performed by checking formulas representing policy and filter by the SMT tool. For restricting policies, p_1 and p_2 it means to find satisfying valuation for $p_1 \wedge f_d$. In SMT syntax this is represented by the following code block:

```
(assert (and f1_permit p_1))
```

Algorithm 1 Computation of a permit filter

Require: An input access-control list L, represented as an ordered list of rules, $r_1, \ldots, r_n \in L$.

$$r_i \in \left[\begin{array}{l} action : \{permit, deny\}, pt : protocol, \\ src.ip : ip\_range, dst.ip : ip\_range, \\ src.pn : port\_range, dst.pn : port\_range \end{array}\right].$$

Ensure: A boolean formula representing the deny filter f_d.

```
f_d := true
R = L.Reverse
while R not empty do
  r := R.Pop
  if r.action = permit then
    p := true
    while r.action = permit & R not empty do
      p := p ∧ ¬φ_r
      r = R.Pop
    end while
    f_d := f_d ∧ p
  else
    d := false
    while r.action = deny & R not empty do
      d := d ∨ φ_r
      r = R.Pop
    end while
    f_d := f_d ∨ d
  end if
end while
```

```
(check-sat)
```

The answer of SMT is `unsat`, which means that the conjunction cannot be satisfied and hence the filter f_1 is correct with respect to policy p_1. In case of policy p_2 the result given by SMT is `sat` and a random model is provided, e.g., an assignment satisfying `(assert (and f1_permit p_1))` is as follows:

```
permit = 2, pt = ICMP, src_ip = #x0a0a6400,
dst_pn = #x0800, dst_ip = #x0a0a0a00
```

Such result contains diagnostic information telling us that policy is violated by ACL because permit rule 2 matches ICMP echo-reply packets originated from 10.10.100.0 and destined to 10.10.10.0. However, these packets should be denied according to the policy.

A cascade of filters is verified by applying essentially the same approach as described in previous sections. permit and deny predicates are computed for each filter. Then these filters are combined to a single formula representing the cascade of filters.

- $f_p^c = f_p^1 \wedge \ldots \wedge f_p^n$,
- $f_d^c = f_d^1 \vee \ldots \vee f_d^n$,

where $f_p^1, \ldots, f_p^n$ are permit filter predicates and $f_d^1, \ldots, f_d^n$ are deny filter predicates. Permit filter is combined using $\wedge$ operator as a packet is permitted if it passes all ACL on the

path. Contrary, a packet can be filtered by any ACL on the path and thus $\vee$ operator is used.

V. System-wide Analysis

In this section, we discuss an extension of a described method for verification of a security policy to system-wide scope. The main goal is to find a network states that violate the given security policy. Recall that security policy is a list of permitted and denied traffic between specified locations. Performing system-wide analysis amounts to check for every pair of network locations specified in a policy rule the permit or deny requirements on the traffic. As there can be multiple paths between these locations these have to be considered. Once we found that a path violates the policy rule it is reported to the user. Considering SCADA network as shown in Fig. 1. Then the topology of this network is capture by the following specification:

```
(declare-const path (Array Int Bool))

;path 1 = ENTP -> F1.1 -> F2.1 -> PCN
(define-fun fp1_permit () Bool
    (and f1_1_permit f2_1_permit))
;path 2 = ENTP -> F1.1 -> DMZ
(define-fun fp2_permit () Bool
    (and f1_1_permit))
;path 3 = PCN -> F2.2 -> F1.2 -> ENTP
(define-fun fp3_permit () Bool
    (and f2_2_permit f1_2_permit))
;path 4 = PCN -> F2.2 -> DMZ
(define-fun fp4_permit () Bool
    (and f_2_2_permit))

; checking violations for policy 1
(assert (or
    (and fp1_permit p1 (select path 1))
    (and fp2_permit p1 (select path 2))
    (and fp3_permit p1 (select path 3))
    (and fp4_permit p1 (select path 4))))
```

We use array to remark which paths violate the policy. The evaluation of SMT specification leads to finding a counter example in case of policy rule violation. The presented encoding brings any counter example depending on the run of SMT algorithm. However, it would be desirable if the produced counter example represent the largest subset of a rule set that violates a security policy. Using this approach the user is not confronted with an arbitrary counter example in case of policy violation, but with a counter-example that, if applied to path based policy checking, violates the greatest number of paths.

The idea of finding the greatest number of paths, which violates the policy rule is based on binary search procedure that guarantees to find the result in $\log_2 N$ steps. The search environment is initialized by introducing a counter array, which keeps the number of paths violating the policy rule. An index in the array is computed as follows:

$$\mathit{sums}\ [i] := \mathit{sums}\ [i-1] + \text{IF}\ path[i]\ \text{THEN}\ 1\ \text{ELSE}\ 0.$$

This initialization is encoded as follows:

```
(define-sort SumT () (Array Int Int))
(declare-const sums SumT)

(assert (= (select sums 0) 0))
```

```
(assert
  (forall ((i Int))
    (ite (select path i)
         (= (store sums i
           (+ (select sums (- i 1)) i)) sums)
         (= (store sums i
           (select sums (- i 1))) sums)
    )
  )
)
```

Note that it is better to unwind the forall statement to avoid dealing with quantifiers. The iteration consists of several steps for i by asserting the following:

```
(assert (= (select sums n) i))
```

Here, n is the total number of paths. Reading *sums* $[n]$ means to get a number of satisfied paths. The iterative steps are guided by the immediate results of SMT executions for the current instance.

VI. Results and Discussion

We experimentally implemented the proposed SMT-based method using Microsoft's Z3 tool. The results of execution of this method on problems of various size are shown in Table II.

The testing set of filtering rules consists of filters generated using the tool called ClassBench [11]. This generator is equipped with templates of filtering rules derived from a collection of real firewall configurations. The tool generates ACLs of different sizes and parameters. For our purpose, we generated filters for different templates, denoted as acl1-3 and fw1 and fw2. These templates differ by the number of conflicting rules. For every template a range of filters of various size was generated. We use rule sets generated for these templates as an imput to our tool that translated them to SMT specification, which was consumed by Z3 tool. We meassured time and memory requirements of the SMT method that checks rule set consistency.

Experiments were performed on a 2.53 Ghz Intel Core 2 Duo machine with 8 GB of RAM running Z3 version 4.3.1 in 64 bit mode. Table II contains results for different sizes of the problem. It can be seen that in most cases the time and memory consumption of the methods increases linearly with the number of rules in firewall configuration. The irregulairites are caused by the different number of conflicting rules in those samples.

VII. Conclusions

In this paper, we presented an approach for verifying ACL configurations by translating them to rule sets, which can be formally analyzed using SMT tools. The proposed method enables network administrators to observe the quality and correctness of firewall configurations, which improves the overall security in administered networks. This technique can be combined with other approaches supposed for securing industrial networks. The overview of security threats in industrial networks were presented by Alcaraz et el in [12] and later by Cardenas et el in [13]. These analyses emphasize the

TABLE II: Time and memory requirements of SMT procedure

Time[s]	10	100	1000	10000	100000
acl1	0.01	0.02	0.11	1.43	13.91
acl2	0.01	0.02	0.10	1.13	14.36
acl3	0.01	0.02	0.11	1.22	39.95
fw1	0.01	0.02	0.13	1.08	30.59
fw2	0.01	0.03	0.11	1.42	13.81

Memory[MB]	10	100	1000	10000	100000
acl1	2.35	2.95	7.81	55.41	459.11
acl2	2.36	2.94	7.73	55.45	460.55
acl3	2.31	2.97	7.84	55.48	456.98
fw1	2.34	2.98	7.84	55.45	455.28
fw2	2.34	3.00	7.89	55.45	458.47

importance of a combination of reactive and proactive methods in order to secure the system against deception and DoS attack.

Description of network security properties is related to the classification of threats and intrusion. There are plenty of different network security problems, such as HTTP attacks, spam, TCP flooding, DoS attacks, Web server misuse, spoofing and sniffing etc. Protection of critical components and network infrastructure is identified as a key requirements for improving security in SCADA system by Hentea in [14].

Analysis of firewall configuration has been intensively studied. Namely, Guttmannin [2] proposed algorithm for computing reachibilty sets based on the firewall configurations. Bera et al in [10] proposed SAT-based methods for verification of security policy. Al-Shaer et al. [15] uses similar approach for representation of ACLs as permit and deny predicates. Their verification methods employ the BDD representation in model-checking procedure.

The network model presented in this paper deals only with IP addresses and services or ports. Therefore, the analysis does not reflect hardware or OS attacks. It also does not examine the contents of TCP/UDP packets. Therefore, this classification only includes selected categories of network security properties. Since it can utilize typical fields from IP, TCP, or UDP headers, namely source/destination IP address and service/port, it allows to specify wide range of different communications to be analyzed in the network.

In this paper we demonstrated the problem of automatic security analysis of IP based industrial networks. The presented verification method aims at validating network design against the absence of security and configuration flaws. The verification technique is based on the encoding problem into SMT instance solved automatically by the solver tool.

REFERENCES

[1] A. Wool, "Trends in Firewall Configuration Errors: Measuring the Holes in Swiss Cheese," *IEEE Internet Computing*, vol. 14, no. 4, pp. 58–65, Jul. 2010.

[2] J. Guttman, "Filtering postures: Local enforcement for global policies," in *IEEE Symposium on Security and Privacy*. IEEE Comput. Soc. Press, 1997, pp. 120–129.

[3] P. Bera, S. Ghosh, and P. Dasgupta, "Formal Verification of Security Policy Implementations in Enterprise Networks," *Information Systems Security*, pp. 117–131, 2009.

[4] E. Al-Shaer, W. Marrero, A. El-Atawy, and K. ElBadawi, "Towards global verification and analysis of network access control configuration," *DePaul University, Chicago, IL, USA, Tech. Rep*, 2008.

[5] L. Cholvy and F. Cuppens, "Analyzing consistency of security policies," in *Security and Privacy, 1997. Proceedings., 1997 IEEE Symposium on*. IEEE, 1997, pp. 103–112.

[6] a. Hari, S. Suri, and G. Parulkar, "Detecting and resolving packet filter conflicts," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2000, pp. 1203–1212.

[7] E. Al-Shaer and H. Hamed, "Discovery of policy anomalies in distributed firewalls," in *Ieee Infocom 2004*. Ieee, 2004, pp. 2605–2616.

[8] S. P. Hidalgo, R. Ceballos, and R. M. Gasca, "Fast Algorithms for Consistency-Based Diagnosis of Firewall Rule Sets," *2008 Third International Conference on Availability, Reliability and Security*, pp. 229–236, Mar. 2008.

[9] G. Xie, D. Maltz, A. Greenberg, G. Hjalmtysson, and J. Rexford, "On static reachability analysis of IP networks," *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, pp. 2170–2183, 2005.

[10] P. Bera, S. Maity, S. Ghosh, and P. Dasgupta, "A Query based Formal Security Analysis Framework for Enterprise LAN," *2010 10th IEEE International Conference on Computer and Information Technology*, no. Cit, pp. 407–414, Jun. 2010.

[11] D. E. Taylor, "ClassBench: A Packet Classification Benchmark," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 135–511, Jun. 2007.

[12] C. Alcaraz, G. Fernandez, R. Roman, A. Balastegui, and J. Lopez, "Secure Management of SCADA Networks," *New Trends in Network Management, Cepis UPGRADE*, vol. 9, no. 6, pp. 22–28, 2008.

[13] A. a. Cardenas, S. Amin, and S. Sastry, "Secure Control: Towards Survivable Cyber-Physical Systems," in *Proceedings of the 28th International Conference on Distributed Computing Systems Workshops*. Ieee, Jun. 2008, pp. 495–500.

[14] I. N. Fovino, A. Carcano, and M. Masera, "A Secure and Survivable Architecture for SCADA Systems," *2009 Second International Conference on Dependability*, pp. 34–39, Jun. 2009.

[15] E. Al-Shaer, H. Hamed, and R. Boutaba, "Conflict classification and analysis of distributed firewall policies," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 10, 2005.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1429–1436

Development of a Cyber-Physical System for Mobile Robot Control using Erlang

Szymon Szomiński, Konrad Gądek, Michał Konarski, Bogna Błaszczyk, Piotr Anielski, Wojciech Turek
AGH University of Science and Technology
Krakow, Poland
Email: szsz@agh.edu.pl

***Abstract*—Design of mobile robot control systems is a huge challenge, which require solving issues related to concurrent hardware access and providing high availability. Existing solutions in the domain are based on technologies using low level languages and shared memory concurrency model, which seems unsuitable for the task. In this paper a different approach to the problem of building a cyber-physical system for mobile robots control is presented. It is based on Erlang language and technology, which support lightweight processes, fault tolerance mechanisms and uses message passing concurrency model with built-in inter-process communication. Created system used a new, open-source robotic platform, which had been designed for scientific and educational purposes. Integrated system has been tested in several scenarios, proving flexibility, durability and high performance.**

I. Introduction

RECENT decades brought impressive growth in capabilities of autonomous mobile robots. Each year new wheeled, walking, swimming and flying devices are being created overcoming another limitations of mechanical devices. However, despite theoretical opportunities, mobile robots are still not widespread in industry or other commercial applications.

Each robotics system consists of two main layers: the hardware platform, which provides certain capabilities and has certain limitations, and the software controlling the hardware in order to fulfill particular tasks. It has been shown many times that even relatively low quality hardware can be used for solving sophisticated tasks if the software controls it correctly. This fact encourages further research on methods for building software systems for managing mobile robots.

Building a cyber-physical system responsible for controlling a robot is a huge challenge. Advanced robotic hardware platform is typically equipped with various sensors and effectors for determining the state of the environment and being able to modify it. The system has to interconnect all hardware devices and manage its functioning concurrently. The requirement of reliable, concurrent hardware access with real-time constraints makes the design and implementation of such systems an extremely hard task. It seems that further popularization of mobile robots in real-life applications depends on finding proper technologies and defining methods for rapid development of high quality cyber-physical systems for robotics applications.

Research in the domain of mobile robotics, which have been evaluated using real robots, has always been considered more valuable than the results obtained in simulation. Even solutions to relatively simple problems, like cooperative box pushing [1] or formation control [2], required a huge amount of work to verify using hardware. It seems that development of a control system dedicated for a particular robot and a particular task is definitely an inefficient approach.

Reusability of high level control software components can be achieved by using the software agents paradigm. Defining several layers of abstraction in a solution of a complex problem makes it possible to implement high level algorithms without depending on the specific hardware [3], [4].

Recent years brought some attempts to build an abstraction layer over hardware components, which should accelerate robot control software development. The Player/Stage [5] platform succeeded by the Robot Operating System [6] share the same idea: to provide a set of drivers for particular devices and a uniform method for accessing the hardware. Using a typical hardware with the ROS is definitely far simpler than creating hardware drivers from scratch.

There is no doubt, that the C++ language used by the ROS for implementing hardware drivers is the appropriate choice. However using the same language and technology for providing reliable and concurrent access to the hardware layer and for writing control applications may raise doubts. Shared memory concurrency model, used in C++, is based on pthreads library [7], which does not provide any high level constructs or high availability mechanisms. Error in any fragment of such application causes whole system failure which is an extremely undesirable feature in real-time mobile robot control programs.

In this paper a different approach to the problem of building cyber-physical system for mobile robots control is proposed. It is based on a message passing concurrency model adopted from the software agent paradigm. The model is applied in the hardware management layer in order to make the system more resistant to hardware failures. The implementation of the presented system has been created using Erlang language and technology [8], which provides built-in inter-process communication and failure recovery mechanisms.

The system has been tested on a new wheeled robotic platform developed in the Department of Computer Science, AGH University of Science and Technology. The platform will hopefully become very popular in educational and scientific applications. It is fully open-source, built of relatively cheap and common components, powerful and extendible. Documen-

978-83-60810-53-8/$25.00

tation is available at address http://capo.iisg.agh.edu.pl/.

In the following section the details on the hardware platform are presented. In the following section the complete Erlang-based cyber-physical system for controlling the robot is described. Finally results of preliminary experiments are provided.

II. Mobile Platform Hardware Design

The wheeled robotic platform required for testing the Erlang-based cyber-physical system for mobile robot control has to meet several requirements. It has to be equipped with relatively powerful on-board computer, capable of running Erlang virtual machine and Linux operating system. It has to provide precise velocity control and long lasting power source. One of the key features was extensibility – hopefully the system can be used in many different scientific applications, which may require different sensors and effectors. More over it should be relatively inexpensive.

Significant development and miniaturization of electronic components over last decades resulted in creation of several advanced commercial and open-source mobile robotic platforms. The platforms are designed with different applications in mind, like transportation, exploration of unknown or unsafe area, inspection of inaccessible places such as water pipes or in the buildings security.

Most of these solutions are designed to solve a particular problem. The most widespread group, which can be found in our homes, are cleaning robots, like Roomba, Scooba and Myrrh [11]. Although these robots are quite advanced, they are hardly extendible and it is hardly possible to use the platforms to other purposes.

Avatar III is an advanced platform, which belongs to a group of robots whose main task is to detect potential intruder[12]. This type of robots is characterized by very good parameters in comparison to experimental projects, but it still does not support flexibility in adding extensions.

There are several more flexible mobile platforms available, like Komodo [13] or much bigger Husky [14], which provide large spectrum of available extensions. However, it is hard to determine what kind of on-board computer do they use. Moreover, these solutions definitely do not meet the inexpensiveness requirement.

Designing a robotic hardware platform is a very complex task. Many unexpected problems have to be solved, making the process surprisingly slow. Final solution has to include designing a robot body, its components, power management, physical and electrical interfaces, integration with remote devices etc. Building individual components, like motor drivers or power distribution systems, requires a lot of time and expenses along with further design problems and delays.

In the presented platform simplicity was the key factor. For this reason the robot was built from off-the-shelf components integrated within suitable chassis. Beside of time saving, the most important advantage of this approach is the simplicity of building new units – off-the-shelf components are relatively easy to assemble. Another advantage of this solution is the possibility of independent testing of individual parts of the system which greatly simplifies failure diagnosis. On the other hand, ready-made parts often cause problems during integration because they were designed for other purposes. Selecting, testing and integrating proper components is probably the most important outcome of the presented work.

Designed mobile platform is composed of the following components:

- chassis,
- power supply
- control unit,
- motor drivers,
- sensors and other peripherals.

Selected chassis is a Lynxmotion A4WD1 four-wheel body, 30 cm long and width. The platform is shown in figure 1.

Fig. 1. Designed robotic platform based on A4WD1 chassis.

Power is supplied by two LiPo batteries connected in parallel with the nominal voltage level 14.8 V and single battery capacity 5000mAh, which is sufficient for several hours of continuous operation. Once the robot runs out of energy, batteries can be replaced without restarting the control unit.

The main robot control unit is Pandaboard [9]. Pandboard is a low-power low-cost single board computer based on the OMAP4430 dual core processor. Platform gives access to many of the powerful features of the multimedia processor while maintaining low cost. This will allow the user to develop software and use available peripherals in many configurations. The major components available on the PandaBoard, which can be used in the robot, are as follows:

- Power Management Companion Device,
- Audio Companion Device,
- Mobile LPDDR2 SDRAM Memory,
- HDMI Connector,
- SD/SDIO/MMC Media Card Cage,
- UART via RS-232 interface via 9-pin D-Sub Connector,
- LS Research Module 802.11b/g/n, Bluetooth, FM,
- Camera Connector,
- LCD Expansion Connectors,
- Generic Expansion Connectors,
- Composite Video Header.

The device runs Linux kernel with either popular distribution. The most basic task of the Pandaboard is to control the motor drivers – the RoboClaws [10].

The RoboClaw 2X15 Amp is an extremely efficient, versatile, dual channel synchronous regenerative motor controller. It supports dual quadrature encoders and can supply two brushed DC motors with 15 amps per channel continuous and 30 amp peak. With support for dual quadrature decoding it get greater control over speed and velocity is automatically maintains speed even if load increases. RoboClaw uses PID calculations with feed forward in combination with external quadrature encoders to make an accurate control solution. RoboClaw is easy to control with several built in modes. It can be controlled from a standard RC receiver/transmitter, serial device, microcontroller or an analog source, such as a potentiometer based joystick.

To control the speed of motor RoboClaw uses pulse width modulation (PWM). Pulse width modulation is a method of adjusting the current or voltage signals, which consists of changing the pulse width of constant amplitude, used in amplifiers, switching power supplies and systems control the operation of electric motors. PWM powers the system directly or through a low pass filter which smoothes the voltage waveform or current.

Because the Pandaboard and the RoboClaw works with different logic levels, a converter is required. For this purpose KAmodLVC [15] logic level converter has been used. KAmodLVC module is an 8-bit bi-directional converted voltage levels. The converter can be used to connect two digital systems operating with different voltages (like 1.8V and 5.0V in this case).

The basic orientation sensors embedded in the robot includes a gyroscope, accelerometer and magnetometer. The sensor can be used to determine the position of the robot in two planes. The diagram of components connections and relations is presented in Fig 2. The alignment of the components in the chassis is shown in Fig 3.

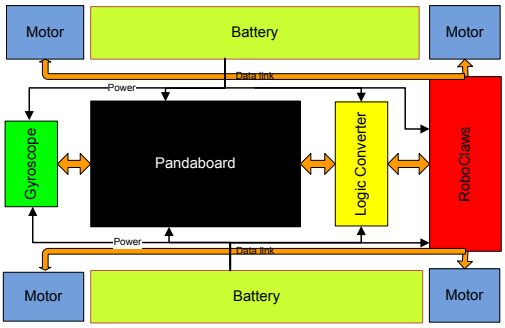


Fig. 2. The block diagram of the robot components.

The central point of control and communication is the Pandaboard. This board has several communication interfaces which are to control the robot effectors and to collect information from the sensors. Communication bus between Pandaboard and motor controller was realized using RS232 interface. For the purpose of control only lines RxD and TxD are used. There is no hardware flow control, because communication with the Pandaboard and RoboClaw is realized in inquiry respond method and it is always initiated by the Pandaboard. Therefore, if the control program waits for data from the controller it is not necessary to control rate. The data rate of this link is set to 38400bps.

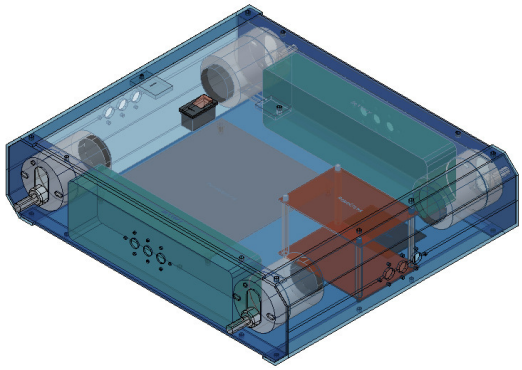


Fig. 3. Internal design of the robot components.

The orientation sensor uses serial I2C bus. To communicate with this bus the system uses duplex line Serial Data Line (SDA) and one-way line Serial Clock Line (SCL). Both lines are pull-up to power line so it is easy to detect transmissions collision using hardware. In robotic system this bus combines simplicity and functionality in one at a low investment of hardware and software to give the desired effect.

Robot communication with the surrounding environment is based on the built-in wireless card: Pandaboard WiFi. Each robot has its own unique MAC address so it is possible to communicate with the selected robot even if a group of robots is working in the same network.

Robot design provides an easy way for extending the range of sensors or effectors. It has been tested with ultrasonic sensors, laser rangefinders, cameras and Microsoft Kinect sensor. Further extensions are possible using various interfaces: USB, COM, I2C or SPI.

To determine the exact position of the robot can use the Global Positioning System (GPS) receiver or the more accurate indoor marker-based Hagisonic Stargazer [16] system. Stargazer uses markers placed on the ceiling and on the basis of their positions it can determine the location of the robot with high accuracy.

Ten units have been built so far for testing and further development purposes. The cost of all parts for a single unit does not exceed 900 USD, which is a very low price for the capabilities. The robot can develop speed of 3 m/s, it can put itself into vertical position by climbing a wall. It includes an on-board computer with 2-core CPU, running ordinary Linux OS and providing large variety of extension ports. It meets all defined requirements for testing the Erlang-based cyber-physical system for mobile robot control.

III. Control System Architecture

On the top of robot's hardware there is a need for a control software layer that allows users to interfere with it. Due to the fact that the robot was designed from scratch, control system was also chosen to be created from the ground up instead of using existing solution in order to fit the needs perfectly. Main aims of the software layer were to:

- provide high level, easy to use and consistent programming interface to low-level robot's peripherals,
- allow multiple client applications to run simultaneously on one robot, taking into account concurrency, timing, performance and other possible issues,
- allow users to write their client application in different programming languages,
- give an ability to put client application either on robot's on-board computer or on a separate network-reachable machine,
- ensure flexibility by allowing to add other external devices in the future.

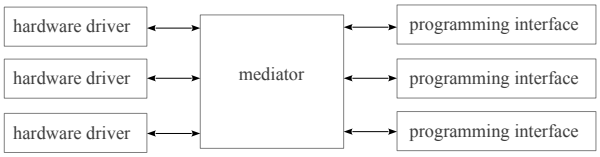


Fig. 4. Control system architecture schema

Control system has been divided into three parts (as shown on Fig 4):

- hardware drivers,
- mediator,
- programming interfaces.

Hardware drivers are standalone programs that interfere directly with robot's hardware. Being written in C++, they provide full compatibility with low-level Linux communication mechanisms.

Mediator connects hardware drivers and programming interfaces, handling communication between those two parts and controlling the whole system. It was developed in Erlang/OTP [8] due to the fact that Erlang was designed to be a solution for message passing and orchestrator applications.

Programming interfaces are libraries that end users include in their programs. They provide a consistent API to robot's hardware and can be implemented in virtually any language. There have been developed exemplary interfaces in Erlang and Java. This part of the system with be described in the next section.

System's internal communication has been based on Protocol Buffer [21] library, because it offers easy and reliable way of specifying and using custom binary protocols and has support for many popular programming languages.

A. Erlang in Embedded Systems

Erlang is a programming language created in 1986 at Ericsson Telecom AB to 'provide a better way of programming telephony applications' and "was designed for writing concurrent programs that 'run forever' " [17]. At that time telephony applications tackled atypical problems and so had unusual requirements. That applications were highly concurrent, had "soft real-time" constrains, had to be changed "on the fly" and—most importantly—had to be highly fault–tolerant, because "when the software that controls telephones fails, newspapers write about it".

Modern web servers have very similar requirements: high availability, ability to serve multiple concurrent clients, low latency and low downtime. As recent study shows [18], Erlang is well suited for such servers. It allowed writing Data Mobility server in $\frac{1}{3}$ of code and to obtain twice the throughput of C++ implementation. It's worth to note that the C++ server crashed when overloaded while Erlang just slowed down.

Since its birth, Erlang was designed as a practical tool. It is a dynamically typed, functional language with garbage collector to facilitate prototyping and ease programming. To greatly improve robustness, it implements language–level lightweight processes in shared–nothing architecture[23]. Communication is done exclusively with messages. Moreover, Erlang easily integrates with programs and libraries written in other languages. Finally it has a low memory footprint and people "successfully run the Ericsson implementation of Erlang on systems with as little as 16MByte of RAM. It is reasonably straightforward to fit Erlang itself into 2MByte of persistent storage"[24]. With all that in mind and with soft real-time characteristics of its scheduler, Erlang appears to be a perfect fit for modern embedded systems.

Embedded systems have to deal with hardware, but currently more and more sophisticated logic has to be implemented as well. Functional aspect of the language allows it to create great abstractions over hardware, algorithms and data structures. That is why it is considered that "Erlang programmers are not happy with design patterns as a convention, they want a solid abstraction"[20]. One positive effect of that is code reuse increases and the programmer can concentrate on the problem itself.

Interoperability is also very important in embedded world. Erlang has few methods for that. One of them is to write so–called "NIF"s – Native Implemented Functions. Another method is to use port drivers – communication method based on stdin/stdout streams. The latter has some advantages, most important of them is the separation of processes: even if the external program crashes for whatever reason (hardware failure or system bug), Erlang run–time is not harmed and can make attempt to recover.

In 2008 Erlang gained a SMP scheduler that allows it to scale on multiple cores/CPUs. This is a great feature, as it allows to fully use modern hardware like 64–core Parallella platform. In conjunction with multiple independent processes and message passing, this is a great advantage over most

programming languages. To compare briefly:

- Standard system processes are heavy – in practice it's not feasible to create more than few hundred of them. Erlang processes on the other hand are lightweight: each one occupies only 309 words of memory. Some tests showed that it's possible to run 136.000 Erlang processes on Raspberry Pi[19].
- Concurrency is *very hard* – while using low–level tools like locks, monitors and semaphores, programmer must deal with hard problems like deadlocks, process starvation, priority inversion. Using higher–level tools, exploiting scheduler that Erlang provides and using generic structures from standard libraries allows to avoid those problems most of the time and to facilitate reasoning about process' safety and liveness.

B. Mediator

Mediator is a central part of a system. It's a thin middleware that gives much flexibility:

- Abstracts messaging between components.
- Communicates with components using standard methods, so endpoints can be written independently in most modern languages.
- Supervises each component and takes actions in case of failures.
- It is a central part of a system – only one place where configuration needs to be done.

. During start, mediator reads configuration, creates supervision tree and spawns hardware drivers (Fig. 5). Next it runs a server for communication with, possibly remote, logic system.

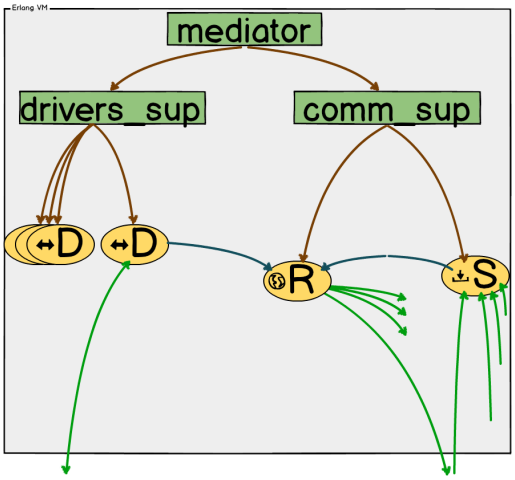


Fig. 5. Mediator is divided into: D – component mediating with software driver; R – central component, routing messages between components; S – server component, communicates with logic.

Communication with hardware drivers is done with Unix pipes. If a software or hardware has an error, mediator tries a simple yet effective tactic: restart and try again. After a number of failures in a row, it is assumed that such system is not recoverable. What is important, other subsystems are not affected and can continue to work, while defective system is turned off.

To communicate with logic, UDP protocol was chosen:

- Usually, if some part of transmission is lost, there is no need for retransmission as newer data will be available.
- UDP allows for communication locally and between computers in exactly the same way.
- When performed on localhost, packets could be lost only in case of UDP buffer overflow.
- UDP is fast and easier to use for programmers than TCP.
- Most modern programming languages have capability to communicate via UDP.

C. Hardware Drivers

As mentioned above hardware drivers are the part of the system that lays right next to robot's peripherals. There are different driver implementations for each type of supported device. They handle all low-level communication with external devices using hardware-specific protocols. This is also the only place were device logic is implemented.

Drivers are relatively simple programs spawned by the mediator and communicating with it using Unix pipes. Because software that interferes with hardware is always exposed to different kinds of failures, it is crucial to make the system as easy to recover from such issues as possible. Therefore drivers are designed as lightweight programs that can be quickly killed and restarted in case of any problems. This approach is, of course, not a perfect solutions for all types of possible exceptional situations (e.g. hardware malfunction), but makes system much more error-tolerant.

Due to the fact the drivers are separate and independent programs it is easy to add support for other devices, protocols and interfaces in the future.

All original requirements have been met in described robot's control system. The software is robust, error-tolerant, fast, flexible, perfectly suited to given hardware and ready to be used in future applications.

IV. Robot Programming Interfaces

Programming interfaces are the part of the system that end user uses directly. They communicate with the mediator using UDP sockets. Therefore client application can be run on any machine that has a network connection with robot, especially on the robot itself. UDP protocol has been chosen because it introduces small delays and low transmission overhead.

Programming interfaces can be implemented in any language that supports UDP sockets and has Protocol Buffer bindings. Thus it is possible to provide API in popular, easy to learn languages like Java or Python and allow less experienced users to work with the robot.

First implementation of programming interface was written in Java, which is a high-level, widely spread and well documented programming language that can be run on many different types of computers and other devices including

mobile phones and tablets. This fact extends the number of possible applications in which robot can be used.

Second implementation was written in Erlang. It's conceptually similar to Java's implementation, but it's written idiomatically to allow programmer fully benefit features of Erlang/OTP platform. Moreover, if mediator and logic are to be both running on the same unit, they can be run on one virtual machine, thus reducing memory usage. Finally, this allows fast prototyping and experimenting using REPL (Read–Eval–Print Loop, interactive environment with command line shell).

Apart from running programs on the robot there is a possibility of testing them in a simulation. The platform was integrated with ROBOSS simulation framework [22]. A model of a physical robot is described in XML and its visual representation in ROBOSS is shown on Fig 6.

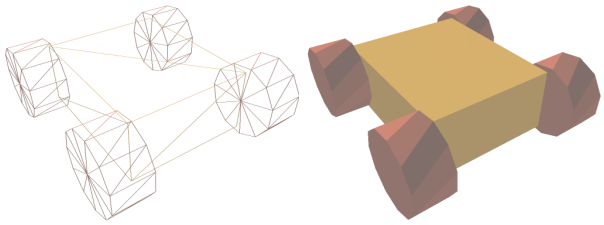


Fig. 6. Visualization of a robot model in ROBOSS simulation framework.

The use of custom Erlang module behaviour allowed to expose a simple interface which is implemented by specific Erlang modules (one for the simulation and one for the robot). As a result, the program containing logic can be run both in simulation and on the physical robot without any changes. The decision what target driver module should be used is made with regard to the configuration files.

V. Examples and Tests

In order to test the concept of building the Erlang-based cyber-physical platform and to prove that it can be used in solving real world problems, the system has been tested in a number of different applications. There were two basic groups of examples:

- on-board - when controlling program is running on robot's on-board computer,
- remote - when robot is controlled from other machine.

A. Basic Tests

In the first example robot was remotely controlled by user moving a joystick plugged into a standalone laptop computer connected to local wireless network. Moreover, real-time data read from 9DOF sensor was constantly transmitted back to the laptop and visualised on the screen as charts (see Fig 7). Therefore the user is able to see the immediate change of data charts while robot is moving.

This test showed that control software itself generates very small delays and provides enough performance to control robot manually. Actual latency is mostly dependent on WiFi

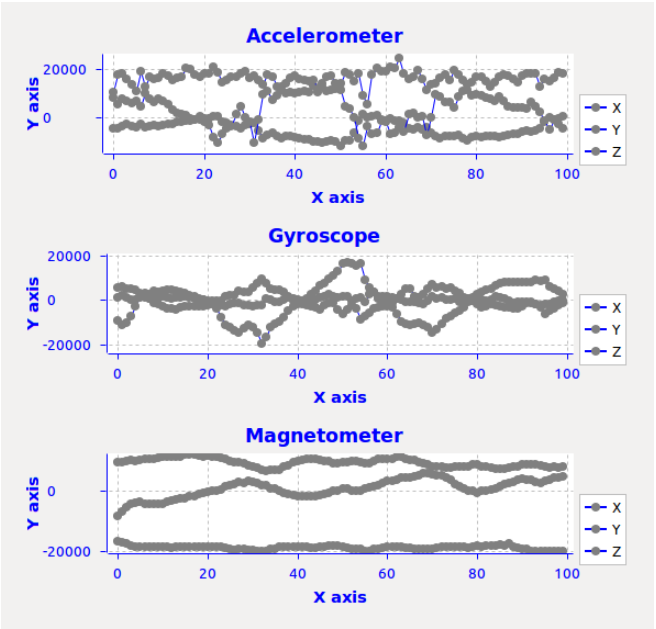


Fig. 7. Accelerometer, gyroscope and magnetometer sensors reading received from the robot during motion tests.

connection quality – some noticeable delays were observed on wireless router. This suggests that all time-critical decisions should be made on the on-board computer, while robot management or monitoring can be performed remotly.

To verify autonomous control algorithms using localization and motors controllers an advanced Trajectory Follower algorithm has been designed and implemented in Erlang. The algorithm was supposed to control robot's movements in order to reach specified locations in particular moments in time. A marker-based localization system (Hagisonic Stargazer) was used for finding current localization.

The algorithm is fully reactive. In an infinite loop it calculates most suitable control using localization and specified trajectory. The movements of the robot are smoothed according to specified algorithm parameters.

The Trajectory Follower is designed to be used both with physical robots and in a simulation. It is also desired to run on both remote and onboard nodes. Used simulation framework [22] is .NET based and this is why testing and running the trajectory follower in a simulation requires it working correctly on Windows operating system. Nonetheless, Windows OS is not required to run this component on the physical robot.

B. Trajectory Follower Algorithm

The entry point to the algorithm is a desired path to follow, expressed as a list of line segments and time constraints. On this basis, for each cycle of a control loop invoked by localisation update, desired robot speed is calculated. To preserve abstraction over the physical layer of robot, output of the algorithm is expressed as a pair of desired angular and

linear velocity. Those values are later converted to velocities on respective motors by a dedicated Erlang module, called driver.

In general, it is transparent to the driver whether it communicates with physical device or simulated robot, but it is responsible for translating control and localization. It must adapt abstract values to actual robot configuration: number of independent wheels or tracks, wheels distance and radius.

Velocity calculation algorithm is based on PD controller. Considered parameters – robot's angular distance d_α from the desired robot orientation and linear distance d_{track} from the followed trajectory, with respectful weights w_α, w_{track}, are used to obtain turn radius R:

$$\frac{1}{R} = w_\alpha d_\alpha + w_{track} d_{track} \quad (1)$$

The value of d_{track} can be treated as the P term while d_α can be treated as the D term in PD controller.

The input trajectory consists of successive path segments. The final R value is a weighted average of radiuses for corresponding segments. The number of segments taken into account and weight depend on the distance to them. Maximal cut-off distance is specified by $lookahead$ parameter in the configuration file.

Behaviour of the algorithm depends on two sets of settings: description of a robot and algorithm parameters. First one defines the name of the dedicated driver, robot's physical dimensions, localization update interval (in case of polling type of driver) and path to the simulation agent, if one is used. The latter one allows to modify weights of respective factors of PD controller, $lookahead$ parameter, maximal centripetal acceleration and maximal linear velocity.

The input can be also defined as a list of control points of Bézier splines which will result in much smoother path with no rapid turns. In this case the number of segments sampled from smoothed Bézier curve has to be defined. If the $lookahead$ parameter is too small, the robot can sometimes perform tougher turns. To ensure robot stability, centripetal acceleration of the robot must stay below certain limit.

C. Results

Example run performed in a simulation is shown in Fig 8.

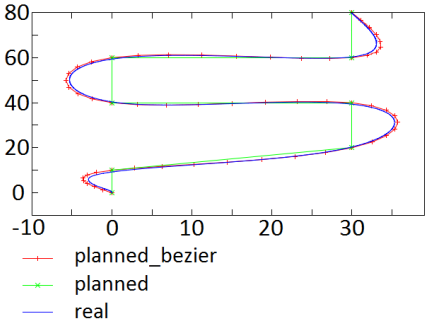


Fig. 8. Visualization of a run performed by a simulated robot. The input trajectory was smoothed with Bézier splines. Units in meters.

There was no difference or time overhead observed in communication while running the application from the remote and onboard node.

During the tests on a real robot, an issue with marker-based localization systems occurred. There were several strong light sources in the testing room. As a result, the robot tended to perform better runs with lights turned off. Exemplary run is presented in Fig 9. The robot managed to successfully read the destination within specified time, however, there is place for improvements. It is possible to reduce the noise and make measurements of localizer more precise by introducing a dead reckoning technique, i.e., applying Kalman filter.

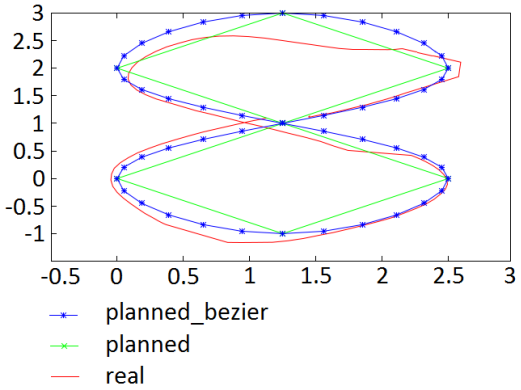


Fig. 9. Visualization of a run performed by a real robot. The input trajectory was smoothed with Bézier splines. Units in meters.

The system has demonstrated stability and performed well during the tests. It generated very small (unnoticeable) and constant delays even when several applications were using hardware components simultaneously. Erlang's functional nature seems to match high level abstraction what could be experienced during the design and implementation of the applications.

VI. Conclusions and Further Work

The implementation and performed test suggest that the Erlang language and technology is a suitable basis for building cyber-physical systems for mobile robots control. The developed system has been intensively tested in several applications. The results are promising and further work on this approach is definitely justified.

Mobile platform, which has been used for testing the approach, met all specified requirements. Designed robot is relatively inexpensive to build, offers good performance and is very easy to extend. Optional sensors and effectors will be introduced to the systems in order to increase its abilities and the range of applications. Hopefully the platform will become popular among robotics researchers.

Acknowledgements

The research leading to this results has received founding from the Polish National Science Centre under the grant no. 2011/01/D/ST6/06146.

REFERENCES

[1] C. R. Kube, H. Zhang, *Collective robotic intelligence.* Proceedings of: Simulation of Adaptive Behavior, Honolulu, Hawai USA, 1992, pp. 460–468.

[2] T. Balch, R. Arkin, *Behavior-based Formation Control for Multi-robot Teams.* IEEE Transactions on Robotics and Automation, 14, 1999, pp. 926–939.

[3] W. Turek. *Extensible Multi-Robot System.* In: Computational Science - ICCS 2008, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 574–583.

[4] W. Turek, K. Cetnarowicz, and W. Zaborowski. *Software Agent Systems for Improving Performance of Multi-Robot Groups.* Fundamenta Informaticae, 112(1), 2011, pp. 103–117.

[5] B. Gerkey, R. T. Vaughan, A. Howard, *The player/stage project: Tools for multi-robot and distributed sensor systems.* In Proceedings of the 11th International Conference on Advanced Robotics, 2003, pp. 317–323.

[6] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, A. Ng, *ROS: an open-source Robot Operating System.* ICRA Workshop on Open Source Software, vol. 3 (2), 2009.

[7] B. Nichols, D. Buttlar, J. Farrell, *Pthreads programming: A POSIX standard for better multiprocessing.* O'Reilly Media, Inc. 1996.

[8] F. Cesarini, S. Thompson. Erlang Programming. A Concurrent Approach to Software Development. O'Reilly Media, 2009.

[9] PandaBoard Documentation, http://pandaboard.org/, 05.2013.

[10] RoboClaw Documentation, http://www.basicmicro.com/, 05.2013.

[11] iRobot Products, http://store.irobot.com/, 05.2013.

[12] Avatar III Security Robot, http://robotex.com/, 05.2013.

[13] Komodo Robot Specification, http://www.robotican.net/#!komodo/c9sa, 05.2013.

[14] Husky Robot Technical Specification, http://www.clearpathrobotics.com/husky/tech-specs/, 05.2013.

[15] KAmodLVC module technical documentation, http://www.kamami.pl/dl/kamodlvc.pdf, 05.2013.

[16] J. Lopez Fernandez, C. Watkins, D. Perez Losada, M. Diaz-Cacho Medina, *Evaluating different landmark positioning systems within the RIDE architecture*, Journal of Physical Agents, 7(1), 2013, pp. 3–11.

[17] J. Armstrong. *A history of Erlang*, Proceedings of the third ACM SIGPLAN conference on History of programming languages, San Diego, California, 2007, pp. 6–26.

[18] J. H. Nyström, P. W. Trinder, D. J. King, *High-level distribution for the rapid production of robust telecoms software: comparing C++ and ERLANG*, Concurr. Comput. : Pract. Exper. 20(8), 2008, pp. 941–968.

[19] O. Kilic, *136.000 Processes on a Pi.*, http://www.erlang-embedded.com/2012/05/episode-3-\%E2\%80\%93-136-000-processes-on-a-pi/, 05.2013.

[20] F. Hébert, *Learn You Some Erlang for Great Good!: A Beginner's Guide*, No Starch Press, Incorporated, 2013.

[21] Protocol Buffer library homepage, https://code.google.com/p/protobuf/, 05.2013.

[22] W. Turek, R. Marcjan, K. Cetnarowicz. *A Universal Tool for Multirobot System Simulation*, Knowledge-Driven Computing, Springer, 2008, pp. 289-303.

[23] J. Armstrong *Proceedings of the third ACM SIGPLAN conference on History of programming languages*, ACM, 2007,pp. 6-26.

[24] Erlang FAQ, *Implementation and ports of Erlang*, urlhttp://www.erlang.org/faq/implementations.html, 05.2013.

Performance of Business Database Applications

MODERN business applications process large amounts of persistent data. Thus, a key aspect influencing the speed of such applications is the query processing time. This time can be optimized in a variety of ways, ranging from classic - such as automatic query rewriting or the usage of a variety of indexing techniques, through the variety of cashing techniques, automated database tuning, etc. On the other hand, application maintenance costs are closely related to the quality of the application architecture. Together, both those aspects lead to the search for optimization techniques that integrate with different intermediate layers and the use of alternative data persistence solutions. The development of graphics processing units and their unique computing power-price ratio also suggests searching for the acceleration of data processing based on GPUs. In addition to universal optimization techniques it is also worth looking into the use of domain knowledge in optimization mechanisms. The proposed workshop is planned to review the research on this broader query optimization process.

Topics

Topics include (but are not limited to):

- Extending the capabilities of object relational mappings
- Usage of the GPUs in query processing
- Indexing techniques
- Usage of metadata and domain knowledge in optimization processes
- Usage of fuzzy sets and rough sets in databases
- Cashing techniques
- Data snapshots
- Non-relational DBMS
- Column-oriented DBMS
- Hybrid DBMS
- Storage and analysis of time series
- Automated database tuning
- ETL optimization
- Business Intelligence optimization

Event Chair

Wisniewski, Piotr, Nicolaus Copernicus University, Poland

Program Committee

Alsabbagh, Jamal, Grand Valley State University, United States

Bała, Piotr, Nicolaus Copernicus University, Poland

Bouchakri, Rima, National High School of Computer Science, Algeria

Burzańska, Marta, Nicolaus Copernicus University, Poland

Fischer, Simon, Rapid-I GmbH, Germany

Gervasi, Osvaldo, University of Perugia, Italy

Hackney, Michael, Infobright, United States

Kaczmarski, Krzysztof, Warsaw University of Technology, Poland

Stencel, Krzysztof, University of Warsaw, Poland

Terlecki, Paweł, Tableau Software, United States

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1439–1446

On Redundant Data for Faster Recursive Querying Via ORM Systems

Aleksandra Boniewicz
Faculty of Mathematics
and Computer Sciences
Nicolaus Copernicus University
Toruń, Poland
Email:grusia@mat.umk.pl

Piotr Wiśniewski
Faculty of Mathematics
and Computer Sciences
Nicolaus Copernicus University
Toruń, Poland
Email:pikonrad@mat.umk.pl

Krzysztof Stencel
Institute of Informatics
University of Warsaw
Warsaw, Poland
Email:stencel@mimuw.edu.pl

***Abstract*—Persistent data of most business applications contain recursive data structures, i.e. hierarchies and networks. Processing such data stored in relational databases is not straightforward, since the relational algebra and calculus do not provide adequate facilities. Therefore, it is not surprising that initial SQL standards do not contain recursion as well. Although it was introduced by SQL:1999, even now it is implemented in few selected database management systems. In particular, one of the most popular DBMSs (MySQL) does support recursive queries yet. Numerous classes of queries can be accelerated using redundant data structures. Recursive queries form such a class. In this paper we consider four materialization solutions that speed up recursive queries. Three of them belong to the state-of-the-art, while the fourth one is the contribution of this paper. The latter method assures that the required redundant storage is linearithmic. The other methods do not guarantee such a limitation. We also present thorough experimental evaluation of all these solutions using data of various sizes up to million records. Since all these methods require writing complex code if applied directly, we have prototyped an integration of them into Hibernate object-relational mapping system. This way all the peculiarities are hidden from application developers. Architects can simply choose the appropriate materialization method and record their decisions in configuration files. All necessary routines and storage objects are then generated automatically by the ORM layer.**

I. Introduction

DATA models of numerous business enterprises encompass recursive data structures in the form of hierarchies and networks. They store data on e.g. railway networks, bill of material and product categorization. Their actual storage format can be chosen from a plethora of proposals [1]. There are various ways to query such data. Obviously, a dedicated 3GL client code can be written. Then, the data processing is done on the client side. In this case a significant amount of complex source code must be created, debugged and maintained. This usually causes a noteworthy increase of the budget and a shift in the delivery schedule. Therefore, a server side solution is called for. It was proposed as extensions to SQL, e.g. Oracle's CONNECT BY clause or recursive Common Table Expressions eventually adopted in SQL:1999. Such extensions have been implemented in numerous database systems [2]. Simultaneously the academia worked on optimization methods for such queries [3], [4], [5]. However, there are still database managements systems that do not support recursion in queries, e.g. MySQL. Since they are widely adopted and used, applications programmers often face the question how to query their recursive data. As noted above, they can choose to hardcode suitable logic in the application. In spite of deceptive simplicity of this solutions, it causes merely troubles: lower efficiency, increased cost and complexity, as well as reduced maintainability.

On the other hand, object-relational mapping systems (ORM) [6], [7] are a possible way to solve the above problem. They bridge the gap between data models of relational storage and object-oriented code [8], [9]. Besides this basic functionality, they also establish a thick abstraction layer that can be augmented with abundant features. In our research, we have prepared proof-of-concept extensions to Hibernate that realize recursive queries [10], [11], [12], partial aggregation [13] and functional indices [14]. In particular, we experimented with adding recursion on top of database systems that do not implement it directly [15]. In order to accelerate processing recursive queries in such a setting we proposed adding redundant data.

In this paper we describe another format of redundant data called *logarithmic paths*. Its advantage lays in its linearithmic size, while most state-of-the-art methods possibly lead to squared space complexity. We also describe our proof-of-concept implementation of this new method and three known techniques to build redundant data that facilitate recursive querying. They are *nested sets*, *materialized paths* and *full paths*. We show results of extensive performance experiments to verify the quality of these solutions. They have shown that there is no dominating method. All of them have advantages and disadvantages. We summarize them and present recommendations when each of them seems to be the most suitable.

The contributions of this paper are as follows:

- a novel (linearithmic in space) method to build redundant data that accelerate recursive querying,
- a proof-of-concept implementation of this method in Hibernate assisted with the implementation of three state-of-the-art methods,

This work was supported by the Polish National Science Centre grants 2011/01/B/ST6/03867

978-83-60810-53-8/$25.00

```
> SELECT * FROM emp;

 eid |   fname   |   sname   | bid
-----+-----------+-----------+-----
   1 | John      | Travolta  |
   2 | Bruce     | Willis    |   1
   3 | Marilyn   | Monroe    |
   4 | Angelina  | Jolie     |   3
   5 | Brad      | Pitt      |   4
   6 | Hugh      | Grant     |   4
   7 | Colin     | Firth     |   3
   8 | Keira     | Knightley |   6
   9 | Sean      | Connery   |   1
  10 | Pierce    | Brosnan   |   3
  ...
```

Fig. 1. Example persistent data on the hierarchy of employees in a company.

- a thorough experimental evaluation of the performance of these four methods,
- an analysis of their quality and circumstances under which each of them is recommended.

The paper is organized as follows. In Section II we address the related work. Section III describes the integration of the proposed method with Hibernate object-relational mapping system. In Section IV we present the new materialization method that uses only linearithmic space. Section V reports the results of an experimental evaluation of four methods to build redundant data for recursive queries. Section VI contains recommendations when each of the considered methods is most suitable. Section VII concludes.

II. Related work

Recursive relationships between entities can be implemented with an additional database table or by a single foreign key in case of hierarchies (many-to-one association). If nodes and edges are stored in the same table, querying such data can be more efficient. There are numerous optimisation methods for recursive queries [3], [4], [5]. The survey [2] summarizes implementations of recursive queries in commercial and open-source database management systems.

As noted above, a single table with self-referencing foreign key is the most straightforward way to store hierarchical data. In all sections of this paper we use a hierarchy of employees in a company as the running example. The number of levels of the hierarchy is not limited. Therefore, there exists no number n such that all leaves of the hierarchy are no further than n hops from the root. Figure 1 contains data on an example hierarchy recorded in the table `emp`. Figure 2 shows the schema of this table. The standard SQL:1999 query that retrieves all records from the subtree spanned by a particular record is presented on Figure 3.

A. Unrolling

There are database management systems that do not execute recursive queries with MySQL as the most famous example.

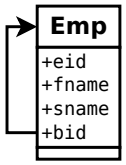


Fig. 2. The schema of the table `Emp`.

```
WITH RECURSIVE rcte (
  SELECT eid, fname, sname, bid,
         0 as level
    FROM Emp  WHERE sname = 'Travolta'
 UNION
  SELECT e.eid, e.fname, e.sname, bid,
         level +1  as level
    FROM Emp e JOIN rcte r
      ON (e.bid = r.eid)
    WHERE r.level<4
)
SELECT e.eid, e.fname, e.sname, bid
  FROM rcte
```

Fig. 3. This query retrieves all subordinates of `Travolta`.

The wide adoption of the LAMP paradigm of web development make us convinced that numerous web enterprises have to write extra code that queries recursive data. Most of such projects have similar data, e.g. nested categories of stock items, posts in discussion forums or revenue sharing chains in multi-level marketing. In order to address such applications, we have created a number of methods to run queries to such data even against database management systems that lack this feature [15]. Since the resulting solutions are non-trivial, we have prepared appropriate extensions to Hibernate. Our intent has been to hide the details from applications programmers and offer them a uniform API regardless of the chosen backend storage. Both methods considered in [15] have been integrated with API as presented on Figure 6. The first method, called *horizontal unrolling*, joins the subject table `maxlevel` times. Figure 4 shows the horizontal unrolling up to the 3$^{\text{rd}}$ level.

```
SELECT *
  FROM Emp l0
    LEFT JOIN Emp l1
      ON (l0.eid = l1.bid)
    LEFT JOIN Emp l2
      ON (l1.eid = l2.bid)
    LEFT JOIN Emp l3
      ON (l2.eid = l3.bid)
  WHERE l0.sname = 'Travolta'
```

Fig. 4. The horizontal unrolling of the query from Figure 3 up to the third level.

The other method, called *vertical unrolling*, uses temporary tables. It sends a number of queries and constructs the answer

from partial results. Both unrolling methods are notably more efficient than the potential naïve method that loops over nodes of the graphs and poses a separate query for each encountered node. Our experiments indicate that the horizontal variant is faster.

An application programmer/designer/architect chooses the required method of unrolling using the annotation `@unrolling`. Its parameter `method` can have two values: `"horizontal"` or `"vertical"`. The second (vertical) variant is the default since this method yields the same form of result as standard recursive queries. The result of horizontally unrolled query is slightly different.

B. Redundant data

If an application frequently queries hierarchical data, the abovementioned unrolling methods will not be efficient. However, in such cases designers can impose using redundant materialized data. As mentioned in Section I a number of such methods has been proposed [16]. Here we consider three of them. The first method called *nested sets* and the second method called *materialized paths* change the definition of base tables. They require adding a new column. The third method called *full paths* leaves the base table intact and puts materialized data into an additional table. In the following subsections we analyze their details.

1) Nested Sets : If the *nested sets* are used, two columns will be added to the base table. These are the columns `left` and `right`. The values of these columns satisfy the following constraints:

- $e.\texttt{left} < e.\texttt{right}$ for every tuple e,
- If a tuple e is in the subtree spanned by a tuple b, then it is true that $e.\texttt{left} > b.\texttt{left}$ and $e.\texttt{right} < b.\texttt{right}$.

Figure 5 shows the values of these two columns for the sample data from Figure 1. Arrows present how the values grow.

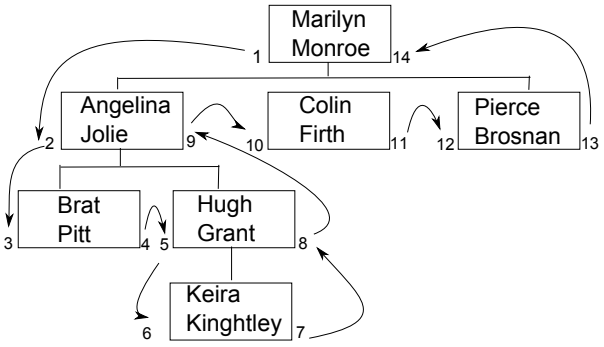


Fig. 5. Values of the redundant columns `left` and `right` in the method *nested set* computed for sample data.

The most significant advantage of this method is the possibility to query for descendants by means of a plan vanilla join. In order to find subordinates of `Travolta` we just execute the following query:

```
SELECT e.eid, e.name
  FROM emp e, emp b
  WHERE b.name = 'Travolta'
    AND e.left BETWEEN b.left AND b.right
    AND e.right BETWEEN b.left AND b.right
```

2) Materialized Paths: The method of *materialized paths* stores the whole path from the node to its root in the additional column `paths` as shown below:

```
 eid | fname |   sname    | bid | paths
-----+-------+------------+-----+------
   6 | Hugh  | Grant      |   3 | [4,3]
   7 | Colin | Firth      |   3 | [3]
   8 | Keira | Knightley  |   6 | [6,4,3]
```

This methods has proven to be noteworthy more universal that *nested sets* as discussed in Section V. Unfortunately, such materialization of paths breaks the first normal form. The following query enumerates all subordinates of `Travolta` when the method *materialized paths* is applied.

```
SELECT *
  FROM emp
 WHERE path_string LIKE (
     SELECT concat(path_string,'%')
       FROM emp
      WHERE sname = 'Travolta'
```

3) Full Paths : The method of *full paths* has been studied in [17]. That paper contains also a comparison of effectiveness of the full paths approach against unrolling methods. The full paths are similar to the *materialized paths*. However, the method of full paths stores redundant data in an extra table. We assume the name of this table to be `fullpaths`. It contains a distinct row for every step in any path towards the root. For `Keira Knightley` and `Colin Firth` the table *fullpaths* will contains the following rows.

```
 eid | bid | pl
-----+-----+----
   7 |   3 | 1
   8 |   6 | 1
   8 |   4 | 2
   8 |   3 | 3
```

The column `pl` (path length) contains the number of steps in the path.

This method is particularly universal. However, if the structure is deep, the size of the table *fullpath* can even be square with respect to the size of the base tables.

If the method of full paths is used, the query for `Travolta` subordinates will have the following form.

```
SELECT e.eid, e.name
  FROM emp e JOIN fullpaths fp USING (eid)
    JOIN emp b ON (fp.bid = b.eid)
  WHERE b.name = 'Travolta'
```

III. ORM CAN HIDE RECURSIVE PECULIARITIES

The layer of object-relational mapping [6], [7] facilitates cleaning the architecture and reducing the complexity of a system. However, an additional overhead between the application logic and the storage layers can hindrance the performance. In our research we assume that it is not the case. In our opinion an additional mapping layer *can* help optimizing the system. We can put there disparate algorithms and redundant data structures that aid improving the communication with the backend storage. The ORM layer also hides most of the peculiarities of such features from application programmers.

A. Hibernate interface for recursive queries

In our research we have presented the integration of recursive queries with object-relational mapping systems [10], [11], [12], [18]. In particular, our API for Hibernate allows defining recursive queries by XML annotations to Java entity classes [12]. Figure 6 shows a sample annotated entity class. For this class, ORM produces the table `Emp` presented in Figure 2.

```
package sample.recursive.mapping;
import org.ncu.hibernate.annotations.*;
@RecursiveQuery (maxLevel = 4)
@Tables (name = "Emp")
@RecursiveCondition (on = "Emp.bid",
                     to = "Emp.eid")
@Filter (seed = "Emp.sname = $Param(sname)")
public class Suboridnates {
    @Column(name = "Emp.eid")
    public String id;
    ...
}
```

Fig. 6. This annotation of an entity class causes generation of recursive facilities.

Consider the following scenario. The class from Figure 6 is registered in Hibernate. If the backend database connected to Hibernate implements recursive queries, the programmer can call the API function `getRecursive(String)`. This function sends the recursive query to the storage layer. If PostgreSQL is the backend, it will process the query from Figure 3.

B. Setting the support for recursive queries

If a programmer wants to use some of the above-mentioned methods to query recursive structures, e.g. unrolling or materialization, he/she just adds the annotation `@unrolling(method = "`*method name*`")`. The name of available methods are: `horizontal unrolling`, `vertical unrolling`, `full paths`, `nested sets`, `materialized paths` and `logarithmic paths`. The method of logarithmic paths is described in Section IV.

If no unrolling is specified and the database does not support recursive queries, the vertical unrolling will be used by default. More details of horizontal and vertical unrolling can be found in [15].

If a method based on materialization is chosen, the ORM layer will build the required redundant data at the first recursive access to the data. Depending on the method chosen, the main table will be altered (for nested sets or materialized paths) or an additional table will be created (for full paths and logarithmic paths). Then, redundant data get populated. Eventually, ORM automatically creates all necessary triggers. Thus, a programmer does nothing but chooses the method and specifies it in the `@unrolling` annotation.

IV. LOGARITHMIC PATHS

Section II-B presents three methods to build redundant materializations that facilitate querying recursive structures efficiently. In this Section we describe another such materialization method, called *logarithmic paths* or shortly *log paths*. This method is a kind of a compromise between full paths and vertical unrolling [15].

The idea of logarithmic paths is to store only those paths whose length is a power o 2. We assume that the data on such paths is stored in the redundant table `logpaths`. For `Keira Knightley` and `Colin Firth` the table *logpaths* will contains the following rows.

```
 eid | bid | pl
-----+-----+----
   7 |   3 | 1
   8 |   6 | 1
   8 |   4 | 2
```

If the data contains n tuples stored in trees of depth m, then the table `logpaths` will contain $O(n \log m)$ tuples. Therefore, we keep only $O(\log m)$ tuples for each arbitrary tuple of the base table, while *full paths* store $O(m)$ redundant tuples for each base tuple. For a user query the method of *log paths* issues $O(\log m)$ database queries for $O(1)$ columns, while the *horizontal unrolling* sends $O(1)$ queries for $O(m)$ columns.

A. Building the table `logpaths`

In order to populate the table `logpaths`, the paths of length 1 are copied from base table:

```
INSERT INTO logpaths
  SELECT eid, bid, 1 AS pl
    FROM emp
```

Next for $n = 1, 2, 4, 8, \ldots$ (powers of 2) the following query is executed as long as it adds new tuples. The population process will certainly finish since at each step strictly longer paths are created. Finite hierarchical data can contain only finite paths. In fact the number of those executions is $O(\log m)$ where m is the maximum depth of the hierarchy.

```
INSERT INTO powpath
  SELECT e.eid, b.bid, 2n AS pl
    FROM powpath e
      JOIN powpath b ON (e.bid = b.eid)
    WHERE e.pl = n and b.pl = n
```

B. Querying

At the query time, we have to reconstruct the information on all paths (as in *full paths*). We use the following auxiliary query:

```
SELECT lp1.eid, lpk.bid,
  lp1.pl + lp2.pl + ... + lpk.pl AS pl
  FROM logpaths lp1
    JOIN logpaths lp2
      ON (lp1.bid = lp2.eid)
    ...
    JOIN logpaths lpk
      ON (lp(k-1).bid = lpk.eid)
  WHERE lp1.pl < lp2.pl
    AND ...
    AND lp(k-1).pl < lpk.pl
```

This query reconstructs data on paths whose length (`p1`) has exactly k ones in its binary representation. Therefore, we have to run this query for $k = 1, 2, ..., \log m$ and merge their results using set union. Obviously, it will generate all paths and no path will be repeated. The resulting union query will be a part of the eventual user query. If it contains selections, the optimizer should push them down the query tree. Thus, with logarithmic paths only a fraction of the potential content of the table `fullpaths` will be actually computed.

C. Using ORM

This method has also been integrated into the Hibernate framework. If the annotation `@unrolling(method = "logarithmic paths")` is present, the table `logpath` will be automatically created and populated at the first recursive query. All necessary triggers are also created automatically.

V. Performace

The methods discussed in this paper has been tested on a computer with AMD Phenom II 3,4GHz, 8GB RAM and 2 Caviar Black 7400rpm 500 GB HDDs. The test has been run against Hibernate with a standard installation of MySQL as the backend database. We used six data sets of various sizes. Three of them contain 100 000 records organized in trees of depths 10, 15 and 20. The other three contain 1000000 records organized in trees with the same depths, i.e. 10, 15 and 20. We have tested seven usage scenarios.

The tables that report the results are organized as follows. Each table is divided into two parts. The first part presents results for data sets of 100 000 records, while the second part corresponds to data sets composed of 1 000 000 records. The first column of each table shows the depths of the trees. The second column presents times of evaluation for the presented methods. The names of two methods, namely *nested sets* and *materialized paths* are abbreviated to *ns* and *mp* respectively.

A. Building redundant materialization

In the first test we examine the time required to build redundant data structures that accelerate perspective recursive queries. We assume that the base table contains appropriate data and the derived table is empty. The results are presented in Table I.

TABLE I
Time necessary to build materialized data

100 000 rec				
	full path	logpath	ns	mp
10	00:00:47,94	00:00:15,90	00:00:29,41	00:00:12,59
15	00:01:26,20	00:00:17,67	00:00:29,26	00:00:13,23
20	00:02:54,21	00:00:21,82	00:00:29,55	00:00:13,54
1 000 000 rec				
	full path	logpath	ns	mp
10	00:22:13,21	00:04:53,27	00:34:59,03	00:05:16,97
15	01:00:24,07	00:05:30,94	00:39:47,73	00:05:10,29
20	02:03:50,40	00:06:24,73	00:39:25,69	00:05:31,39

B. Finding subordinates of root

We assume that we have an object representing a root in the hierarchy. We want to enumerate all nodes in the subtree below this root. The results are presented in Table II.

TABLE II
Time necessary to find subordinates of a root

100 000 rec				
	full path	logpath	ns	mp
10	00:00:01,57	00:00:07,99	00:00:00,51	00:00:00,56
15	00:00:01,91	00:00:08,90	00:00:00,51	00:00:00,60
20	00:00:03,64	00:00:05,27	00:00:00,51	00:00:00,64
1 000 000 rec				
	full path	logpath	ns	mp
10	00:00:30,16	00:04:32,97	00:00:10,27	00:00:09,70
15	00:00:40,09	00:07:08,23	00:00:10,96	00:00:10,09
20	00:00:49,72	00:07:44,32	00:00:10,94	00:00:10,08

C. Finding a subordinate of an arbitrary node

We assume that we have an arbitrary object in the hierarchy. We want to find an example node in the subtree below this node. The results are presented in Table III.

TABLE III
Time necessary to find subordinates of an arbitrary node

100 000 rec				
	full path	logpath	ns	mp
10	00:00:00,01	00:00:00,02	00:00:00,02	00:00:00,02
15	00:00:00,02	00:00:00,02	00:00:00,01	00:00:00,02
20	00:00:00,07	00:00:00,01	00:00:00,01	00:00:00,02
1 000 000 rec				
	full path	logpath	ns	mp
10	00:00:00,27	00:00:00,06	00:00:00,03	00:00:00,48
15	00:00:00,40	00:00:00,12	00:00:00,03	00:00:00,44
20	00:00:00,36	00:00:00,12	00:00:00,03	00:00:00,42

D. Finding the root for an arbitrary node

We assume that we have an arbitrary object in the hierarchy. We want to find the root of the tree that contains the given object. The results are presented in Table IV.

TABLE IV
TIME NECESSARY TO FIND THE ROOT FOR AN ARBITRARY NODE

100 000 rec				
	full path	logpath	ns	mp
10	00:00:00,00	00:00:00,01	00:00:00,01	00:00:00,01
15	00:00:00,00	00:00:00,05	00:00:00,01	00:00:00,01
20	00:00:00,01	00:00:00,05	00:00:00,01	00:00:00,01
1 000 000 rec				
	full path	logpath	ns	mp
10	00:00:00,01	00:00:00,01	00:00:00,01	00:00:00,01
15	00:00:00,01	00:00:00,07	00:00:00,01	00:00:00,01
20	00:00:00,01	00:00:00,08	00:00:00,01	00:00:00,01

TABLE VII
TIME NECESSARY TO MOVE A SUBTREE

100 000 rec				
	full path	logpath	ns	mp
10	00:00:08,53	0:00:26,60	00:00:19,28	00:00:00,61
15	00:00:16,01	00:00:51,14	00:00:08,70	00:00:00,64
20	00:00:52,75	00:01:28,36	00:00:24,15	00:00:01,19
1 000 000 rec				
	full path	logpath	ns	mp
10	00:04:30,89	00:13:37,25	00:00:22,51	00:00:00,47
15	00:07:49,17	00:35:06,68	00:01:20,40	00:00:00,08
20	00:13:38,99	00:57:39,43	00:00:28,07	00:00:00,92

E. Inserting new nodes

The next three tests are devoted to the assessment of the overhead imposed by all four methods when updates of the structure occur. Table V presents efficiency measures for the operation of inserting new rows into an existing base table. In this test we only added leaves to the hierarchy.

TABLE V
TIME NECESSARY TO INSERT NEW NODES TO THE HIERARCHY

100 000 rec				
	full path	logpath	ns	mp
10	00:00:00,14	00:00:00,08	00:00:05,22	00:00:00,06
15	00:00:00,20	00:00:00,09	00:00:04,35	00:00:00,06
20	00:00:00,20	00:00:00,08	00:00:16,54	00:00:00,05
1 000 000 rec				
	full path	logpath	ns	mp
10	00:00:00,15	00:00:00,11	00:07:47,67	00:00:00,05
15	00:00:00,17	00:00:00,12	00:08:27,64	00:00:00,05
20	00:00:00,16	00:00:00,12	00:08:16,85	00:00:00,05

F. Deleting nodes

In this test we examine the time needed to complete the delete operation. As in the previous test, we deleted leaves only. Deleting rows from the structure. The results are presented in Table VI.

TABLE VI
TIME NECESSARY TO DELETE A NODE FROM THE HIERARCHY

100 000 rec				
	full path	logpath	ns	mp
10	00:00:00,05	00:00:00,08	00:00:05,57	00:00:00,06
15	00:00:00,05	00:00:00,08	00:00:06,67	00:00:00,05
20	00:00:00,04	00:00:00,09	00:00:16,92	00:00:00,04
1 000 000 rec				
	full path	logpath	ns	mp
10	00:00:00,04	00:00:00,07	00:09:53,84	00:00:00,05
15	00:00:00,04	00:00:00,08	00:09:37,94	00:00:00,04
20	00:00:00,04	00:00:00,09	00:08:59,66	00:00:00,04

G. Updating nodes

In this test we measure the time needed to move a subtree to some other place. The results are presented in Table VII.

VI. ANALYSIS

In this Section we analyze the results of performance tests presented in Section V. We also formulate recommendations for which usage scenarios each of the analyzed methods is the most suitable. We presented four methods to build redundant data for efficient recursive queries. We divide them into two groups of two methods each. The first group contains methods that store data in additional tables, i.e. *full paths* and *logarithmic paths*. Since they do not require altering the original database schema, it will be easier to introduce them into an existing installation. The size of redundant data for *full paths* is $O(nm)$ where n the size of the data and m is the maximum depth of the hierarchy. In case of *logarithmic paths* this size is only $O(n \log m)$. Thus the method is feasible also in case of deep trees, i.e. when $m = O(n)$.

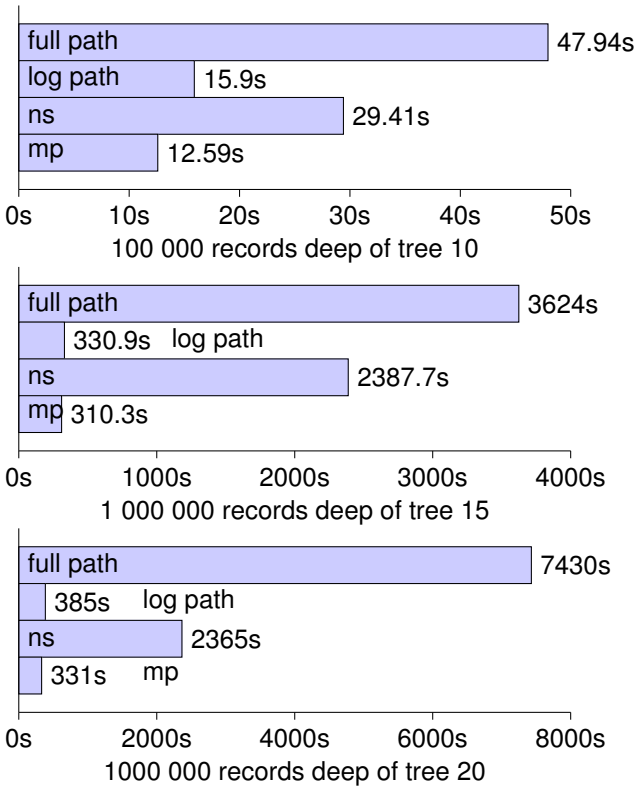


Fig. 7. The comparison of times necessary to build materialized data.

A significant advantage of *logaritmic paths* is the time necessary to construct the redundant data structure. For each analyzed size of data this method has proven to be fast and

comparable only to *materilized paths*. Figure 7 shows this comparison. Unfortunately, the efficiency of queries with this methods is notably lower. It is caused by multiple equijoins that are required to assemble all paths. For depth 20, *logarithmic paths* execute four subqueries combined by the set union. The most complex component of this union is a 4-way self join of the table `logpath`. The tests have shown that if the depth of the hierarchy grows, so does the execution time of queries. Figure 8 presents the comparison of times necessary to run the query that retrieves subtrees.

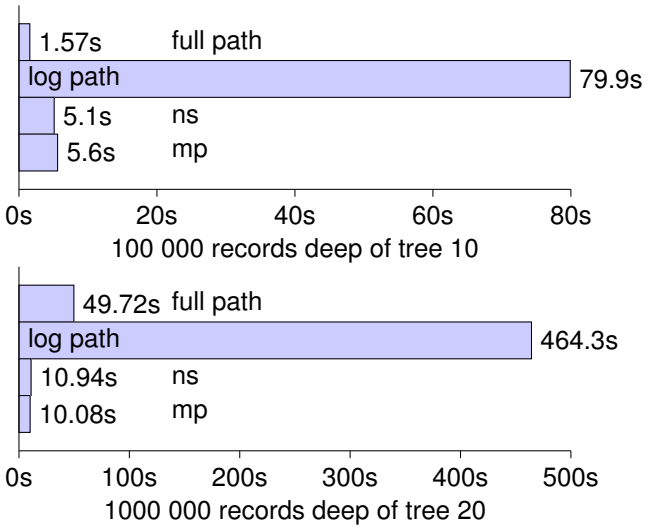


Fig. 8. The comparison of times necessary to retrieve whole subtrees.

On the other hand, *full paths* are useful when the data is collected incrementally as presented on Figure 9. Both creation and synchronization of redundant data is reasonably fast. However, in presence of deletes and updates that thoroughly *change* the structure of the tree, *full paths* are too inert and such operations are exceptionally costly.

The second group of methods modifies the base table by adding extra columns. Such an intervention into the schema may be too severe in existing deployments and probably will never be accepted in such circumstances. This group of methods contains *nested sets* and *materialized paths*. The first of them is optimized for retrieving whole subtrees. It requires careful allocation of identifiers. Furthermore, it is definitely the most expensive for maintenance in case of updates. The operations that modify the structure are two orders of magnitude slower than in case of other methods. On the other hand, *nested sets* are extremely fast for queries that search subtrees.

However, also for such scenarios *materialized paths* are slightly faster. This method is usually more efficient than all other methods. Unfortunately, it is also the only one that uses *semistructured columns*, i.e. non first normal form. If we accept this drawback, we will get the fastest querying method. Its disadvantages show at updates changing the logical structure of the tree.

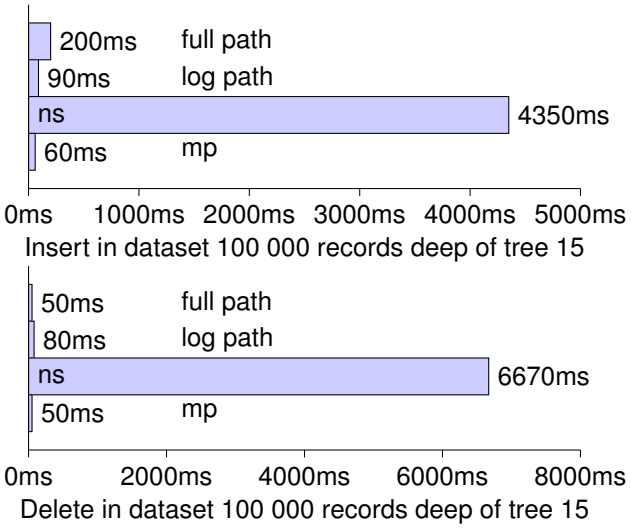


Fig. 9. The comparison of times necessary to insert and delete leaves.

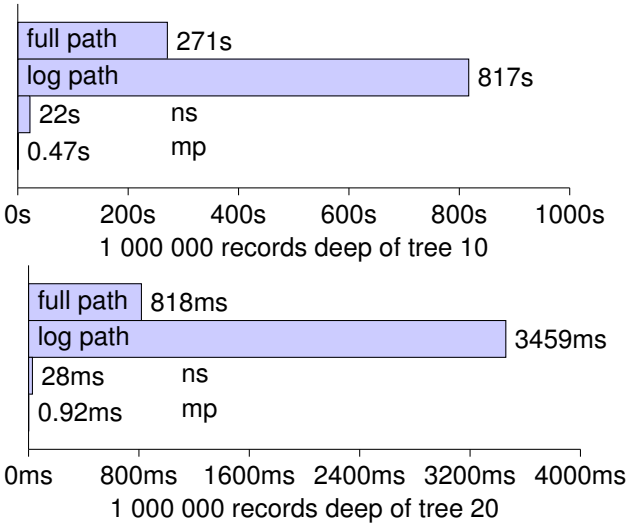


Fig. 10. The comparison of times necessary to move subtrees.

VII. Conclusions

In this paper we discussed four materialized data structures that accelerate recursive queries to hierarchical data. One of them called logarithmic paths is an original contribution of this paper. Logarithmic paths is the place where other methods meet halfway. Its efficiency is worse by a logarithmic factor than other methods that consume square space. However, logarithmic paths consume only linearithmic space.

We reported the results of experimental evaluation of all the four methods. None of them proved to be always worse or better than another tested method. For each of them, there are scenarios where it is the recommended materialization. We summarized our advises when to use each of the methods.

All these methods have been prototypically implemented as part of Hibernate, i.e. the most popular Java object-relational mapping system. This allows (1) hiding all the peculiarities of

these solutions from application programmers and (2) offering architects and tuners an easy choice of the materialization that is the most suitable for the application at hand.

REFERENCES

[1] D. Brandon, "Recursive database structures," *J. Comput. Sci. Coll.*, vol. 21, no. 2, pp. 295–304, Dec. 2005. [Online]. Available: http://dl.acm.org/citation.cfm?id=1089053.1089098

[2] P. Przymus, A. Boniewicz, M. Burzańska, and K. Stencel, "Recursive query facilities in relational databases: A survey," in *FGIT-DTA/BSBT*, ser. Communications in Computer and Information Science, Y. Zhang, A. Cuzzocrea, J. Ma, K.-I. Chung, T. Arslan, and X. Song, Eds., vol. 118. Springer, 2010, pp. 89–99.

[3] A. Ghazal, A. Crolotte, and D. Y. Seid, "Recursive sql query optimization with k-iteration lookahead," in *DEXA*, ser. Lecture Notes in Computer Science, S. Bressan, J. Küng, and R. Wagner, Eds., vol. 4080. Springer, 2006, pp. 348–357.

[4] C. Ordonez, "Optimization of linear recursive queries in sql," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 2, pp. 264–277, 2010.

[5] M. Burzańska, K. Stencel, and P. Wiśniewski, "Pushing predicates into recursive sql common table expressions," in *ADBIS*, ser. Lecture Notes in Computer Science, J. Grundspenkis, T. Morzy, and G. Vossen, Eds., vol. 5739. Springer, 2009, pp. 194–205.

[6] S. Melnik, A. Adya, and P. A. Bernstein, "Compiling mappings to bridge applications and databases," *ACM Trans. Database Syst.*, vol. 33, no. 4, 2008.

[7] W. Keller, "Mapping objects to tables. a pattern language," in *EuroPLoP*, 1997, pp. 1–26.

[8] E. J. O'Neil, "Object/relational mapping 2008: Hibernate and the Entity Data Model (EDM)," in *SIGMOD Conference*, J. T.-L. Wang, Ed. ACM, 2008, pp. 1351–1356.

[9] C. Bauer and G. King, *Java Persistence with Hibernate*. Greenwich, CT, USA: Manning Publications Co., 2006.

[10] M. Burzańska, K. Stencel, P. Suchomska, A. Szumowska, and P. Wiśniewski, "Recursive queries using object relational mapping," in *FGIT*, ser. Lecture Notes in Computer Science, T.-H. Kim, Y.-H. Lee, B. H. Kang, and D. Ślęzak, Eds., vol. 6485. Springer, 2010, pp. 42–50.

[11] P. Wiśniewski, A. Szumowska, M. Burzańska, and A. Boniewicz, "Hibernate the recursive queries - defining the recursive queries using Hibernate ORM," in *ADBIS (2)*, ser. CEUR Workshop Proceedings, J. Eder, M. Bieliková, and A. M. Tjoa, Eds., vol. 789. CEUR-WS.org, 2011, pp. 190–199.

[12] A. Szumowska, M. Burzańska, P. Wiśniewski, and K. Stencel, "Efficient implementation of recursive queries in major object relational mapping systems," in *FGIT*, ser. Lecture Notes in Computer Science, T.-H. Kim, H. Adeli, D. Slezak, F. E. Sandnes, X. Song, K.-I. Chung, and K. P. Arnett, Eds., vol. 7105. Springer, 2011, pp. 78–89.

[13] M. Gawarkiewicz and P. Wiśniewski, "Partial aggregation using Hibernate," in *FGIT*, ser. Lecture Notes in Computer Science, T.-H. Kim, H. Adeli, D. Slezak, F. E. Sandnes, X. Song, K.-I. Chung, and K. P. Arnett, Eds., vol. 7105. Springer, 2011, pp. 90–99.

[14] A. Boniewicz, M. Gawarkiewicz, and P. Wiśniewski, "Automatic selection of functional indexes for object relational mappings system," *accepted to International Journal of Software Engineering and Its Applications*, vol. 7, 2013.

[15] A. Boniewicz, K. Stencel, and P. Wiśniewski, "Unrolling SQL:1999 recursive queries," in *Computer Applications for Database, Education, and Ubiquitous Computing*, ser. Communications in Computer and Information Science, T.-h. Kim, J. Ma, W.-c. Fang, Y. Zhang, and A. Cuzzocrea, Eds. Springer Berlin Heidelberg, 2012, vol. 352, pp. 345–354.

[16] J. Celko, *Joe Celko's Trees and hierarchies in SQL for smarties, second edition*. Elsevier/Morgan Kaufmann, 2012.

[17] A. Boniewicz, P. Wiśniewski, and K. Stencel, "On materializing paths for faster recursive querying," in *ADBIS*, 2013, p. to appear.

[18] A. Szumowska, M. Burzańska, P. Wiśniewski, and K. Stencel, "Extending HQL with plain recursive facilities," in *ADBIS (2)*, ser. Advances in Intelligent Systems and Computing, T. Morzy, T. Härder, and R. Wrembel, Eds., vol. 186. Springer, 2012, pp. 265–272.

[19] T.-H. Kim, H. Adeli, D. Slezak, F. E. Sandnes, X. Song, K.-I. Chung, and K. P. Arnett, Eds., *Future Generation Information Technology - Third International Conference, FGIT 2011 in Conjunction with GDC 2011, Jeju Island, Korea, December 8-10, 2011. Proceedings*, ser. Lecture Notes in Computer Science, vol. 7105. Springer, 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1447–1454

Java Interface for Relaxed Object Storage

Michal Danihelka, Michal Kopecký, Petr Švec and Michal Žemlička
Faculty of Mathematics and Physics, Charles University in Prague
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
Email: Michal.Danihelka@seznam.cz, kopecky@ksi.mff.cuni.cz, petsvec@tiscali.cz, zemlicka@sisal.mff.cuni.cz

***Abstract*—Most development tools manipulate objects by changing values of their attributes. If the object should change more radically, problems arise. The amount of available information can vary from instance to instance and can be collected incrementally. It can happen that there exists no class suitable for all known attributes and so even movement of the instance to another class can be complicated. We can create exhaustive number of classes to cover all predicted variants, but still some other combinations of data can occur. To solve this situation, appearing often during processing of heterogeneous and mutable data, the model of relaxed objects was invented. It is based on the idea that object classes should be defined loosely in form of conditions – presumptions on data content or availability – and that instances should belong implicitly to all classes that are currently met. Methods associated with such classes assure that each instance is provided by all currently executable methods and its behavior change dynamically with changes of its content. The paper describes the Java-based object interface for this model, its effectivity, and the domain index suitable for efficient data searching.**

I. Introduction

MOST of the object models are designed for situations when we work with groups of objects described by a known set of attributes – with classes. This approach can handle changes of attribute values but can get in troubles when new attributes should be stored and handled. The practice shows that it is sufficient for many cases. The real life is however more complicated and thus the real-life objects can fit to various classes during the time. Sometimes there could be collected some additional data, while other data could be missing. Handling such situations in most class-based object models is very complicated.

Some of the situations could be expected and supported by the application (it could be expected that person can become both parent and driver at once), some could stay hidden (person being a child under 15 and pensioner at once) and could appear during the applications service. An example of an additional attribute could be an e-mail or ICQ for an application being designed sooner than these media got known enough.

The original motivation was the need to implement a historical map that should present given area as it looked like at given time in the past. All data had to be stored temporally together with the period of their validity. The obtaining of historical information from various sources was extremely complicated. A lot of information is lost and thus amount of available/known data differed from object to object and from period to period.

Therefore we wanted an application able storing and processing such data including the originally unexpected ones. Current models and tools were not matching our needs sufficiently. We therefore developed a corresponding object model together with corresponding store ourselves [1].

The *Relaxed object model* was introduced in [2], [3]. It has been developed for handling data of such extensible applications. We have tested it also for prototyping of applications handling complex data and compared it with the currently dominating approach (the use of relational or object-relational database supported by an object-oriented language). We have focused on the development speed, variability of the solution and its resulting efficiency (throughput; it could be interesting in the case when we create single-purpose applications processing large amount of non-trivial data).

II. Background

The relaxed object model presents inner content of objects as an arbitrary set of values assigned to a subset of predefined set of attributes $A = \{a_1, a_2, \ldots, a_n\}$. While the class-based object models define strict set of classes C and assigns each object instance to one of them, the relaxed object model define classes using arbitrary Boolean condition on attribute values. We can thus imagine class *Person* as any object defining values for attributes *Name*, *Surname* and *BirthDate*. Similarly as a *Driver* can be considered any object defining values for attributes *Name*, *Surname*, *BirthDate* and *DriversLicenseNr*, i.e. any *Person* with assigned driver's license. On the other hand, class *Child* can be defined as a *Person* having *BirthDate* greater than current date minus eighteen years. We can see that both *Driver* and *Child* classes are subclasses of *Person* class.

An important difference is the very loose coupling between object instances and classes. The class or classes are assigned to object instances not explicitly by its declaration, but implicitly according to their inner content. If the object gain value of another attribute or current values of the object are changed, the object can be immediately considered belonging to different set of classes.

Methods could be formally expressed as a finite set $F = \{f_1, f_2, \ldots, f_m\}$ of data manipulation functions.

Each method $f_i \in F$ defines a condition $req(f_i)$ of its executability on a given object instance. So the method *RevokeDriversLicense* can require objects having attribute *DriversLicenseNr* set. As the condition on class *Driver* implicates the condition required by a *RevokeDriversLicense* method, this method belongs to the *Driver* class interface.

978-83-60810-53-8/$25.00

Thanks to the concept of Relaxed Object, where the interface $iface(o)$ of given object instance o contains all methods $f_i \in F$ for whose its content fulfills the requirement $req(f_i)$, it is easy to access and use all methods available for a given object at a given moment. The availability can depend not only on physical availability of data, but also on their accessibility with respect to users' rights to individual attributes. Again, the client application can still provide users with the maximum available functionality.

The concept of Relaxed Object promises solving of some problems arising during transformation of real world entities to their abstract object model.

This paper describes basics of first implementation of these concepts that provides programmers writing in Java language with an extension allowing easy access and manipulation with relaxed object stored in the Oracle relational database. The environment selection has been made with respect to its potential enterprise deployment.

Following section discusses related work. Section 3 concerns on main issues of implementing relaxed objects in Java language and presents its results from performance point of view. Section 4 discusses enhancements currently available for the database layer that allows further performance enhancements. The last chapter gives the final conclusion.

III. Related Work

The model presumes data storage by individual columns. The *C-Store* database system [4] deals with this idea. It shows that this approach can be under some conditions quite efficient for data retrieval. To optimize reading operations, *C-Store* keeps data copies organized with respect to particular queries and so slows-down data manipulation. The *Datapile* system [5] targets to be a backup system for heterogeneous operational databases. It stores attribute values including their time validity into a single table. Data can be only imported from or exported to the classical operational databases. The approach does not allow direct data manipulation. A Java interface for relaxed objects [6] was designed to allow direct manipulation with stored relaxed objects, using classical relational database storage and still provide acceptable response time.

Software development method named Design by contract [7] tries to decrease complexity of implemented applications. It is based on the concept of a contract between two modules. Each module guarantees that if all requirements for input parameters are met, the output fulfills declared output conditions. This approach is not much used in the practice today. The problem is that definitions of requirements are not mandatory. Relaxed objects will require definitions of input conditions, as this test checks if the object belongs to supposed class or not and if the method is available. Still the condition check should be optional on production systems for performance reasons. It should be possible to switch the checking off in the code if it is obvious, that the condition is already met. The conditions could be, for example, already checked and the object was not changed afterwards.

As the implementation had to use relational database store, it was necessary to implement object-relational mapping between Java representation of relaxed objects and the database. Most of the known frameworks (for example *Hibernate*[1] for Java), provides more mapping methods or their combinations, as well chosen mapping can substantially increase the application performance and decrease the lag caused by SQL communication. The relaxed objects use quite different approach to data storage, none of common mappings – neither *vertical*, *horizontal* nor *filtered* [8] is applicable. It appeared more efficient to implement a special mapping, optimized for its specific object manipulation. One of the advantages is the possibility to add new attributes at the runtime. As the mapping performance was one of major risks, it was tested in detail.

IV. Relaxed Object Implementation in Java

This chapter describes basic ideas of relaxed object interface in Java. Among others, the solution handles:

- Optimal description of conditions for methods and their parameters.
- Optimal way of checking conditions at runtime.
- Possibility to find a correct implementation of a given method.
- Finding relaxed objects having required characteristics, i.e. fulfilling of a given condition.

A. Relaxed Object Representation

The representation of relaxed objects in the memory has to be chosen first. The focus was held on the speed of in-memory operations and the size of the data representation as we wanted to add as little overhead as possible.

We distinguish primitive data types and object data types in the further text. Primitive data types as int contain data value directly and do not need any space overhead. On the other hand, object data types store data wrapped into objects and reference to them using four byte references.

From the test results shown in the table I is obvious that the creation time depends on the object representation memory size. The object representation sizes are everytime rounded up to 8 Bytes. Class *Object* itself requires 8 Bytes of memory. *Integer* value represented as an Integer instance requires 16 Bytes of memory plus 4 Bytes for each reference. Fortunately, the memory overhead for objects with more attributes is relatively smaller. The results for *String* types have shown to be comparable with int objects (not counting memory for string characters), and so we tested mainly *int* parameters.

The results show average values from five runs, where each run executes the operation one hundred thousand times. The first two rows correspond to the cases where attributes were declared as public and so the manipulation can be done directly, not through getters and setters. A *final* modifier denies further changing of a value once it is already set.

The last three columns compare object creation time without attribute initialization, with attributes set by constructor

[1] http://www.hibernate.org

Table I
IN-MEMORY MANIPULATION WITH JAVA OBJECTS ([6])

Class Content	Size (B)	Get Time (ns)	Set Time (ns)	Creation Time (ns)		
					+Constr.	+Setters
public final int	16	10	n/a	—	25	n/a
public int	16	11	13	—	25	26
Int	16	11	17	18	25	42
2 x int	16	12	16	19	26	44
8 x int	40	20	29	46	51	98
32 x int	136	25	52	—	196	402
32 x int + sync. crit. section	136	89	91	—	189	417
32 x int + sync. method	136	92	95	—	194	428

parameters and with attribute initialization using setters after the object is created. The required initialization time increased with the growing number of parameters.

The last two rows show the time consumption in the case when the attribute access is exclusive and when it forbids parallel processing. The recommended variant that uses synchronized access only for critical section is better than the variant that requires synchronization for the whole method. In both cases the throughput of the initialization phase decreases significantly.

The relaxed objects allow any combination of set attributes what leads to at least $2^{|A|}$ theoretically possible classes. To avoid this, the relaxed object instance representation should allow arbitrary number of attributes and adapt its behavior according to actual inner content. We supposed that from the speed point of view the *HashMap* representation of attribute name – value pairs will be optimal while representation in array will provide more efficient storage. The *HashMap* representation was – surprisingly – 2–5 times slower than array representation for tens of attributes. Moreover the *HashMap* representation takes 80B plus additional 24B for each pair, while arrays require 24B plus 8B for each new pair.

Table II
MANIPULATION WITH RELAXED OBJECT INSTANCES ([6])

Class Content nr. of int	Get Time (ns)	Set Time New (ns)	Set Time Known (ns)	Creation +constr. Time (ns)	Creation +setters Time (ns)
1	30/30	106/106	57/57	30/58	156
2	36/37	107/152	56/60	29/71	296
4	41/54	107/181	59/61	31/89	623
8	53/65	106/322	68/75	30/133	1416
16	80/95	109/650	92/107	32/230	4527
32	104/188	109/650	120/208	30/502	11818
32*Integer	147/190	115/641	162/201	32/677	13060

Table II shows time required by the operations for an array representation. The values were of primitive type int. The last row shows a comparison with the values stored as an object type *Integer*. Most of the results is shown in the format best/worst case. Setting values distinguishes between changing an already existing value and a (slower) setting value for an unknown attribute. Again, the creation of an object instance distinguishes between the creation of the initialized instance and the creation of an empty object instance and the additional setting attributes one by one. The first case corresponds to the situation where attribute values are known in advance while the other models are more probable in the situation when the attributes are set after the instance creation one by one. This approach is up to 30 times slower than the manipulation with classical Java objects.

The overall memory consumption is approximately 2 times larger than in the case of classical object models.

It is possible to suppose that the users will spend significantly more time by reading and processing information than by its modification. Under such condition it was reasonable to implement two different interfaces of the relaxed object instance – *RFObject*[2] and its extension *RMObject*[3]. The first of them represents immutable object instance and provide methods *find*, *get*, *isSet* and *setCallMode*. The second one inherits it and adds additional methods *set*, *setValue*, *unset*, *amend*, *store*, and *delete*.

RFObject instance has all its values set during initialization, which makes checking conditions easier, as they can be checked only once. Methods of this class are thread-safe. The *RMObject* implementation is thread-unsafe and so it was implemented a wrapper that overrides all methods as synchronized. The programmer then can easily choose whether he/she prefers more efficient or safer object manipulation.

Attributes are defined as members of an enumerated type. Their data type is determined by the interface it implements. This way the compiler can check the type consistency. The description of the type is done using annotation. The definition then can look like:

```
// Example 1 - Attribute definition
public enum String implements RString {
  @Info ("State description")
  StateDescription
}
```

Getters and setters take attribute identification as their first parameter. So getters are declared as *get(RBytes): byte[]*, *get(RBoolean): boolean*, *set(RLong,long): void*, *set(RString,String): void*, etc. This allows compiler to check type consistency. Getters return primitive types to decrease memory and time overhead. The *amend(String): void* method loads attributes of object from the database. The *store(): void* method stores object including referenced objects to the database. The *unset(...): void* method marks an instance attribute as deprecated and the subsequent store invocation removes its value from the database.

The most important is the method *find(class<Type>): Type* that allows finding proper implementation of the object according to its inner state and required type. Return value is stored in the object cache. Method *setCallMode* allows programmer to choose if the next method invocation should:

- evaluate the conditions and execute the method, (the default behavior), or

[2]Relaxed Final Object
[3]Relaxed Mutable Object

- execute the method without checking, or
- only checks the executability condition.

B. Sets and Object Lists

As proposed in [2], the sets and lists are represented as collections of references to relaxed objects. Again, the implementation was split to the final *RFSet* and *RFList* interfaces, and their mutable *RMSet* and *RMList* extensions. The *unset* method marks the reference to the object as unused. To remove the referenced object as well, the implementation extends the interface by the method remove. The method *filter* unsets all objects that do not fulfill given condition.

C. Object persistency

Attribute declaration can add other requirements as uniqueness constraint or index enforcement for the attribute. An application can contain a method that goes through all the available classes, creates missing tables for the attributes in the database and registers all the attributes in the catalog. The implementation supports also transient attributes of instances that can hold temporary data during object stay in the memory and that are not stored in the database. To define a transient attribute, *Info* annotation from the example above should be changed to *Transient* annotation.

Transaction support allows setting of isolation level of the transaction or set the transaction as read only. Attributes and/or instances can be explicitly locked in either shared or exclusive mode. An application can request locking objects in the order of the increasing object ID's, which decreases probability of deadlock occurrence.

D. Methods and Class Definitions using Conditions

One of the main goals was to implement suitable way for declaration of conditions for method executability and for their parameters. As a solution the annotations were chosen similarly to attribute declaration. The next interesting problem was the enforcing of a transparent condition checking. The instrumentation of bytecode was chosen as the most feasible implementation. This special Java feature allows us to modify the bytecode of the class in the time of loading it to the memory. The last main problem was the way allowing relaxed object instances calling any method available according to their inner state. The solution through proxies was chosen here. Each method defines its interface. The Java then allows generation of an implementation for a given set of interfaces.

```
// Example 2 - Method definition
public interface Foo {
  Object bar (Object obj) throws BazException;
}
public class FooImpl implements Foo {
  Object bar (Object obj) throws BazException {
    // ...
  }
}

// Example 3 - Proxy Object Implementation
public class DebugProxy implements
  java.lang.reflect.InvocationHandler {
  private Object obj;
  public static object newInstance(Object obj) {
  return java.lang.reflect.Proxy.newProxyInstance (
    obj.getClass().getClassLoader(),
    obj.getclass().getInterfaces(),
    new DebugProxy(obj));
  }
  private DebugProxy(Object obj) { this.obj = obj; }
  public Object
  invoke(Object proxy, Method m, Object[] args)
    throws Throwable
  {
    Object result;
    try {
      system.out.println("before method " +
          m.getName());
      result = m.invoke(obj, args);
    } catch (InvocationTargetException e) {
      throw e.getTargetException();
    } catch (Exception e) {
      throw new RuntimeException(
            "unexpected invocatÝon exception: "
            + e.getMessage());
    } finally {system.out.println("after method "
            + m.getName());
    }
  return result;
  }
}

// Example 4 - Proxy Object Usage
Foo foo =
  (Foo) DebugProxy.newInstance (new FooImpl());
  foo.bar(null);
```

E. Queries

The relaxed object interface in Java tries to hide the database and provides to the developer an object oriented interface. Queries are thus provided through methods taking the conditions and returning either *RFList* or *RMList* implementations according to the necessity to modify results programmatically.

F. System Catalogue

The library keeps information about defined attributes in a class defined as one of relaxed object classes. The *RSetAttributes* attribute represents set of attributes having a defined value in a given instance. Due to the reflexivity of the system catalogue definition, the application can query the catalogue using the same way as other stored data.

```
//Example 5 - System Catalogue Class Definition
public class systemCatalogue {
  public enum String implements RString {
      @Info(value="Name of attribute", unique=true)
    RAttribute,
      @Info("Type of attribute")
    RType,
      @Info("Unit of attribute (e.g. ms,s,CZK,USD)")
    RUnit,
      @Info(
        value="Description of attribute",
        uniqu=true
        )
    RDescription
  }
  public enum Integer implements RInteger {
      @Info(
         value="Index of attribute
                for RSetAttributes",
        unique=true)
    RSetIndex
  }
```

```
  public enum Boolean implements RBoolean {
      @Info(
      "Information if values of attribute
       have to be unique"
      )
    RUnigue
      @Info(
      "Information if index is created
       for attribute"
      )
    RIndex
  }
  public enum Bytes implements RBytes {
      @Info(
     "Bit array of set attributes of relaxed object"
      )
    RSetAttributes
  }
}
```

The main weakness of the provided implementation is the performance of a persistent layer due to splitting of objects to more tables for individual attributes. Searching for k attributes of a given relaxed object requires joining of at least k tables according to equal ID. The search effectiveness in relaxed object storage without support of any specially designed index is shown on figures 1 and 2.

Results show that the time needed for searching grows significantly with the number of stored objects. The querying is complicated also due to the fact, that search can be done according to at most one indexed column and other conditions have to be checked against data stored in other tables and joined by object ID.

To allow effective searching, special type of index had to be designed and implemented. We needed index embedded to some widespread enterprise database management system to support large number of applications and application environments. We choose an Oracle database for its support of extensible indexing through data cartridges[4]. Its first implementation was described in [9]. The index has been significantly evolved and further optimized later.

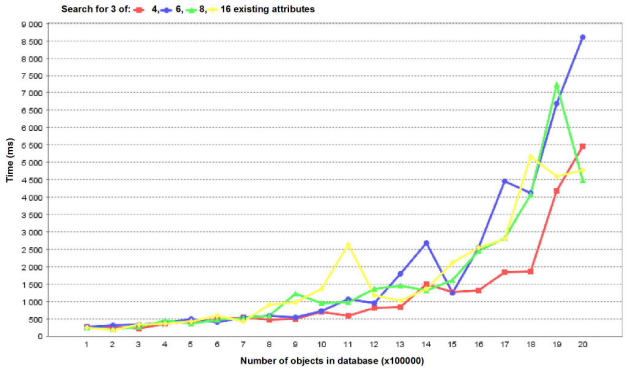


Figure 1. Object search in Relaxed Object Database - 3 attributes

[4]http://download.oracle.com/docs/cd/B19306_01/appdev.102/b14289/toc.htm

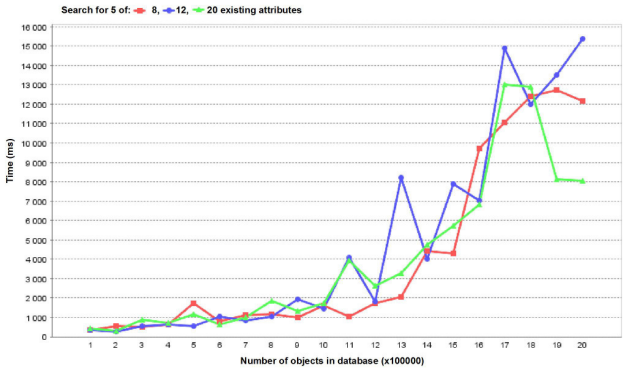


Figure 2. Object search in Relaxed Object Database - 5 Attributes

V. Temporal Relaxed Object Index

Due the restriction, given by predefined domain index interface in Oracle our indexes are created in two phases. First, they are created as *abstract* ones. In this phase the index holds only the metadata, not the data from the indexed attribute set (table set). In this state it is possible to attach new attributes to it, or detach unnecessary ones. At this moment the index is switched to *real* index, it becomes available for query optimization. Real indexes store data in R-tree structure [10] with parameters M and N. The parameters can be set at the creation time. Otherwise the optimal values are set automatically using built-in heuristics. The heuristics tries to find such a largest N that one node can be read by one I/O operation.

Let us begin with an example over Relaxed Object model storing information about people in attributes *Name*, *Surname*, *Degree*, *DrivingLicense*, *NrOfChildren*, and *Income*. The statistical office would like regularly check relation between education degree and the income, while some company would like to find potential customers. For these purposes two different multi-table indexes could be created:

```
-- Abstract index COMPANY creation
EXECUTE INX.CONSTRUCT('COMPANY');

-- Adding attributes
EXECUTE INX.ADD_ATTRIBUTE(
  'COMPANY','DrivingLicense');
EXECUTE INX.ADD_ATTRIBUTE(
  'COMPANY','NrOfChildren');
EXECUTE INX.ADD_ATTRIBUTE(
  'COMPANY','Income');

-- Switch index to real state and index data
EXECUTE INX.MAKE_REAL('COMPANY');

-- Create and populate index for statistics
EXECUTE INX.CONSTRUCT('STATISTICS');
EXECUTE INX.ADD_ATTRIBUTE(
  'STATISTICS', 'Degree');
EXECUTE INX.MAKE_REAL('STATISTICS', 20, 50);

-- Switch index back to abstract one,
-- add new attribute
-- and switch back to real index
EXECUTE INX.MAKE_ABSTRACT('STATISTICS');
EXECUTE INX.ADD_ATTRIBUTE(
  'STATISTICS', 'Income');
```

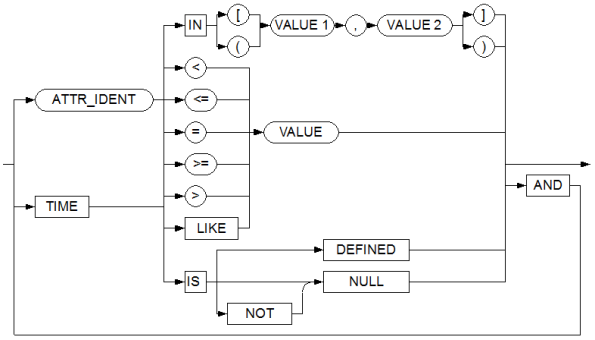


Figure 3. Query language grammar

```
EXECUTE INX.MAKE_REAL('STATISTICS');
```

All R-trees over all sets of columns are held within one domain index instance built on column ID of table $OBJECTS$. The select statements then looks like

```
SELECT * FROM OBJECTS
WHERE MATCH(ID,
  'conjunction of conditions on attribute values'
  ) = 1.
```

The query language used within the string parameter of this operator was defined to be as similar to SQL language as possible. With the exception for standard Boolean operators and comparison operators it allows easy definition of interval queries.

Index search then choose the best R-tree available to search object instances according to known attribute values. This is advantageous as the user need neither know all sets of indexed columns nor decide the specific one to be used.

Queries can either let the datastore choose optimal index, hint appropriate indexes, or force it to explicitly use given R-tree index. The SELECT statement then could look as follows:

```
SELECT *
FROM OBJECTS
WHERE MATCH(ID,
'Income >= 8000 AND NrOfChildren IN[1, 99)
 AND DrivingLicense IS NOT NULL') = 1;

-- Hinting index quality
-- by adding one point to COMPANY
-- and decreasing two points to STATISTIC
SELECT *
FROM OBJECTS
WHERE MATCH(ID,
'/* +COMPANY, --STATISTICS */ Income >= 8000
 AND NrOfChildren IN[1, 999)
 AND DrivingLicense LIKE "%B%"
 AND DEGREE = "MGR."') = 1;

-- Forcing search by index STATISTICS
SELECT *
FROM OBJECTS
WHERE MATCH(ID,
'/* STATISTICS */ Income >= 8000
 AND NrOfChildren IN[1, 999)
 AND DrivingLicense = "B" AND DEGREE = "MGR."') = 1
```

A. Search Language Grammar

The Figure 3 shows the grammar schema. The n-dimensional query is represented by a conjunction of conditions over more attributes – dimensions. Each atomic conjunction consists of the name of attribute or keyword *TIME*, of the comparison operator and of the value (1a). Each dimension can set at most one lower and at most one upper limit (1b). If the boundary for given attribute is not set, the value *MIN_VAL*, respectively *MAX_VAL*, defined internally inside the library for each database type and representing $+\infty$ and $-\infty$ is used.

The equality operator sets both upper and lower limit for given attribute (1c). To simplify interval queries, the *IN* operator was introduced. It defines both limits for the dimension delimited by a comma. Square brackets are used for closed interval boundaries, while the rounded brackets define open boundaries (1d). The value of *VALUE1* has to be less or equal to the value of *VALUE2*. The operator *IS NULL* is used to test equality with *NULL* value as it is usual in the SQL language. The *IS NOT NULL* operator is equivalent to search over closed interval *[MIN_VAL, MAX_VAL]* (1e). The *IS DEFINED* operator is used to find all objects with the value set to any value including the *NULL* value. It is equivalent to "*IS NULL* or *IS NOT NULL*". The *LIKE* operator is allowed only for textual attributes and works accordingly to its SQL equivalent. The following examples show different forms of queries.

```
-- (1a)
'NrOfChildren >= 1 AND NrOfChildren < 999
AND DrivingLicence = "B"'
-- (1b) - error: lower limit set twice
'NrOfChildren >= 1 AND NrOfChildren > 1'
-- (1c) - equiv. to 'DrivingLicence = "B" '
'DrivingLicence >= "B" AND DrivingLicence <= "B" '
-- (1d) - equiv. to 'NrOfChildren IN[1,999)'
'NrOfChildren >= 1 AND NrOfChildren < 999'
-- (1e) - equiv. to 'Degree IN[MIN_VALUE,MAX_VALUE]'
'Degree IS NOT NULL'
```

Syntax of values depends on their data type family – *integer*, *float*, *text* or *datetime*. Details for different database types are shown in Table III.

Strings have to be enclosed in double quotas ". The ordering and comparison is defined by Oracle function *NLSSORT* and is case insensitive. The date and timestamp values are enclosed in double quotas as well. If the length of the literal is shorter than expected, it is completed automatically with respect to the required format. So e.g. the value "2013-05" is by default completed to the timestamp value "2013-05-01 00:00:00.000000000".

B. Heuristics Language Grammar

All hints affecting the heuristics are optional and if present, they must be enclosed in comment brackets /* and */ at the beginning of the query string. The hint grammar is declared on Figure 4. Hint *NONE* explicitly forbids the use of any existing R-tree and data are searched programmatically. If the user wants to use one particular R-tree index, he or she can force its usage by writing its name to the comment. It is possible to write down the list of more R-tree indexes separated by commas. Each index name can be prefixed by a sequence of plus or minus signs. The heuristics then favors, respectively suppresses their usage accordingly to the prefix lengths.

Table III
SUPPORTED DATA TYPES

DB Type	Oracle format	Max	Min_Val	Max_Val
NUMBER(9)	FM999999999	10	-999999999	999999999
NUMBER(12,3)	FM999999999.000	14	-99999999.999	+99999999.999
TIMESTAMP(6)	FXYYYY-MM-DD HH24:MI:SS.FFF	21	0000-01-01 00:00:00.000	9999-12-30 23:59:59.999
VARCHAR2(16)		16	CHR(0)	16xCHR(255)
VARCHAR2(32)		32	CHR(0)	32xCHR(255)
VARCHAR2(64)		64	CHR(0)	64xCHR(255)
VARCHAR2(128)		128	CHR(0)	128xCHR(255)

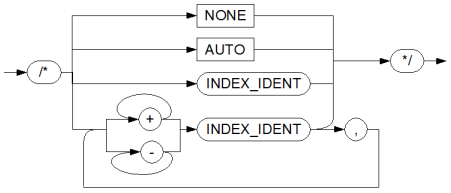


Figure 4. Hint comment grammar

C. The Embedded Heuristics

The optimal R-tree that allows optimal evaluation of given query is chosen by the embedded heuristics. It takes into account two basic factors. First it tries to avoid filter search phase. Second it tries to minimize the number of I/O operations between the R-tree stored in the BLOB and the operational memory. More formally, the heuristics tries to use such a suitable R-tree that provides maximal value of expression (1). Weights H_1 to H_4 are still subject of further research and change.

$$H_1 \frac{N_{Ind}}{N_{Attr}} + H_2 \frac{P_{Ind}}{P_{Tot}} + H_3 \frac{N}{S_{Ch}} + H_4 U_{Pref} \qquad (1)$$

The first expression $H_1(N_{Ind}/N_{Attr})$ tries to identify the biggest subset of indexed attributes. N_{Ind} represents the number of indexed attributes in the query. N_{Attr} stands for the number of attributes in the query. Optimally the number of indexed attributes corresponds with the number of attributes in the query and so $N_{Ind}/N_{Attr} = 1$. The H_1 value is currently set to 3.

The second expression $H_2(P_{Ind}/P_{Tot})$ searches for the index with as much pages read into the cache as possible. P_{Ind} denominates the number of pages already read into the index cache. P_{Tot} then holds the number of all pages that forms the index. Current value of H_2 is set to 2.

The expression $H_3(N/S_{Ch})$ prefers indexes with smaller sizes of attributes, as they can store more rows of indexed data into one index node. N is a parameter of the R-tree, while S_{Ch} stands for system chunk-size. H_3 is set to 1.

The last expression $H_4 U_{Pref}$ takes into account the user preferences. H_4 is currently set to 1. U_{Pref} is defined by the length of plus/minus prefix in the hint. For example, in the case of the hint /* ++COMPANY, -STATISTICS */ the U_{Pref} value is equal to +2 for R-tree *COMPANY* and to -1 for R-tree *STATISTICS*.

Following examples show the influence of hints and further aspects to the R-tree selection:

```
-- index COMPANY will be used (by heuristics)
SELECT * FROM OBJECTS
WHERE MATCH(ID,
'Income >= 8000 AND NrOfChildren IN[1, 9)
  AND DrivingLicense IS NOT NULL') = 1;

-- with LIKE, the STATISTICS will be used;
-- without it, the COMPANY one
SELECT * FROM OBJECTS
WHERE MATCH(ID,
'/* ++COMPANY, -STATISTICS */ Income >= 8000
  AND NrOfChildren IN[1, 9)
  AND DrivingLicense LIKE "%B%"
  AND DEGREE = "MGR."') = 1;
```

D. Combined Search of Data

All searches supported by domain index are primarily done through R-tree indexes. As R-tree indexes are not and cannot be built on all possible subsets of attributes, it is necessary to combine index search with additional comparisons of found object instances with the query.

In this case the search runs in two phases. The first index search phase – *index search* – uses the best available real R-tree index to find out object instance candidates. In the second phase – *filter search* – compares remaining attributes if they match to the user query or not. The obtained result set is then propagated to the application. The filter search evaluation is more time-consuming, but fortunately it is necessary to use it usually for relatively small amount of objects candidates only. If the sufficient R-tree index exists, it is not used at all.

E. Performance Tests

We tested the search speed using real domain indexes with optimal value of N parameter against searches, written in standard SQL both without any index support as well as using a join over *ID* columns supported by standard B-tree indexes on them. Tests run on *Oracle* 11gR2 XE on the PC with the dual-core processor Intel® Pentium® at 2.4 GHz and 4GB RAM DDR3. Third we tried to measure the search with *NONE* hint, so all instances were searched by filter search. Search times achieved over the database are shown in Table IV. Times in seconds spent by domain index search are listed in the column *REAL*. Search times needed by standard SQL search are in column *ORACLE*. The last variant is shown in column *NONE*.

```
-- REAL index STATISTICS
SELECT * FROM OBJECTS
```

```
WHERE MATCH(ID,
'/* STATISTICS */ Income >= 8000 AND DEGREE = "MGR."
AND TIME < "2100"') = 1;

-- ORACLE search
SELECT COUNT(DISTINCT Income.ID)
FROM Income I INNER JOIN DEGREE D
  ON (I.ID = D.ID
      AND ((I.VALID_TO > D.VALID_FROM
            AND I.VALID_FROM < D.VALID_TO)
           OR
           (D.VALID_TO > I.VALID_FROM
            AND D.VALID_FROM < I.VALID_TO)
     ))
WHERE I.VALUE >= TO_NUMBER('8000',
                           TYPES.GET_FORMAT('INT'))
  AND D.VALUE = "MGR."
  AND I.VALID_FROM < TYPES.TO_DATETIME(2100)
  AND D.VALID_FROM < TYPES.TO_DATETIME(2100);

-- NONE R-tree index
SELECT * FROM OBJECTS
WHERE MATCH(ID,
'/* NONE */ Income >= 8000 AND DEGREE = "MGR."
AND TIME < "2100"') = 1;
```

The database was populated with 10, 20, respectively 30 thousands of object instances with two attributes each. The attributes have had 8 to 15 unique values. Queries returned approximately 25% of the database content.

The table shows that the time spent by index creation is relatively large due to tree balancing. Optimal balancing has then positive influence to further index search times.

Table IV
SEARCH TIMES COMPARISON

Objects	Items	Create time (s)	REAL (s)	ORACLE (s)	NONE (s)
10 000	118.582	174.9	0.510	13.730	25.625
20 000	237.695	370.5	1.160	28.020	51.640
30 000	356.974	574.2	1.410	41.950	80.640

The same results are shown graphically on 5. It can be easily seen the major differences in times. Note that the results for Oracle SQL search were obtained in database without additional B-tree indexes on attribute values. Introducing them speeds up the Oracle SQL queries 3 to 4 times in the case of low dimensional queries. The queries on more attributes where the SQL join takes most of the time the speed-up was not so evident. The graph shows also an apparent increase of time between 10 and 20 thousands of object instances. It is caused by returning results in batches what implied more time-consuming calls from the PL/SQL to the C++ code. The tests were run with the batch size of 2000 objects.

VI. CONCLUSION

Using technologies as *Annotations*, *Annotation Processing* and *Java Instrumentation API*, available in Java 6.0 language, the concept of relaxed objects was successfully implemented including polymorphism that was originally not considered. Future planned extensions of Java that suppose storing method names together with parameter names and types in the byte-code in the retrievable way would make condition definitions

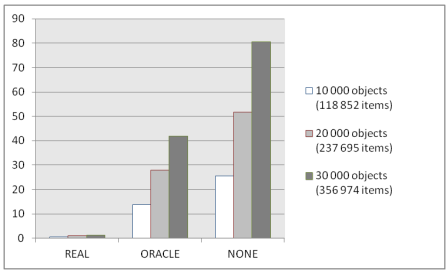


Figure 5. Search Times Comparison in Graphical Form

even simpler. The suitability of the above described implementation was proved by the pilot application – storage for software test management.

Advantage of the model is its ability of maintain heterogeneous data. The checking of method availability on given instance made the application more fault-tolerant. It appears that relaxed objects seem to be a good choice for prototyping applications having complex or heterogeneous data. Current state of implementation of specialized domain index promises even better performance and greater usability of this concept in the near future.

ACKNOWLEDGMENT

This research was partially supported by Charles University research funds PRVOUK as program P46.

REFERENCES

[1] M. Žemlička, J. Anděl, M. Bělocký, A. Buble, R. Doulík, P. Daněček, and D. Veselý, "gmap," FreeGIS CD v1.1 by Intevation GmbH, 2001.
[2] M. Kopecký and M. Žemlička, "Rozvolněné objekty (in Czech: Relaxed objects)," in *DATAKON 2004*, K. Ježek, Ed. Brno, Czech Republic: Masaryk University, 2004, pp. 243–252.
[3] ——, "Relaxed Objects - Object Model for Context-Aware Applications," in *2009 IEEE 33RD International Computer Software and Applications Conference, Vols 1 and 2*, ser. Proceedings - International Computer Software & Applications Conference, IEEE. 345 E 47th st, New York, NY 10017 USA: IEEE Computer Society, 2009, Proceedings Paper, pp. 898–903, IEEE 33rd International Computer Software and Applications Conference, Seattle, WA, JUL 20-24, 2009.
[4] M. Stonebraker, D. J. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. R. Madden, E. J. O'Neil, P. E. O'Neil, A. Rasin, N. Tran, and S. B. Zdonik, "C-Store: A Column-Oriented DBMS," in *VLDB*, Trondheim, Norway, 2005, pp. 553–564.
[5] D. Bednárek, D. Obdržálek, J. Yaghob, and F. Zavoral, "Data integration using datapile structure," in *Proceedings of the 9 th East-European Conference on Advances in Databases and Information Systems, ADBIS 2005*, Tallinn, Estonia, 2005, pp. 178–188.
[6] M. Danihelka, "Úložiště pro rozvolněné objekty (in Czech: Data Store for Relaxed Objects)," Master's thesis, Charles University, Prague, 2009.
[7] J. Rieken, "Design by contract for java - revised," Master's thesis, Department für Informatik, Universität Oldenburg, Apr. 2007.
[8] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen, *Object-Oriented Modeling and Design*. Englewood Cliffs, New Jersey 07632: Prentice-Hall, 1991.
[9] P. Švec, "Datové úložiště pro temporální rozvolněné objekty (in Czech: Data Store for Temporal Relaxed Objects)," Master's thesis, Charles University, Prague, 2010.
[10] A. Guttman, "R-trees: a dynamic index structure for spatial searching," in *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '84. New York, NY, USA: ACM, 1984, pp. 47–57. [Online]. Available: http://doi.acm.org/10.1145/602259.602266

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1455–1462

Enhanced Rough SQL for Correlated Subqueries

Marcin Kowalski*†, Dominik Ślęzak*† and Piotr Synak†
*Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
†Infobright Inc.
ul. Krzywickiego 42/219, 02-078 Warsaw, Poland
{marcin.kowalski,dominik.slezak,piotr.synak}@infobright.com

***Abstract*—We discuss some enhancements of approximate SQL extensions available in Infobright's database technology. We explain how these new enhancements can speed up execution of complex correlated subqueries, which are quite popular in advanced database applications. We compare our research to the state-of-the-art solutions in the area of analytic databases. We also show in what sense our technology follows the principles of rough sets and granular computing.**

***Index Terms*—Analytic Databases, Data Granulation, Approximate SQL, Correlated Subqueries.**

I. Introduction

COLUMNAR databases provide a number of benefits with regard to both data storage (e.g.: data compression [1]) and data processing (e.g.: on-demand materialization [2]). Their characteristics are particularly advantageous for exploratory sessions and ad hoc analytics. The principles of columnar stores can be also combined with a pipelined and iterative processing [3], leading toward modern analytic engines able to handle rapidly growing data sets.

Infobright's technology discussed in this article combines the benefits of columnar architectures with utilization of a metadata layer aimed at limiting data accesses while resolving queries [4]. In our solution, the content of each data column is split onto collections of values of some consecutive rows. Each data pack created this way is represented by its statistics. Such statistics are utilized by algorithms identifying data packs sufficient to complete particular stages of a given query execution [5].

The above solution is an example of a more general strategy of scaling complex computations on large data sets. Following the principles of rough sets [6] and granular computing [7], this methodology can be expressed by the four following steps: 1) Decompose data onto granules; 2) Create statistical snapshots for each of granules; 3) Do approximate computations on snapshots; 4) Whenever there is no other choice, access some of granules.

Certainly, such methodology should be compared to other approaches based on decomposing and merging computational tasks (see e.g. [8], [9]). It also requires addressing some details specific for particular applications, such as assuring sufficiently fast data decomposition methods, creating small but sufficiently informative snapshots, redesigning standard computational methods and accessing data granules in a minimized and optimized way.

Once the above challenges are solved, one can achieve a powerful framework where approximate computations assist execution of both standard and novel types of operations over massive data. In particular, in databases, it can be utilized to support both classical SQL statements and their approximate generalizations.

In the case of Infobright's RDBMS products, we distinguish three levels of such approximate assistance: 1) Optimization of data operations, which are components of execution of typical SQL statements [10]; 2) Computation of approximated results of SQL statements that occur in correlated subqueries in order to speed up answering to the main queries [11]; 3) Computation of approximated results visible to end users, which means actually extending standard SQL syntax [12].

In this paper we focus on the last two out of the above-outlined levels. We discuss an enhancement of our previous approximate SQL execution framework, now based on statistical metadata operations combined with accessing a small percentage of heuristically most informative data granules. We also report some performance tests proving that this new implementation can speed up a wide range of practically useful correlated subqueries.

The paper is organized as follows: Section II outlines some examples of understanding approximate SQL. Section III recalls our database technology, that is, Infobright's RDBMS solution. Section IV reports our previous research in the area of approximate SQL. Section V introduces new enhancements into our approximate SQL framework. Section VI recalls basic notions and challenges related to correlated subqueries. Section VII discusses application of enhanced approximate SQL framework to optimization of correlated subquery computation. Section VIII reports some performance tests corresponding to correlated subquery execution. Section IX discusses correlation between data sampling and precision of enhanced approximate query results. Finally, Section X concludes this study.

II. Approximate SQL

In such areas as, e.g., business intelligence or online analytics, there is a discussion whether the answers to SQL

978-83-60810-53-8/$25.00

statements have to be always exact. As an example, in the case of rapidly growing and/or dynamically changing data sets, often with a limited/variable access or limited time/budget resuorces, the outcomes of a standard SQL statement may get practically useless prior to finishing its execution. An analogous dilemma arises for SQL-based versions of some machine learning methods, which work heuristically anyway [13].

There are many aspects of introducing approximate SQL. For instance, one can estimate actual SQL results by executing queries against data samples [14]. One can also rely on data synopses [15]. A database system may build numerous synopses for various subsets of columns and measures. Each query is then translated and calculated using only synopses. The answer obtained in such a way is returned as approximation. Yet another possibility is to generalize SQL operators in order to provide end users with more flexible answers [16]. Such extensions are especially useful when query constraints turn out to be too restrictive to produce any results or columns' data types are too complex for standard conditions and operations.

The two out of the above aspects can be clearly found in Infobright's technology. Our statistical metadata layer can be utilized to heuristically identify subsets of data packs that form sufficiently informative samples. It can be also interpreted as data synopses, which produces query approximations for both internal and external purposes. We go back to these aspects in Section IV, after getting more familiar with the considered database architecture. With regard to the third above-mentioned aspect, which is tending toward more flexible answers by modifying SQL operators, some of our approximate query functionalities can be regarded as SQL syntax extensions. However, one should remember that Infobright's main inspiration for query approximations is to speed up execution and/or decrease a size of standard SQL outcomes by answering with not fully accurate/complete results.

Let us also mention about one more important direction in the area of SQL approximations, specially related to the enhancements proposed in this paper. It is dedicated to controlling a complex query execution over time, by means of converging outcome approximations [17]. Such a convergence can take different forms, e.g.: monitoring partial query results until the calculation is completely finished, with possibility to stop it at any moment in time, or pre-defining some execution time and/or resource constraints that, when reached, will automatically stop further process even if the given query results are still inaccurate. We go back to this topic in Section VI.

III. Infobright's Architecture

In Infobright's RDBMS, rows loaded into a data table are partitioned onto so called row packs, each consisting of, by default, 2^{16} of rows. Each row pack is partitioned onto data packs, each consisting of 2^{16} values of a column. Thus, each data pack corresponds to a single row pack and a single data column. Data packs are compressed and stored on disk. During query execution selected data packs are read from disk, decompressed and analyzed. Decompression stage commonly constitutes query evaluation time or is its significant factor, so one tends to limit number of data packs decompressions. That is why, among others, some subset of most recently used data packs are available decompressed in memory. Prior to compression, various types of statistics are computed for each of data packs. For each data table, there is so called granulated information system with objects corresponding to row packs and attributes – to statistics. If justified we can consider also information systems with objects related to the pairs of row packs from different tables (see e.g. [5]). Granulated information systems constitute so called Infobright's knowledge grid.

There are various strategies of partitioning rows into row packs. In a general area of data processing and mining, we may refer to this task as to data granulation [7]. Appropriate data organization will have, among others, direct impact on informativeness of data synopses and – in consequence – on engine performance. However, we need to realize that in the case of database solutions expected to analyze large amounts of data being loaded in nearly real time such granulation needs to be very fast, possibly guided by some optimized criteria but utilized rather heuristically. While loading (or reloading, e.g., by an `insert from select` operation) data, one may control the number of rows in row packs (that is, the number of rows does not need to be always 2^{16}). To a certain extent, one may also slightly influence the ordering of rows for the purposes of producing better-compressed row packs described by more meaningful statistics, following analogies to data stream clustering [18].

Knowledge grid can be treated as a metadata layer. Besides simple statistics displayed in Fig. 1, it may also contain more advanced structures [19]. However, the size of all such structures needs to be far smaller than the size of data. Since granulated tables residing in knowledge grid are organized in a columnar way, so their columns, called knowledge nodes, can be selectively employed while querying. Another metadata layer stores information about location and status of data packs. It also contains lower-level statistics assisting in data decompression and, if applicable, translating incoming requests in order to resolve them using only partially decompressed data. There is also one more metadata layer responsible for interpretation of the contents of data packs and knowledge nodes. It contains, e.g., value dictionaries for columns with relatively small number of unique values and domain-specific descriptions of values of long alphanumeric columns, which may assist in their better compression [20].

The most fundamental way of using knowledge nodes during query execution refers to classification of data packs into three categories analogous to positive, negative, and boundary regions in the theory of rough sets [6]: Relevant

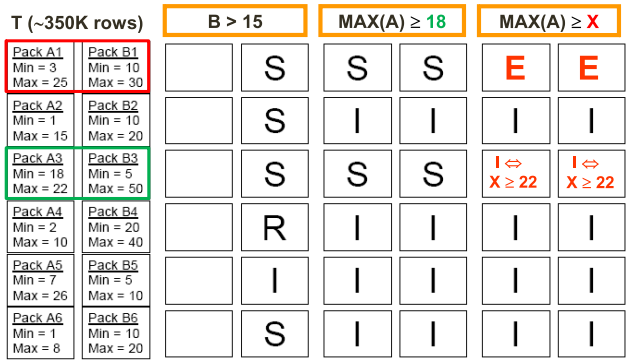


Fig. 1. Illustration for Section III. Simplified min/max knowledge nodes for numeric data columns `a` and `b` in table `T` are presented at the left side. Symbols R, S and I denote relevant, suspect and irrelevant data packs, respectively. E denotes a need of processing at the exact level. I/E means that the decision whether a given pack is irrelevant or requires exact processing will be made adaptively, depending on the outcomes of previous calculations [5].

```
select                  +--------+--------+------+
 min(a), sum(a), b      | min(a) | sum(a) |    b |
 from T                 +--------+--------+------+
 where b > 1            |      2 |      3 |    2 |
 group by b;            |      2 |      2 |    6 |
                        |   null |   null |    5 |
                        |      1 |      3 |    3 |
                        +--------+--------+------+

select roughly          +--------+--------+------+
 min(a), sum(a), b      | min(a) | sum(a) |    b |
 from T                 +--------+--------+------+
 where b > 1            |      1 |      2 |    2 |
 group by b;            |      2 |      3 |    6 |
                        +--------+--------+------+
```

Fig. 2. Example of SQL `select` statement and its rough version. Tables at the right sides display the results of both queries [10].

(R) packs with all elements relevant for further execution; Irrelevant (I) packs with no elements relevant for further execution; Suspect (S) packs that cannot be R/I-classified based on available knowledge nodes.

As an example, consider table `T` with 350,000 rows and columns `a` and `b`. We have six row packs: (A1,B1) for rows 1-65,536, (A2,B2) for rows 65,537-131,072 and so on. Consider knowledge nodes with minimum and maximum values displayed in Fig. 1. Assume there are no nulls and no other types of knowledge nodes available. Consider the following statement: `select max(a) from T where b>15;` [5]. The first execution stage uses min/max knowledge node for column `b` to classify data packs of `b` (and the whole corresponding row packs) onto relevant (B4), irrelevant (B5) and suspect (B1, B2, B3, B6) regions with respect to condition `B>15`. The second stage employs the third row pack to approximate the final result as `MAX(A)` $\geq$ `18`. Thus, only row packs (A1,B1) and (A3,B3) require further investigation. Maximum in A1 is higher than in A3, so, at the third stage, approximation is changed to `MAX(A)` $\geq$ `X`, where X depends on the result of exact processing of the first row pack. If X $\geq$ 22, then there is no need to access the third row pack. Otherwise, we need to proceed with its exact processing to get the precise outcome.

The first realistic implementation of an analogous granulated metadata support for query execution refers to [4], where the idea was to partition data into blocks of consecutively loaded rows, annotate each of such blocks with its min/max values with respect to particular data columns, and use such statistics against `where` clauses in order to eliminate out-of-scope blocks. However, the above example shows that the analysis of fully in-scope row packs is equally useful, as in such cases it is enough to take statistics instead accessing data. Moreover, it is worth re-categorizing data packs as relevant/irrelevant iteratively, which distinguishes us from other approaches.

We refer to the literature for more examples how to utilize knowledge nodes in order to speed up data operations. Let us just mention that in case of multi-column `where` conditions some data packs, within a single row pack, can be identified as relevant/irrelevant while others may remain suspect. Furthermore, statistics and statuses of different data packs can be combined while processing complex (boolean, arithmetic, etc.) expressions treated as dynamically derived data columns [10]. (Actually, results of correlated subqueries investigated in further sections may be interpreted as such complex expressions.) It is also worth noting that knowledge nodes can be computed and efficiently used for intermediate results and structures created during query execution, such as e.g. hash tables storing partial outputs of joins and aggregations [21].

IV. Infobright's Query Approximations

Let us now go back to an outline of approximate SQL approaches that we developed so far. In [5], we introduced informally the notion of rough query, in order to better explain some internal data operations in our database engine. In [22], rough query was formalized for the purposes of correlated subquery optimizations. This topic was continued in [11]. Finally, in [12], we formulated the syntax of rough SQL statements and their results.

The following aspects of rough SQL were studied in [12]: 1) Query execution algorithms; 2) Internal results format; 3) Reporting results to end users. Let us explain those aspects using an example from Section III. Recall that by utilizing min/max knowledge nodes of columns `a` and `b` we could compute that the outcome of `select max(a) from T where b>15;` is at least 18. This is because of row pack (A3,B3), within which there must be at least one row satisfying condition `b>15` and having value not less than 18 on `a`. We obtain approximation in the form of interval $\langle 18, 25\rangle$. This would be actually the answer to query `select roughly max(a) from T where b>15;` within the rough SQL framework introduced in [12].

Every `select` SQL statement returns a set of tuples

labeled with the values of some attributes corresponding to the items after `select`. Approximation of a query answer can be specified as statistics describing attributes of such a tabular outcome. Rough SQL can be assumed to produce a kind of knowledge grid for such attributes. Accordingly, the results of rough queries in [12] were provided as ranges $\langle lower, upper \rangle$ approximating the results of the corresponding standard queries. Here we should notice that, as for now, our approach concerns only numeric columns, however some efforts of analogous representation for alphanumeric columns have been done too.

Consider an example in Fig. 2, where the task is to compute aggregations `min(a)` and `sum(a)` with respect to column `b`. The resulting tuples correspond to the groups induced by `b`. There are three attributes: `min(a)`, `sum(a)` and `b`. Rough version of the same SQL computes ranges for two aggregations and the grouping column. Rough outcome tells us that for each resulting tuple, if its value of `min(a)` is not `null`, then it is for sure between 1 and 2. Similarly, `sum(a)` is in $\langle 2, 3 \rangle$, and `b` is in $\langle 2, 6 \rangle$.

In parallel to our research on rough SQL, in [19] we started to extend Infobright's framework toward a kind of inexact querying. Those studies were further continued in [23]. In this approach, we do not compute query result approximations, neither with full nor partial confidence that they are correct. Instead, we attempt to speed up the query execution process by quickly producing potentially inaccurate results that look in a standard way.

In this case, we follow an expectation that approximate queries yield tuples being almost the same as those resulting from standard queries. We enriched knowledge nodes and the corresponding query processing functions in order to compute degrees of data packs' (ir)relevance. We also investigated inexact knowledge nodes that describe data packs almost correctly, neglecting local outliers in order to provide crisper min/max intervals. Then, basing on some analogies with extensions of rough sets (see e.g. [24]), we loosened the criteria for R/I pack status, i.e., we treated almost not suspect packs as if they were truly (ir)relevant . This way, we created a framework for computing inexact query outputs with a decreased need for data access.

In [23], we also proposed a more probabilistic approach, where so called degrees of (ir)relevance, i.e., dynamically derived coefficients that heuristically estimate how many elements of particular data packs might satisfy query conditions, are employed to randomly select row packs for further processing. In this case, we can work with standard knowledge nodes [5]. However, with some probability proportional to the above degrees, data packs that are potentially almost relevant or irrelevant may be classified as fully relevant or irrelevant, respectively.

The above randomized approach inspired us to develop a number of intelligent sampling techniques. Given the architecture described in Section III, it is important to randomly select a reasonably small subset of row packs and then choose a sample of rows from those row packs only. A challenge is to select row packs providing sufficient representativeness of the final sample. In this case, utilizing knowledge nodes in order to compute degrees of row packs' (ir)relevance with respect to particular query conditions is very helpful. While producing a sample, fully relevant data packs should have the highest chance to get selected while fully irrelevant data packs should not be considered at all. For suspect data packs, their degrees of (ir)relevance should directly influence the probability of taking them into account during sample generation. In further sections, we refer to queries executed over samples generated in this way as to sampled queries.

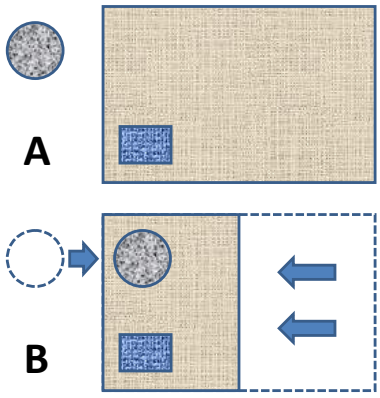


Fig. 3. A: Comparison of a result of a standard `select` statement (symbolized by smaller rectangle) with the results of its sampled (circle) and rough (bigger rectangle) versions; B: Illustration how a sampled result can be modified using knowledge nodes and how rough result may become crisper by accessing some of data packs.

V. Enhanced Rough SQL

Rough SQL reported in [12] provides fast responses, but the spans of approximations are often not informative enough to end users nor internal optimization mechanisms. On the other hand, sampled queries discussed in the end of Section IV provide fast responses (although not as fast as rough queries) that are potentially more informative, but end users have no means to analyze how accurate they are (see Fig. 3A). Certainly, it is crucial to approximate errors occurring for particular data packs and to propagate them through the whole query execution process in order to provide end users with overall estimation. This task gets more complicated along a growing complexity of analytical SQL statements and needs to be considered for all query methods. Thus, a question arises whether it is possible to combine benefits of rough queries and sampled queries within a unified framework.

The above question can be answered in at least two ways (see Fig. 3B). One of possibilities is to enrich sampled queries with an analysis whether their results remain within the bounds produced by rough versions of the same queries. If a given sampled result is not within such bounds, we can quickly move it toward the closest *edge* of rough SQL approximation. Another possibility is to

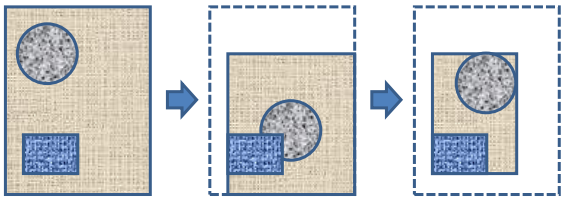


Fig. 4. Illustration how enhanced sampled query results (circles) and enhanced rough query results (bigger rectangles) are expected to behave with respect to a standard query result (smaller rectangle) while accessing more and more data packs.

develop a crisper version of rough SQL that would provide better ranges for both sample-based and fully exact querying, by combining statistical information provided by knowledge nodes with crisp information about opened data packs. Actually, in this case one should expect that rough SQL enriched by additional information about the content of a subset of suspect data packs should provide reasonable ranges for sampled SQL, where a sample is created basing on the same subset of data packs.

From a technical point of view, sampled querying is based on selecting subsets of row packs and then selecting some rows from those row packs. Therefore, it requires decompression of those row packs. Once we have some row packs decompressed, their content can be used to improve rough SQL results in the same time. Realization of this idea is actually very simple: When a data pack is accessed, we can replace its statistics stored in Infobright's knowledge grid with very crisp information about only those of its elements, which are useful at a given stage of query execution (e.g.: which correspond to rows satisfying conditions over other columns in a given `select` statement). This means that knowledge nodes, or rather their copies utilized while executing a given query, can become crisper over time, leading to a crisper final result.

Let us emphasize that from a practical point of view there is no requirement for results of sampled queries to fall within the bounds provided by their rough counterparts. It depends entirely on expectations of end users. For distinction, we will refer to a sampled query mechanism enriched by rough results analysis as to an enhanced sampled querying. Analogously, we refer to a rough query mechanism enriched by information from a subset of suspect data packs as to an enhanced rough querying.

Let us go back for a while to the idea presented in [12], where rough SQL was based entirely on Infobright's knowledge grid. We implemented a number of techniques utilizing knowledge nodes at particular stages of execution of `select` statements. All of them are based on heuristics analogous to the mechanisms of dynamic approximation of standard query outcomes, as shown in Section III. Approximations are often not perfectly precise but can be obtained very fast. However, additional data access may not necessarily lead to a dramatic slowdown of rough SQL execution. For instance, it is worth noting that Infobright caches recently used data packs in memory. Integration of information residing in knowledge nodes with such packs within a framework for enhanced rough SQL may significantly improve precision of results.

Let us also recall that Infobright's query execution should be treated as an iterative process, where knowledge nodes and information acquired from data accessed at a particular stage can be employed for heuristic selection of the very next packs that are likely to mostly contribute to narrowing down the $\langle lower, upper \rangle$ ranges. In general, we can think about several strategies of choosing data packs or row packs to be processed. For instance, to improve enhanced rough SQL results with minimum data accesses, we should choose packs that are expected to maximize information gain for after-select attributes (which may follow exactly the same heuristics as in the example in Section III). On the other hand, in order to increase reliability of sample-based querying, we should rather choose row packs that seem to be statistically representative with respect to the query domain (as discussed in the end of Section IV). Finally, in order to speed up execution without losing quality, we should attempt to combine the above strategies with information about data packs that are currently cached in memory.

Finally, let us note that enhanced rough query result ranges become crisper (or at least not less crisp) when more row packs are taken into account. There is no guarantee that a distance between actual query and approximated query results decreases monotonically. However, while enhanced rough query result ranges become crisper, we intuitively expect to approximate the above distance better and better (see Fig. 4). As mentioned in Section II, there are also some approaches where an end user provides an upper bound for query processing time and acceptable nature of answers (partial or approximate). One can imagine an analogous framework designed for Infobright, wherein a query is executed starting with rough information and then it is gradually refined by decompressing heuristically selected pieces of data. The execution process can be then bounded by means of various parameters, such as time, acceptable errors, or percentage of data accessed. We can also consider a design of Infobright-specific incremental querying, although it would require further extensions of end user interfaces.

VI. Correlated Subqueries

A correlated subquery is a query nested in some other query, called an outer query, which often operates on another table, called an outer table, and which is parameterized by some outer table's values. Correlated subqueries can occur in a number of SQL clauses. For illustrative purposes, consider the following example of the `where` clause of the outer query: `U.x=(select max(T.a) from T where T.b>U.y)`. If `U` and `T` are large, then the execution of the outer query on `T` may be time consuming. However, as pointed out in [22], one can quickly derive rough answers

to particular subqueries and, for each row in `T`, check whether the above condition could be successfully resolved by using such dynamically obtained statistics.

Importance of correlated subqueries is visible in several areas of business intelligence, such as the trend or predictive analysis. Complex nested queries may occur at the reporting stages focused on identifying anomalies and risk indicators. They may be useful also in applications requiring storage of large sets of hierarchical objects in a relational format. In many cases, e.g. for large metadata repositories, a dynamic reconstruction of such objects may be also supported by massive self-join operations or some recursive SQL extensions, if available. However, there are cases where the usage of correlated subqueries seems to be particularly convenient, if they perform fast enough.

Internal rough query algorithms have been utilized in Infobright's RDBMS for a longer time in order to speed up correlated subqueries [11]. Surely, one can adopt here some well-known methods of the correlated subquery optimization [25], such as caching results for last-observed parameters or rewriting nested queries into semi-joins. However, in some cases, the nested part of a query may still need to be processed in its direct form for a huge amount of rows of an external table. Fast production of query approximations may eliminate the need of a standard computation at least for a subset of such rows.

For a subquery in the `where` clause, we launch its rough version with parameters induced by each consecutive row in the outer table. We attempt to use its rough outcome to avoid the exact mode of execution. For the above-mentioned example of the clause `U.x=(select max(T.a) from T where T.b>U.y)`, we can express it in two stages: $(x, \widetilde{s}(y))$ and $(x, s(y))$. The first of them symbolizes, for a given row, the comparison of its value on column x with a rough result of the subquery parameterized by its value on column y. If such a comparison is not sufficient to decide whether that row satisfies the `where` clause, i.e., if the value of `U.x` does not yield fully relevant or fully irrelevant condition when comparing with the result of statement `select roughly max(T.a) from T where T.b>U.y;` then, for that particular row, we need to proceed with the standard subquery execution $s(y)$.

A broader roadmap for rough query-related future optimizations of correlated subqueries was presented in [11]. In all cases presented in that paper, the preciseness of rough query outcomes is crucial for ability to minimize data access and speed up computations.

VII. Optimization of Correlated Subqueries by Enhanced Rough SQL

Efficiency of different forms of approximate or randomized query frameworks can be examined in multiple ways. For example, in [23] we investigated stability and reliability of the results of top-k queries while tuning parameters of random selection of almost suspect data packs. As another example, in [12] we presented some strategies utilizing rough queries to speed up standard SQL calculations. In this paper, we focus on the original idea of using rough SQL to improve performance of correlated subqueries [22]. Our goal is to show that additional employment of enhanced rough SQL mechanisms described in Section VI can lead to further gains in this area. In the nearest future, we plan to extend our tests also onto other aspects of practically most promising enhanced rough SQL applications.

TABLE I
The amount of data packs accessed while querying. The TPCH100 query is a standard benchmark query. The TEST_1 and TEST_2 queries come from Infobright's internal benchmark framework. The SYNAT_1 and SYNAT_2 queries are taken from the SYNAT project [26]. The remaining queries represent Infobright's customers.

SQL statement	rough → exact	rough → enhanced → exact
TPCH100	2869154	185330
TELCO_1	21663	12087
TELCO_2	9579	3
WEBLOGS	51047	28769
TEST_1	40707	40707
TEST_2	40707	3
SYNAT_1	333582	333552
SYNAT_2	252868	251242

At the first glance, one may suspect that accessing data packs slows down producing rough subquery outcomes. However, crisper outcomes let us eliminate more calculations at a higher level. Moreover, we implemented it in such a way that the enhanced rough query execution is triggered only if standard rough query fails. In our previous implementation, we would follow with exact subquery immediately. In the new implementation, we proceed with exact computation only if enhanced rough query fails to give us results that would be precise enough. Therefore, it is not about a trade-off between the amount of accessed packs and precision of results – the way it was implemented assures that we always win by applying this new mechanism. Symbolically, we replace the previous strategy expressed by the pair of stages $(x, \widetilde{s}(y))$ and $(x, s(y))$ by a triple, where $(x, \widetilde{s}(y))$ is followed by $(x, \overline{s}(y))$ and then by $(x, s(y))$, where $\overline{s}(y)$ denotes an enhanced rough execution of inner query s with parameters set up by a value of outer column y. Surely, the other strategies reported in [11] could be extended this way as well.

Mechanism of enhancing is parameterizable. Before running query we can choose how many row packs for single subquery evaluation are permitted to be additionally decompressed. If not specified explicitly, all experiments described in the following section were performed for the only one row pack per subquery evaluation permitted.

VIII. Experimental Results

Table I illustrates one of specific classes of correlated subqueries aimed at selecting and processing extreme cases for some categories. Such nested queries differ from

relatively simpler `group by` statements in a way of considering observations marked as extreme. They are useful for some aspects of time-oriented analysis, extreme behaviour or outliers tracing. Specific tasks corresponding to such queries may look, e.g., as selecting all latest news about each company in the news table

```
select * from news n1 where pub_date
  = ( select max(pub_date) from news n2
      where n1.company_id = n1.company_id )
```

or, for a set of stemmed documents, selecting all stems that are of a maximal TF-IDF value for particular documents (see [26] for comparison)

```
select stem from stemmed_docs x where tf_idf
  = ( select max(tf_idf) from stemmed_docs
      where x.id_doc = id_doc )
```

The results in Table I are displayed by means of data packs that had to be opened. Such measure for considered class of queries, reflects real performance of query evaluation fairy accurately. We will then identify number of decompressed data packs with query performance understood as time spent waiting for the results.

The last column reports the results of our new approach, where the intermediate enhanced rough SQL version of a correlated subquery is computed with only one heuristically selected row pack to be additionally accessed. This shows a huge potential of the proposed method. For some of the considered queries, we were able to provide and test equivalent non-nested `select` statements with additional `group by` or `self-join` operators. Usually, such rewriting significantly improves performance. However, for queries such as TELCO_2 and TEST_2, it was impossible to find an alternative way of execution that would be comparably fast to the strategy based on enhanced rough SQL.

IX. Discussion

Let us now elaborate more on how the data sampling intensity correlates with the precision of enhanced rough query results. First of all, it turns out that queries with correlated subqueries based on MIN/MAX aggregations can particularly gain from enhanced rough SQL as such aggregations are sensitive to information about single observations. Figure 5 illustrates it by presenting enhanced rough queries for selected values of additional decompressions of row packs per query evaluation. Parameter equal to 0 means an ordinary rough SQL (with no extra data accesses). We can see that after accessing nine additional row packs we achieve fully crisp result.

On the contrary, the precision of, e.g., COUNT is expected to be linear with respect to the number of additional accesses [23]. It makes the enhanced rough query sampling strategy for COUNT (as well as SUM and AVG) questionable, although in some cases accessing several data packs may still help in resolving a query condition. Generally, these types of aggregations may be handled better by enhanced sampled queries, although inexactness of their outcomes is not always acceptable by end users.

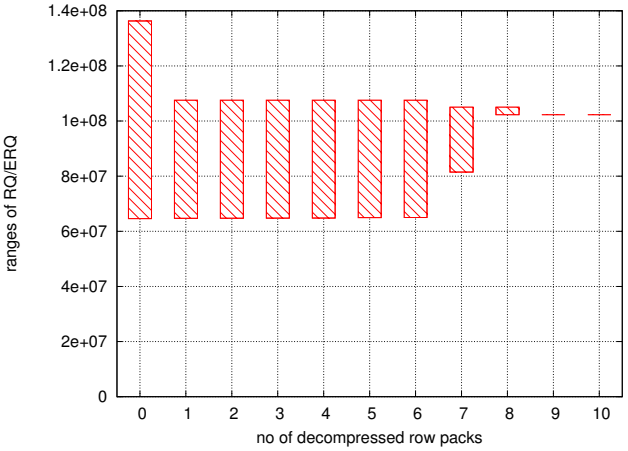


Fig. 5. Exemplary scenario of influence of increasing number of row pack decompressed during enhanced rough query (ERQ) calculation on ranges returned by ERQ performed on customer's data (SELECT ROUGHLY MIN(col) FROM t WHERE cond); 0 stands for original rough query; for 9 row packs we achieve crisp result

It is also worth elaborating on sampling strategies that could decrease a need of additional decompressions by, e.g., achieving crisp results after shorter period of time. For example, one of such strategies for MIN/MAX aggregations is to start decompression from data packs with the most extreme values represented by their knowledge nodes (like in Figure 5). A similar strategy could be followed for COUNT DISTINCT aggregations by referring to data packs with the highest estimated number of distinct values. However, analyzing results from Figure 5 (in particular the little improvement in range estimation between cases with parameter set to 1 and 6) we anticipate the need of enrich sampling strategy development of utilizing information contained e.g. in histograms.

Additional decompressions may be done randomly but a more intelligent procedure is recommended. The presented results also suggest that in case of correlated subquery optimization such mechanism may be complementary to syntactic reformulation of queries (for instance into JOINs). It should be also pointed out that some on-load row reorganization strategies aimed at improving the quality of knowledge nodes may further increase effectiveness of sampling mechanisms [18]. This is because, generally, the increase in the quality of rough information represented by knowledge nodes helps a lot in all aspects of Infobright's querying algorithms.

X. Conclusions

We discussed some enhancements of rough SQL framework developed by Infobright. We reported how such enhancements can speed up execution of correlated subqueries. In our future research, we are going to examine

further cases of practical usage of enhanced rough SQL. We will also attempt to better analyze convergence of query approximations depending on various strategies of sampling blocks of rows during query execution.

References

[1] P. White and C. French, "Database System with Methodology for Storing a Database Table by Vertically Partitioning all Columns of the Table," US Patent 5,794,229, 1998.

[2] D. J. Abadi, D. S. Myers, D. J. DeWitt, and S. Madden, "Materialization Strategies in a Column-Oriented DBMS," in *ICDE*, 2007, pp. 466–475.

[3] P. A. Boncz, M. Zukowski, and N. Nes, "MonetDB/X100: Hyper-Pipelining Query Execution," in *CIDR*, 2005, pp. 225–237.

[4] J. K. Metzger, B. M. Zane, and F. D. Hinshaw, "Limiting Scans of Loosely Ordered and/or Grouped Relations Using Nearly Ordered Maps," US Patent 6,973,452, 2005.

[5] D. Ślęzak, J. Wróblewski, V. Eastwood, and P. Synak, "Brighthouse: An Analytic Data Warehouse for Ad-hoc Queries," *Proc. VLDB Endow.*, vol. 1, no. 2, pp. 1337–1345, 2008.

[6] Z. Pawlak and A. Skowron, "Rudiments of Rough Sets," *Information Sciences*, vol. 177, no. 1, pp. 3–27, 2007.

[7] A. Bargiela and W. Pedrycz, *Granular Computing – An Introduction*. Kluwer Academic Publishers, 2002.

[8] Y. Bu, B. Howe, M. Balazińska, and M. D. Ernst, "HaLoop: Efficient Iterative Data Processing on Large Clusters," *Proc. VLDB Endow.*, vol. 3, no. 1, pp. 285–296, 2010.

[9] M. Szczuka and D. Ślęzak, "How Deep Data Becomes Big Data," in *IFSA-NAFIPS*, 2013.

[10] D. Ślęzak, P. Synak, J. Wróblewski, J. Borkowski, and G. Toppin, "Rough Optimizations of Complex Expressions in Infobright's RDBMS," in *RSCTC*, 2012, pp. 94–99.

[11] D. Ślęzak, P. Synak, J. Borkowski, J. Wróblewski, and G. Toppin, "A Rough-columnar RDBMS Engine – A Case Study of Correlated Subqueries," *IEEE Data Eng. Bull.*, vol. 35, no. 1, pp. 34–39, 2012.

[12] D. Ślęzak, P. Synak, G. Toppin, J. Wróblewski, and J. Borkowski, "Rough SQL – Semantics and Execution," in *IPMU, vol. 2*, 2012, pp. 570–579.

[13] J. Wróblewski, "Analyzing Relational Databases Using Rough Set Based Methods," in *IPMU, vol. 1*, 2000, pp. 256–262.

[14] S. Chaudhuri, G. Das, and V. Narasayya, "Optimized Stratified Sampling for Approximate Query Processing," *ACM Trans. Database Syst.*, vol. 32, no. 2, p. 9, 2007.

[15] K. Chakrabarti, M. N. Garofalakis, R. Rastogi, and K. Shim, "Approximate Query Processing Using Wavelets," *VLDB J.*, vol. 10, no. 2-3, pp. 199–223, 2001.

[16] S. Zadrożny and J. Kacprzyk, "Issues in the Practical Use of the OWA Operators in Fuzzy Querying," *J. Intell. Inf. Syst.*, vol. 33, no. 3, pp. 307–325, 2009.

[17] S. Chaudhuri, V. R. Narasayya, and R. Ramamurthy, "Estimating Progress of Long Running SQL Queries," in *SIGMOD*, 2004, pp. 803–814.

[18] D. Ślęzak, M. Kowalski, V. Eastwood, and J. Wróblewski, "Methods and Systems for Database Organization," US Patent 8,266,147 B2, 2012.

[19] D. Ślęzak and V. Eastwood, "Data Warehouse Technology by Infobright," in *SIGMOD*, 2009, pp. 841–846.

[20] M. Kowalski, D. Ślęzak, G. Toppin, and A. Wojna, "Injecting Domain Knowledge into RDBMS – Compression of Alphanumeric Data Attributes," in *ISMIS*, 2011, pp. 386–395.

[21] D. Ślęzak, P. Synak, J. Wróblewski, and G. Toppin, "Infobright Analytic Database Engine Using Rough Sets and Granular Computing," in *IEEE GrC*, 2010, pp. 432–437.

[22] P. Synak, "Rough Set Approach to Optimisation of Subquery Execution in Infobright Data Warehouse," in *SCKT (PRICAI Workshop)*, 2008.

[23] D. Ślęzak and M. Kowalski, "Towards Approximate SQL – Infobright's Approach," in *RSCTC*, 2010, pp. 630–639.

[24] W. Ziarko, "Probabilistic Approach to Rough Sets," *Int. J. Approx. Reasoning*, vol. 49, no. 2, pp. 272–284, 2008.

[25] M. Elhemali, C. A. Galindo-Legaria, T. Grabs, and M. Joshi, "Execution Strategies for SQL Subqueries," in *SIGMOD*, 2007, pp. 993–1004.

[26] D. Ślęzak, K. Stencel, and H. S. Nguyen, "(No)SQL Platform for Scalable Semantic Processing of Fast Growing Document Repositories," *ERCIM News*, vol. 2012, no. 90, 2012.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1463–1469

Performance Antipatterns of One To Many Association in Hibernate

Patrycja Węgrzynowicz
Institute of Informatics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
Email: patrycja@mimuw.edu.pl

***Abstract*—Hibernate is the most popular ORM framework for Java. It is a straightforward and easy-to-use implementation of Java Persistence API. However, its simplicity of usage often becomes mischievous to developers and leads to serious performance issues in Hibernate-based applications. This paper presents five performance antipatterns related to the usage of one-to-many associations in Hibernate. These antipatterns focus on the problems of the owning side of collections, the Java types and annotations used in mappings, as well as processing of collections. Each antipattern consists of the description of a problem along with a sample code, negative performance consequences, and the recommended solution. Performance is analyzed in terms of the number and complexity of issued database statement. The code samples illustrate how the antipatterns decrease performance and how to implement the mappings to speed up the execution times.**

I. Introduction

HIBERNATE [1] is a remarkably popular implementation of Java Persistence API, i.e., the official standard of object-relational mapping in Java. However, many developers, especially of enterprise systems, complain on the correctness and the performance of Hibernate-based applications. It turns out that even relatively simple Hibernate applications impose a significant overhead on communication with a database, producing too many and/or inefficient SQL statements.

In this paper, we analyze the subject of the most common mapping, namely one-to-many associations of entities. Because of its wide usage as well as popularity of Hibernate, it seemed that the implementation of this association should have been well optimized. Our research revealed that even simple applications of one-to-many associations can result in (1) unexpected SQL statements being executed, (2) too many SQL statement being executed, and/or (3) too many objects being loaded into memory. Therefore, naïve mapping of one-to-many associations can introduce a significant performance overhead in a Hibernate-based application.

The main contribution of this paper consists of five performance antipatterns (i.e., bad practices with a significant impact on performance) related to the usage of one-to-many associations in Hibernate. Each antipattern consists of the description of a problem, performance consequences, the recommended solution, and a sample code.

II. Related Work

Antipatterns are conceptually similar to design patterns as they describe recurring problems and provide solutions to them. A performance antipattern describes a bad practice that has a significant impact on performance. The articles [2] and [3] provide a good explanation of performance antipatterns along with the definition of 14 performance antipatterns. These papers have formed a good basis for further research on performance antipatterns, mainly in the field of automated detection and fixing of performance problems. The paper [4] introduces a framework for automated detection and assessment of performance antipatterns in component based systems. The authors of [5] focus on detection of performance antipatterns in architectural models. The article [6] explains the method to remove performance antipatterns from software by analyzing UML models. The paper [7] describes modelling and analysis of software performance antipatterns along with their constraints and solvability using different models.

As mentioned above there has been work done in the field of performance antipatterns, defining solid means in terms of their definitions, detection, and fixing. These efforts focus on generic and domain-independent antipatterns, mostly in the application layer. However, they do not touch the data layer. Numerous performance problems have their source in an inefficient way of retrieving or storing large amount of data in a database. Therefore, this area is important to correctly identify the source of performance issues in software. It is especially true nowadays, when the amount of data processed continually increases and the use of automated frameworks for object-relational mapping becomes more and more common.

There is a number of popular ORM libraries, like Hibernate [1], EclipseLink [8], Open JPA [9], and Data Nucleus [10] to name a few. Even though they implement the same specification — Java Persistence API, they differ in mapping policies, generate different schemata and SQL statements that have different performance characteristics. The article [11] analyzes the influence of optimizations on the performance of Hibernate. In another paper [12], the same authors compare the performance of an object-relational mapping tool (Hibernate) vs. an object-oriented database (db4o) using OO7 benchmark. Another comparative study of the performance of object-relational mapping tools for .NET platform can be found

978-83-60810-53-8/$25.00

in [13]. However, these comparative studies focus mainly on queries, which usually are as efficient as SQL queries since they are direct translation to SQL. Some authors (e.g., [14]) identify the need to improve the efficiency of ORM tools by utilizing the features provided by the database engines.

Even though there has been work done on comparing performance of ORM tools, we still lack a systematic approach to identification of their strengths and weaknesses in terms of impedance mismatch [15]. In this paper, we aim at the identification of common performance problems for one-to-many associations in Hibernate. We also define five new antipatterns and provide recommendations how to fix them.

III. Antipattern: Inadequate Collection Type on Owning Side

A. Description

In JPA, *@OneToMany* annotation used on the owning side (i.e., the side used to manage persistency of elements) is one of the most common implementations of a one-to-many association between persistent entities. Such an approach is commonly used in enterprise development mainly due its simplicity and little coding overhead. However, it can introduce a serious performance overhead in Hibernate when combined with an inadequate Java collection type.

Table I presents three types of semantics available in Hibernate, dependent on the combination of a Java collection type and JPA annotations. Each of these semantics has a different performance characteristic. Table II shows the numbers of statements issued while persisting a collection of a given semantics after an addition or removal of a single element.

The bag semantics has the worst performance when it comes to the number of operations since it always re-creates the entire collection. Hibernate issues a delete statement to remove all associations of the old collection from the association table. Then, it issues N inserts to add all associations representing the new collection to the association table. Hibernate does not analyze how many elements have been changed in the collection.

In the list semantics, an addition or removal of a single element from a collection of N elements results in a single insert or delete respectively and M updates. The updates are needed to correct the indices of M elements. In case of addition, Hibernate needs to correct the indices of the elements before the one being added. In case of removal, Hibernate needs to correct the indices of the elements after the one being removed.

The set semantics seems to be the most efficient. For a single operation on a collection, it requires only a single database operation. However, it is worthwhile remembering that the Java set semantics requires a uniqueness check on the elements of a set. It implies that, in case of any additions of new elements to a persistent set, all elements of the set must be loaded into the main memory.

The antipattern relates to the usage of an inefficient collection semantics for a given usage pattern of a collection:

- For usage patterns of collections where in a single transaction the collection is usually left unmodified or only a few elements are added or removed, the usage of the bag semantics (i.e., *java.util.Collection* or *java.util.List* without the index or order annotations) is notably inefficient.
- For usage patterns of collections where in a single transaction most of the elements of the collection are removed, the usage of the set or list semantics (i.e., *java.util.Set* or *java.util.List* with the index or order annotation) is inefficient.

TABLE I
THREE SEMANTICS FOR COLLECTIONS WITH @ONETOMANY ANNOTATION IN HIBERNATE.

Semantics	Java Type	Annotation
Bag semantics	java.util.Collection java.util.List	@OneToMany
List semantics	java.util.List	@OneToMany $\wedge$ (@IndexColumn $\vee$ @OrderColumn)
Set semantics	java.util.Set	@OneToMany

TABLE II
THE NUMBERS OF DML STATEMENTS ISSUED WHILE PERSISTING A COLLECTION OF N ELEMENTS WITH A GIVEN SEMANTICS (DEFAULT TABLE MAPPING WITH AN ASSOCIATION TABLE; NO CASCADE OPTION).

Semantics	One Element Added	One Element Removed
Bag semantic	1 delete, N inserts	1 delete, N inserts
List semantic	1 insert, M updates	1 insert, M updates
Set semantic	1 insert	1 delete

B. Consequences

The performance consequences of the usage of an inadequate collection type on the owning side of a one-to-many association include an increased workload on the database engine because:

- For the bag semantics, Hibernate re-creates an entire collection, performing one delete to clear the collection and as many inserts as there are elements in the collection. For the common usage pattern of collections, where only a few elements are added or removed, such a re-creation results in suboptimal performance due to the operations on data which actually have not been changed. The bigger the collection is, the performance overhead is more significant.
- For the list or set semantics, Hibernate performs single delete or insert per each removal or addition respectively. For the usage pattern of collections, where most elements of a collection are removed, such a strategy results in suboptimal performance due to many deletes instead of one delete to clear the collection in one operation.

C. Solution

The solution to this antipattern is to analyze the usage profile of collections in an application and adjust the type of Java collections accordingly:

- If a collection is constant or is the subject to minimal changes or is relatively small, the recommended collection type is *java.util.Set* with the set semantics.
- If a collection is heavily modified, the recommended collection type is *java.util.Collection* or *java.util.List* with the bag semantics.

D. Sample Code

Listing 1 presents an example of the antipattern related to an inadequate collection type on the owning side. The sample code has two persistent entities (`Forest` and `Tree`), which are connected by an unidirectional association (`Forest` has a collection of `Tree` objects). The classes meet all requirements imposed by JPA and Hibernate on persistent entities (e.g., no-arg constructor). We use a minimal set of additional annotation and configuration parameters, instead relying on default values.

To test the persistency mechanism implemented in Hibernate, we execute two transactions. To ensure the same runtime environment (e.g., clear caches), each transaction is executed with a new `EntityManager` instance. The first transaction creates a `Forest` instance and 10000 `Tree` instances planted in the newly created `Forest`, whereas the second transaction finds the previously created `Forest` instance, creates a new `Tree` instance and plants it in the found `Forest`. It turns out that Hibernate for such a piece of code as in the second transaction re-creates the entire collection (i.e., executes one delete and 10 001 inserts). This behavior of Hibernate is not performance-wise since in our example one insert would be enough to synchronize the state of the object in the main memory with the state of the records in the database. For large collections with only a few changes it imposes a significant performance overhead.

Listing 1. The example of an inadequate collection type on the owning side.

```
@Entity
public class Forest {
    @Id @GeneratedValue
    private Long id;
    @OneToMany
    Collection<Tree> trees =
        new HashSet<Tree>();
    ...
    public void plantTree(Tree tree) {
        trees.add(tree);
    }
}
@Entity
public class Tree {
    @Id @GeneratedValue
    private Long id;
    private String name;
    ...
}
// Transaction 1
// creates and persists a forest...
// ... with 10.000 trees
...
// Transaction 2
Tree tree = new Tree("oak");
em.persist(tree);
Forest forest = em.find(Forest.class, id);
forest.plantTree(tree);
```

IV. Antipattern: OneToMany as Owning Side

A. Description

The antipattern relates to the usage of the collection side (i.e., *@OneToMany*) as the owning side of an association for large collections, especially the ones which expect only a few changes in a single transaction.

According to Section III, for such a use case we should use *java.util.Set* to minimize the performance overhead. However, even usage of *java.util.Set* does not guarantee the optimal performance.

First, due to the Java set semantics the entire collection needs to be loaded into the main memory in order to enforce a uniqueness check in case of addition of elements to the collection. It results in an additional database query (or queries due to batch loading) issued and additional processing dedicated to the transformation of each returned row into an object. These additional queries and processing happen even if the application logic does not access the elements of the collection in question.

Second, there might be an overhead connected with transactions and locking, when the entity containing the collection in question uses optimistic locking with versioning. In such cases, Hibernate locks not only the entity but also the collection, which lowers the capability of an application to serve concurrent requests.

B. Consequences

The performance consequences of the *@OneToMany* as the owning side antipattern are as follows:

- There is an increased workload on the database engine. Hibernate issues an additional database query (or a number of queries in case of batch loading) to retrieve the elements of a collection.
- There is an increased workload on the application server (CPU). Hibernate needs to convert each returned row into an object, even if the application logic does not access those objects.
- The memory footprint is significant, especially for large collections. It can result in slower performance due to insufficient memory available, leading to more frequent garbage collections or even page swaps.

- There may be a decreased throughput. Due to transaction and locking issues, the capability of an application to serve concurrent requests may be significantly lowered.

C. Solution

The solution to this antipattern is to manage a collection from the *@ManyToOne* side instead of *@OneToMany* side, i.e., to make *@ManyToOne* the owning side of the association. In case we are not able to change the owning side of an association, we should at least exclude such a collection from locking by using a Hibernate-specific annotation: *@OptimisticLock(excluded=true)*.

D. Sample Code

Listing 2 presents an example of *@OneToMany* as the owning side antipattern. It mimics Listing 1, introducing only a few changes: (1) the Java type of a collection has been changed to *java.util.Set* and (2) version fields have been added in `Forest` and `Tree` classes. The key piece of code is located in the second transaction, which adds a single `Tree` to the previously created `Forest` with 10 000 trees. It turns out that in the second transaction, the entire forest is loaded into memory. Moreover, we are not able to plant trees in parallel because Hibernate locks the entire `Forest` instance along with all its `Tree` associations.

V. Antipattern: Inadequate Collection Type on Inverse Side

A. Description

This antipattern relates to the usage of *java.util.Set* on the inverse side (also known as the mapped collection) of a bidirectional one-to-many association in Hibernate.

In case of bidirectional associations, it is recommended to synchronize the state of the objects (i.e., the mapped collection and the element entities with *@ManyToOne* pointers) in memory. The common implementation to ensure the consistence of the objects relies on an automated update of the mapped collection by the setter method in an element entity responsible for setting the *@ManyToOne* pointer (see Section V-D).

Therefore, when the type of the mapped collection is *java.util.Set*, it means that the entire collections needs to be loaded into the main memory in response to each addition of new elements, even though the application does not access neither the collection nor its elements. It happens due to the Java set semantics mentioned earlier, which requires a uniqueness check on the elements of a set.

B. Consequences

The performance consequences of the usage of *java.util.Set* on the inverse side of a bidirectional one-to-many association with in-memory state synchronization are as follows:

- There is an increased workload on the database engine. Hibernate issues an additional database query (or a number of queries in case of batch loading) to retrieve the elements of a collection.

Listing 2. The example of @OneToMany as the owning side.

```
@Entity
public class Forest {
    @Id @GeneratedValue
    private Long id;
    @Version
    private Integer version;
    @OneToMany
    Set<Tree> trees = new HashSet<Tree>();
    ...
    public void plantTree(Tree tree) {
        trees.add(tree);
    }
}
@Entity
public class Tree {
    @Id @GeneratedValue
    private Long id;
    @Version
    private Integer version;
    private String name;
    ...
}
// Transaction 1
// creates and persists a forest...
// ... with 10.000 trees
...
// Transaction 2
Tree tree = new Tree("oak");
em.persist(tree);
Forest forest = em.find(Forest.class, id);
forest.plantTree(tree);
```

- There is an increased workload on the application server (CPU). Hibernate needs to convert each returned row into an object, even if the application logic does not access those objects.
- For large collections, there is a significant memory footprint. It can slow down performance of an application due to insufficient memory available, more frequent garbage collections, or even page swaps.

C. Solution

The recommended Java types to be used on the inverse side of a one o many association are *java.util.Collection* or *java.util.List*. These types do not force loading the elements into memory until the elements are accessed by client code.

D. Sample Code

Listing 2 presents an example of *java.util.Set* used on the *@OneToMany* inverse side with in-memory state synchronization. The example continues the previously described `Forest` and `Tree` classes presented in Sections III and IV. Here, `Tree` is the owning side of an association, while `Forest`

Listing 3. The example of an inadequate collection type on the inverse side.

```
@Entity
public class Forest {
    @Id @GeneratedValue
    private Long id;
    @OneToMany (mappedBy = "forest")
    Set<Tree> trees = new HashSet<Tree>();
    ...
    public void plantTree(Tree tree) {
        trees.add(tree);
    }
}
@Entity
public class Tree {
    @Id @GeneratedValue
    private Long id;
    private String name;
    @ManyToOne
    Forest forest;
    ...
    public void setForest(Forest forest) {
        this.forest = forest;
        this.forest.plantTree(this);
    }
}
// Transaction 1
// creates and persists a forest...
// ... with 10.000 trees
...
// Transaction 2
Tree tree = new Tree("oak");
Forest forest = em.find(Forest.class, id);
tree.setForest(forest);
em.persist(tree);
```

is the inverse side. Therefore, to save changes in a database, we need to set a right `Forest` reference in a `Tree` instance. While setting the `Forest` instance in a `Tree`, the collection of `Trees` is automatically updated to include the new tree (the invocation of `plantTree`). Such a pattern is widely used and recommended to ensure up-to-date states of objects. However, it results in a significant performance overhead, when combined with *java.util.Set* on the inverse side. In the second transaction, which only adds a new tree and does not access other trees in the forest, Hibernate has to load the entire collection into the main memory.

VI. Antipattern: Lost Collection Proxy on Owning Side

A. Description

The antipattern relates to the assignment of a new collection object to a persistent field, representing the owning side of a one-to-many association. Thus, a collection proxy returned by Hibernate is lost.

In such a case, Hibernate is not able to track what has been changed in a collection and its policy is to re-create the entire collection regardless of the actual modifications. Therefore, even if the elements of the collection have not been changed, Hibernate issues a delete to remove the associations from the association table and then performs as many inserts as there are elements in the collection.

B. Consequences

The performance consequences of a lost proxy on the owning side are as follows:

- There is an increased workload on the database engine. There are unnecessary database operations performed. Hibernate re-creates an entire collection, performing one delete to clear the collection and as many inserts as there are elements in the collection. Such a re-creation results in suboptimal performance due to the operations on data which actually have not been changed.

C. Solution

The recommended solution to the antipattern is to operate on collection objects returned by Hibernate as in most cases it is a much more efficient approach. However, it might be performance-wise to re-create the entire collection in cases where most elements of the collection have been removed. On the other hand, it is one of the places where Hibernate could apply a smarter policy, especially as it has all required data available.

D. Sample Code

Listing 4 presents an example of a lost (according to Hibernate) collection proxy on the owning side. The example consists of two persistent entities: `Hydra` and `Head`. `Hydra` is a mythical creature that re-grows three heads in place of one head cut off. In order to model this feature, we need to provide strict encapsulation of heads. Therefore, `getHeads` returns an unmodifiable wrapper over the mutable collection of heads. In the first transaction, we create and persist a `Hydra` instance with three `Heads`. In the second transaction, we simply read the previously stored instance. It turns out that for this piece of code Hibernate executes two selects, one delete and three inserts, even though the code is purely read-only. The problem lies in the way Hibernate checks whether or not a property is dirty during the commit of a transaction. In order to check the dirtiness of a collection, Hibernate compares the references on the actual collection and the proxy originally loaded. Unfortunately, in our example Hibernate uses `getHeads` method to access the collection (due to property access mapping). The method returns an unmodifiable wrapper over the original proxy returned by Hibernate. Obviously, it returns a different Java object that the original one loaded by Hibernate. Thus, Hibernate decides that the collection has been changed and re-creates the entire collection.

Listing 5 presents a more straightforward example of a lost collection proxy on the owning side. Again we implement two

Listing 4. The example 1 of a lost collection proxy on the owning side.

```
@Entity
public class Hydra {
  private Long id;
  private List<Head> heads =
    new ArrayList<Head>();
  ...
  @Id @GeneratedValue
  public Long getId() {...}
  protected void setId() {...}
  @OneToMany(cascade=CascadeType.ALL)
  public List<Head> getHeads() {
    return Collections.
          unmodifiableList(heads);
  }
  protected void
      setHeads(List<Head> heads)
  {...}
}

// Transaction 1
// creates and persists the hydra...
// ...with 3 heads
...
// Transaction 2
Hydra found = em.find(Hydra.class, id);
```

Listing 5. The example 2 of a lost collection proxy on the owning side.

```
@Entity
public class Hydra {
  @Id @GeneratedValue
  private Long id;
  @OneToMany(cascade=CascadeType.ALL)
  private List<Head> heads =
    new ArrayList<Head>();
  ...
  public Long getId() {...}
  protected void setId() {...}
  public List<Head> getHeads() {
      return heads;
  }
  public void setHeads(List<Head> heads) {
    this.heads = heads;
  }
}

// Transaction 1
// creates and persists the hydra...
// ...with 3 heads
...
// Transaction 2
Hydra found = em.find(Hydra.class, id);
List<Head> currentHeads =
  new ArrayList<Head>(found.getHeads());
found.setHeads(currentHeads);
```

persistent classes: `Hydra` and `Head`. However, we do not introduce strict encapsulation. Instead we implement simple getters and setters for all fields. In the second transaction, we create a new collection containing the current heads of our `Hydra` instance. In terms of our business model, nothing has changed — the heads are the same heads as originally loaded. However, Hibernate observes a different collection reference and applies the policy of re-creation of the entire collection.

VII. Antipattern: One-By-One Processing of Collection

A. Description

The antipattern refers to a sequential processing of a persistent collection, i.e., a piece of code iterates over the collection and for each element in a collection, it may perform a database operation.

Best practices of database programming highlight the need to operate on the sets of records instead of single records. While SQL provides the means to such a paradigm switch in programming, Java is an object-oriented language without support to relational algebra. Therefore, it is a common approach in the Java world to iterate over persistent collections and process their elements one-by-one. Such an approach often leads to a significant performance overhead, considering the number of database round-trips and the volume of data passed between a database and an application.

B. Consequences

The performance consequences of one-by-one processing of a persistent collection are as follows:

- A high number of database operations is executed. It is proportional to the size of the collection.
- RDBMS engine is used ineffectively.
- Network latency can sum up to a significant performance overhead.

C. Solution

The solution to this antipattern is to utilize the capabilities of a relational database by the usage of bulk statements and aggregate functions. Frequently it requires a different object model.

D. Sample Code

Listing 6 presents an example of a one-by-one processing of a collection. The example continues the previous examples consisting of `Forest` and `Tree`. Here, in the second transaction we want to delete the entire `Forest`. A simple remove causes a constraint violation exception since there are trees associated with the forest to be removed. Therefore, we need to unbind the trees first. Unfortunately, in Hibernate there is no other way to do this than setting the `Forest` reference in each `Tree` instance to `null`. In our example it results in

Listing 6. The example of a one by one processing of a collection.

```
@Entity
public class Forest {
  @Id @GeneratedValue
  private Long id;
  @OneToMany (mappedBy = "forest")
  Set<Tree> trees = new HashSet<Tree>();
  ...
  public void plantTree(Tree tree) {
    trees.add(tree);
  }
}
@Entity
public class Tree {
  @Id @GeneratedValue
  private Long id;
  private String name;
  @ManyToOne
  Forest forest;
  ...
}
// Transaction 1
// creates and persists a forest...
// ... with 10.000 trees
...
// Transaction 2
Tree tree = new Tree("oak");
Forest forest = em.find(Forest.class, id);
for (Tree tree : forest.getTrees()) {
  tree.setForest(null);
}
em.remove(forest);
```

10 000 updates. To fix this inefficiency in Hibernate, we need to change the object model and introduce an explicit class representing the association.

VIII. Conclusion

In this paper, we presented five performance antipatterns related to one-to-many associations in Hibernate. Each antipattern consists of the description of a problem, performance consequences and the recommended solution, as well as a sample code to better illustrate the problem. The identified antipatterns introduce a significant performance overhead in terms of the number of SQL statements executed as well as the number of objects loaded into the main memory. These are two critical factors that have serious impact on the performance of applications. The number of SQL statements executed directly increases the load on the database engine. Usually it also introduces an additional performance overhead due to the network latency which is important in modern multi-tiered applications, where applications and databases are located in different servers/tiers. High memory consumption has indirect impact on the performance as it usually leads to more frequent garbage collection or even page swaps.

The presented antipatterns show that the usage of Hibernate is not as simple as it looks at first glance. Even plain use cases can significantly decrease the performance of an application. The antipatterns explain how to use Hibernate efficiently and what policies should be improved in Hibernate in order to shorten the execution time.

References

[1] Hibernate. [Online]. Available: http://www.hibernate.org

[2] C. U. Smith and L. G. Williams, "Software performance antipatterns; common performance problems and their solutions," in *Int. CMG Conference*, 2001, pp. 797–806.

[3] ——, "New software performance antipatterns: More ways to shoot yourself in the foot," in *Int. CMG Conference*. Computer Measurement Group, 2002, pp. 667–674.

[4] T. Parsons and J. Murphy, "A framework for automatically detecting and assessing performance antipatterns in component based systems using run-time analysis," in *The 9th International Workshop on Component Oriented Programming, part of ECOOP*, 2004.

[5] C. Trubiani and A. Koziolek, "Detection and solution of software performance antipatterns in palladio architectural models," in *Proceedings of the 2nd ACM/SPEC International Conference on Performance engineering*, ser. ICPE '11. New York, NY, USA: ACM, 2011, pp. 19–30. [Online]. Available: http://doi.acm.org/10.1145/1958746.1958755

[6] V. Cortellessa, A. Di Marco, R. Eramo, A. Pierantonio, and C. Trubiani, "Digging into uml models to remove performance antipatterns," in *Proceedings of the 2010 ICSE Workshop on Quantitative Stochastic Models in the Verification and Design of Software Systems*, ser. QUOVADIS '10. New York, NY, USA: ACM, 2010, pp. 9–16. [Online]. Available: http://doi.acm.org/10.1145/1808877.1808880

[7] V. Cortellessa, A. Di Marco, and C. Trubiani, "Software performance antipatterns: modeling and analysis," in *Proceedings of the 12th international conference on Formal Methods for the Design of Computer, Communication, and Software Systems: formal methods for model-driven engineering*, ser. SFM'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 290–335. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-30982-3_9

[8] Eclipselink. [Online]. Available: http://www.eclipse.org/eclipselink/

[9] Openjpa. [Online]. Available: http://openjpa.apache.org/

[10] Datanucleus. [Online]. Available: http://www.datanucleus.org/

[11] P. van Zyl, D. G. Kourie, L. Coetzee, and A. Boake, "The influence of optimisations on the performance of an object relational mapping tool," in *Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, ser. SAICSIT '09. New York, NY, USA: ACM, 2009, pp. 150–159. [Online]. Available: http://doi.acm.org/10.1145/1632149.1632169

[12] P. van Zyl, D. G. Kourie, and A. Boake, "Comparing the performance of object databases and orm tools," in *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, ser. SAICSIT '06. Republic of South Africa: South African Institute for Computer Scientists and Information Technologists, 2006, pp. 1–11. [Online]. Available: http://dx.doi.org/10.1145/1216262.1216263

[13] S. Cvetković and D. Janković, "A comparative study of the features and performance of orm tools in a .net environment," in *Proceedings of the Third international conference on Objects and databases*, ser. ICOODB'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 147–158. [Online]. Available: http://dl.acm.org/citation.cfm?id=1926241.1926257

[14] A. Szumowska, M. Burza$#324;ska, P. Wi$#347;niewski, and K. Stencel, "Efficient implementation of recursive queries in major object relational mapping systems," in *Proceedings of the Third international conference on Future Generation Information Technology*, ser. FGIT'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 78–89. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27142-7_10

[15] P. Wiśniewski, M. Burzańska, and K. Stencel, "The impedance mismatch in light of the unified state model," *Fundam. Inform.*, vol. 120, no. 3-4, pp. 359–374, 2012.

4th Workshop on Advances in Programming Languages

PROGRAMMING languages are programmers' most basic tools. With appropriate programming languages one can drastically reduce the cost of building new applications as well as maintaining existing ones. In the last decades there have been many advances in programming languages technology in traditional programming paradigms such as functional, logic, and object-oriented programming, as well as the development of new paradigms such as aspect-oriented programming. The main driving force was and will be to better express programmers' ideas. Therefore, research in programming languages is an endless activity and the core of computer science. New language features, new programming paradigms, and better compile-time and run-time mechanisms can be foreseen in the future.

The aims of this event is to provide a forum for exchange of ideas and experience in topics concerned with programming languages and systems. Original papers and implementation reports are invited in all areas of programming languages.

Topics

Major topics of interest include but are not limited to the following:

- Automata theory and applications
- Compiling techniques
- Domain-specific languages
- Formal semantics and syntax
- Generative and generic programming
- Grammarware and grammar based systems
- Knowledge engineering languages, integration of knowledge engineering and software engineering
- Languages and tools for trustworthy computing
- Language theory and applications
- Language concepts, design and implementation
- Markup languages (XML)
- Metamodeling and modeling languages
- Model-driven engineering languages and systems
- Practical experiences with programming languages
- Program analysis, optimization and verification
- Program generation and transformation
- Programming paradigms (aspect-oriented, functional, logic, object-oriented, etc.)
- Programming tools and environments
- Proof theory for programs
- Specification languages
- Type systems
- Virtual machines and just-in-time compilation
- Visual programming languages

Steering Committee

Lukovic, Ivan, University of Novi Sad, Serbia, Serbia
Mernik, Marjan, University of Maribor, Slovenia
Slivnik, Bostjan, University of Ljubljana, Slovenia

Event Chair

Janousek, Jan, Czech Technical University, Czech Republic

Program Committee

Aycock, John, University of Calgary, Canada
Chen, Haiming, Chinese Academy of Sciences, China
Henriques, Pedro Rangel, Universidade do Minho, Portugal
Horvath, Zoltan, Eotvos Lorand University, Hungary
Ivanovic, Mirjana, University of Novi Sad, Serbia
Kardas, Geylani, Ege University International Computer Institute, Turkey
Kollar, Jan, Technical University of Kosice, Slovakia
Kosar, Tomaz, University of Maribor, Slovenia
Liu, Shih-Hsi "Alex", California State University, United States
Lukovic, Ivan, University of Novi Sad, Serbia, Serbia
Mandreoli, Federica, University of Modena, Italy
Martínez López, Pablo E. "Fidel", Universidad Nacional de Quilmes, Argentina
Mernik, Marjan, University of Maribor, Slovenia
Milasinovic, Boris, University of Zagreb Faculty of Electrical Engineering and Computing, Croatia
Moessenboeck, Hanspeter, Johannes Kepler Universitat Linz, Austria
Papaspyrou, Nikolaos, National Technical University of Athens, Greece
Pereira, Maria Joao Varanda, Instituto Politecnico de Braganca, Portugal
Poruban, Jaroslav, Technical University of Kosice, Slovakia
Rodriguez, Jose Luis Sierra, Universidad Complutense de Madrid, Spain
Slivnik, Bostjan, University of Ljubljana, Slovenia
Splawski, Zdzislaw, Wroclaw University of Technology, Poland
Watson, Bruce, Stellenbosch University, South Africa

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1473–1476

Magnify – a new tool for software visualization

Cezary Bartoszuk, Grzegorz Timoszuk, Robert Dąbrowski, Krzysztof Stencel
Institute of Informatics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland

***Abstract*—Modern software systems are inherently complex. Their maintenance is hardly possible without precise up-to-date documentation. It is often tricky to document dependencies among software components by only looking at the raw source code. We address these issues by researching new software analysis and visualization tools.**

In this paper we focus on software visualisation. *Magnify* is our new tool that performs static analysis and visualization of software. It parses the source code, identifies dependencies between code units and records all the collected information in a repository based on a language-independent graph-based data model. Nodes of the graph correspond to program entities of disparate granularity: methods, classes, packages etc. Edges represent dependencies and hierarchical structure. We use colours to reflect the quality, sizes to display the importance of artefacts, density of connections to portray the coupling. This kind of visualization gives bird's-eye view of the source code. It is always up to date, since the tool generates it automatically from the current revision of software. In this paper we discuss the design of the tool and present visualizations of sample open-source Java projects of various sizes.

I. Introduction

THE complexity of software systems and their development processes have rapidly grown in recent years. In numerous companies the high level structure of dependencies between software components is kept only in the heads of developers. This makes the development process fragile, since teams composed of humans are inherently volatile. Therefore, development teams need methods and tools to inspect *current* states of complex systems and their fragments.

In this paper we describe a software visualization tool *Magnify* that caters for these needs. It shows top level views of software entities of disparate granularities: systems, components, modules, classes etc. These views contain graphic presentation of the importance, the quality and the coupling of subcomponents.

Given a source code bundle *Magnify* parses and analyzes it in order to create a graph-based model [1]. This model is then persisted in the architecture warehouse that provides data for software intelligence [2]. One of its methods is visualization. *Magnify* reads the data from the warehouse and produces a graph laid out on a plain. Nodes of the graph are entities (classes, packages etc.) of the provided source code.

The size of a node reflects the importance of the corresponding artefact. In the current version of *Magnify* we use PageRank to assess importance. The colour of a node represents the quality of the corresponding piece of code. In the examples presented in this paper, we use the numbers of lines per class. Any software metrics accompanied with threshold values for green and red can be employed.

Edges represent relationships of two kinds: structural inclusion and dependency.

The paper is organised as follows. In Section II we address the related work. In Section III we recall the theoretical foundations for our research. In Section IV we describe the design of *Magnify* and possible extension points in its architecture. In Section V we present visualizations of sample open-source Java projects of various sizes. Section VI concludes.

II. Motivation

The idea described in this paper has been contributed to by several existing approaches and practices.

A unified approach to software systems and software processes has already been presented in [3]. Software systems were perceived as large, complex and intangible objects developed without a suitably visible, detailed and formal descriptions of how to proceed. It was suggested that software process should be included in software project as parts of programs with explicitly stated descriptions; software architect should communicate with developers, customers and other managers through software process programs indicating steps that are to be taken in order to achieve product development or evolution goals.

Multiple graph-based models have been proposed to reflect architectural facets, e.g. to represent architectural decisions and changes [4], to discover implicit knowledge from architecture change logs [5] or support architecture analysis and tracing [6]. Graph-based models have also become helpful in UML model transformations, especially in model driven development (MDD) [7].

Visualization of software architecture has been a research goal for years. The tools like Bauhaus [8], Source Viewer 3D [9], Gevol [10], JIVE [11], evolution radar [12], code_smarm [13] and StarGate [14] are interesting attempts in visualization.

However none of them simultaneously supports aggregation (e.g. package views), drill-down, picturing the code quality and dependencies. Moreover, all of them are significantly more complex when compared to our proposal. In our opinion noteworthy better effects can be achieved with simpler facilities. Movies and the third dimension are not necessary to quickly assess the quality, robustness and resilience of an architecture.

978-83-60810-53-8/$25.00

III. Theoretical foundations

A. Model

We recall the theoretical model [1] for unified representation of architectural knowledge. Definition of the model is based on directed labelled multigraph. According to the model, the *software architecture graph* is an ordered triple $(\mathcal{V}, \mathcal{L}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices that reflect all artefacts created during a software project, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{L} \times \mathcal{V}$ is the set of directed edges that represent dependencies (relations) among those artefacts, and $\mathcal{L}$ is the set of labels which qualify the artefacts and their dependencies.

Example 1. Each artefact can be described by a set of labels. A method can be described by labels showing that it is a part of project source code (*code*); written in Java (*java*); its revision is 456 (*r:456*); it is *abstract* and *public*. Edges are directed and may have multiple labels as well, e.g.: a package *contains* a class; a method *calls* another method.

The transformations and metrics recalled below give the foundation for the layer of *software intelligence* tools [2].

B. Transformations

Our graph model is general and scalable, fits both small and huge projects [15], and has been tested in practice [16].

The tests proved that in case of a large project its graph model is too complex to be human-tractable as a whole. This has confirmed that transformations and views of the graph model are a must.

Example 2. For a given software graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$ and a subset of its labels $\mathcal{L}' \subseteq \mathcal{L}$, its *filter* is a transformation $\mathcal{G}|_{\mathcal{L}'} = (\mathcal{V}', \mathcal{E}', \mathcal{L}')$ where $\mathcal{V}'$ and $\mathcal{E}'$ have a label in $\mathcal{L}'$.

C. Metrics

For complex projects their quantitative evaluation is a must. The graph-based approach is in line with best practices for metrics [17], [18], allows for easy translation of existing metrics into graph terms [19], ensures they can be efficiently calculated using graph algorithms. It also allows designing new metrics that combine both software system and software process artefacts [20].

Example 3. For a given software graph $\mathcal{G} = (\mathcal{V}, \mathcal{L}, \mathcal{E})$, its *metric* is a transformation $m : \mathcal{G} \mapsto \mathcal{R}$ where $\mathcal{R}$ denotes real numbers and m can be effectively calculated by a graph algorithm on G.

IV. Magnify

We implemented *Magnify* as a server system, based on a graph database, with web front-end written in Scala. *Magnify* functionality allows loading source code bundles, analyzing them and displaying the resulting graphs of software components. Figure 1 shows a sample of *Magnify*'s GUI.

The analytical backend is composed of three main modules: the parser, the graph storage and the analysis engine.

The parser is the only part of *Magnify* that must be programming language specific. Its responsibility is to transform

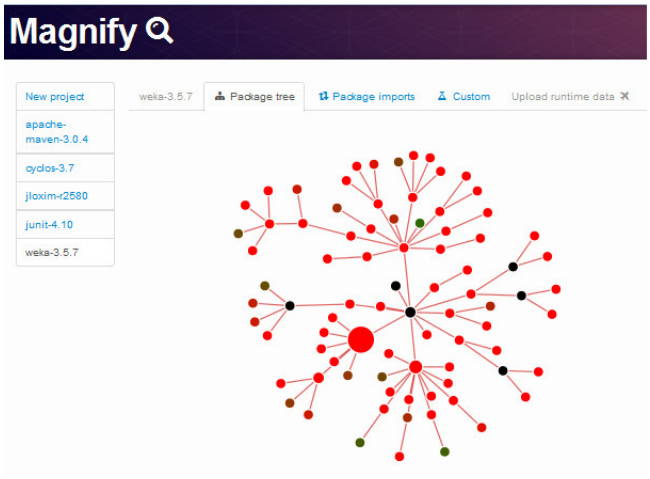


Fig. 1. *Magnify* functionality - main view

contents of a source code bundle into an abstract software graph. This graph is then persisted. Currently the implementation contains a Java 5 parser based on the `javaparser` library. The graph storage is based on Tinkerpop Blueprints specification. When the source code is converted and stored in the abstract language-agnostic form, the analysis engine will run. The implementation computes PageRank of every node in the graph and code quality metrics for classes. These metrics are then escalated to the package level. Figure 2 shows the referential deployment diagram of *Magnify*.

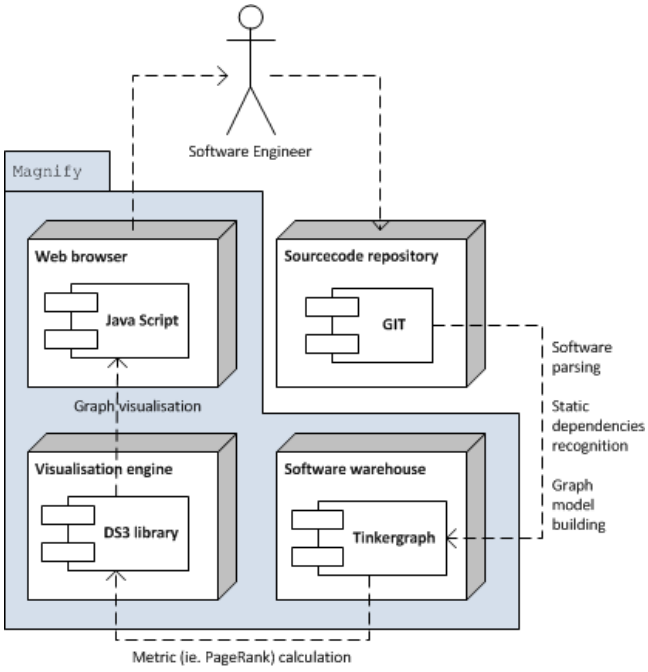


Fig. 2. General architecture of *Magnify*

The architecture of *Magnify* is flexible enough to replace any of its components and add new constituents. As noted

above, we can replace the repository with any graph database that conforms to Blueprints API. New data providers can be added, like parsers for more programming languages, runtime profilers, analyzers of version control systems, and analyzers of web data (e.g. forum discussions).

V. Evaluation

In this Section we show sample visualizations produced by *Magnify*. We have chosen five open-source projects that significantly vary in size and quality. All of them have a noteworthy number of users. They are well adopted by the software development community.

These are Cyclos, Play, Spring and Karaf. For each system we present its top-level visualization created by *Magnify*. Then, we analyze the resulting images and enumerate conclusions that can be drawn from them.

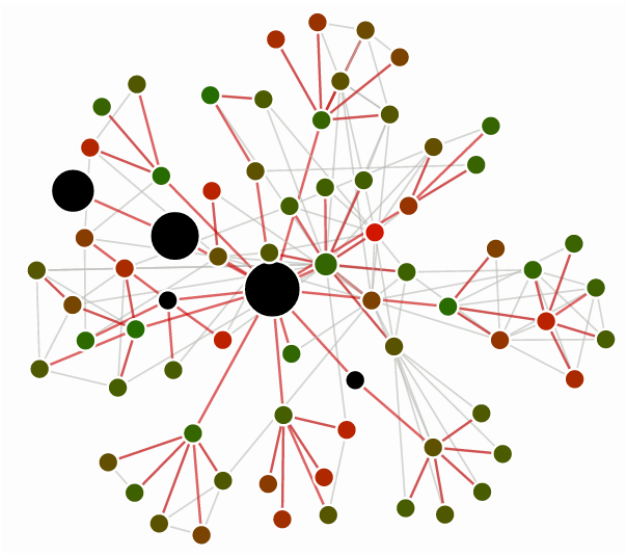


Fig. 3. The visualization of Spring context 3.2.2 produced by *Magnify*

A. Spring context 3.2.2

Spring is one of the most popular enterprise application frameworks in the Java community. It provides an infrastructure for dependency injection, cache, transactions, data base access and many more. Figure 3 shows the visualization of Spring produced by *Magnify*.

The structure of dependencies implies that Spring is well designed. The graph is notably sparse. The only packages detected as important are empty vendor packages. All the packages that do contain classes are of the same importance. This indicates a well balanced software. Figure 3 contains no brightly red packages. This means that on average the classes are small in most of packages. Thus, the overall quality is satisfactory.

B. Cyclos 3.7

Cyclos is a complete on-line payment system. Figure 4 presents visualization of this system produced by the *Magnify* tool.

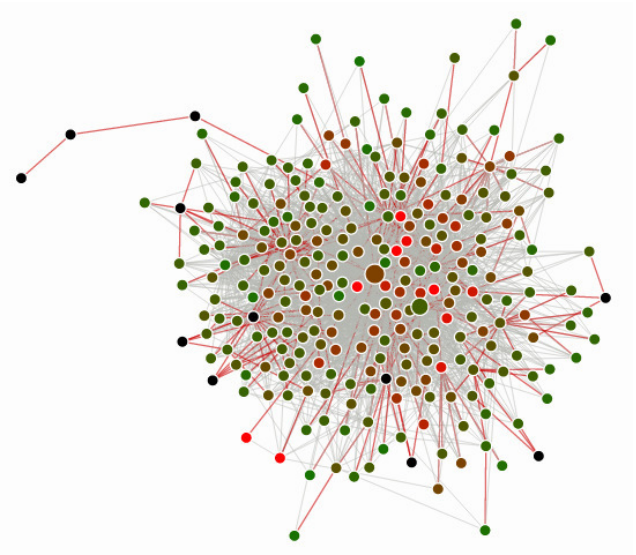


Fig. 4. The visualization of Cyclos 3.7 produced by *Magnify*

Unfortunately, this time the dependency graph is exceptionally dense. The software engineering experience indicates that the development and maintenance of software systems with so tight coupling is difficult, costly and error-prone. On the other hand, Figure 4 shows few packages in which classes are big on average. That means that overall complexity of the classes themselves is acceptable.

Cyclos is a profound example of a system that should be split into orchestrated group of communicating systems. This kind of refactoring will significantly improve the quality of this software. It will also reduce the cost of further development and maintenance.

C. Play 1.2.5

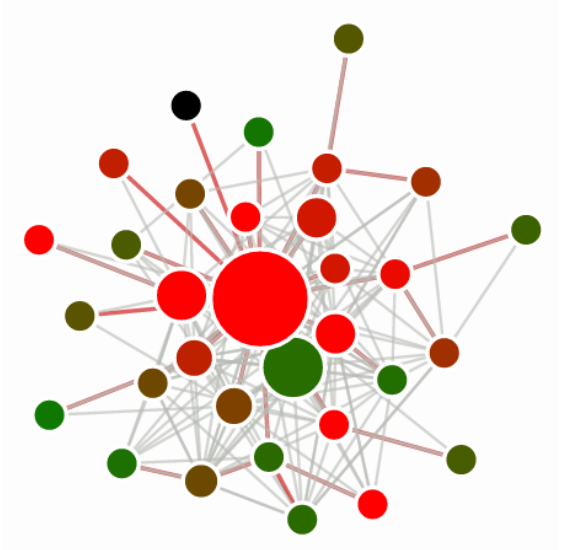


Fig. 5. The visualization of Play 1.2.5 produced by *Magnify*

Play is a popular Scala and Java web framework. Figure 5 shows the visualization of Play using *Magnify*.

It presents a small project with decent amount of dependencies. The flat package structure is typical for dynamic languages. The biggest node corresponds to the project root package `play`. Brightly red packages reveal potentially high complexity of their classes.

D. Apache Karaf 3.0.0 RC1

Apache Karaf is a small OSGi container to deploy various components and applications. Even though it is split into many packages, the number of dependencies is small. Many subtrees of the package hierarchy have only a single dependency on the rest of the system. Thus, Karaf is well packaged.

Figure 6 shows that overall code quality in Karaf is good. There are only a few packages where the average class size is alarming.

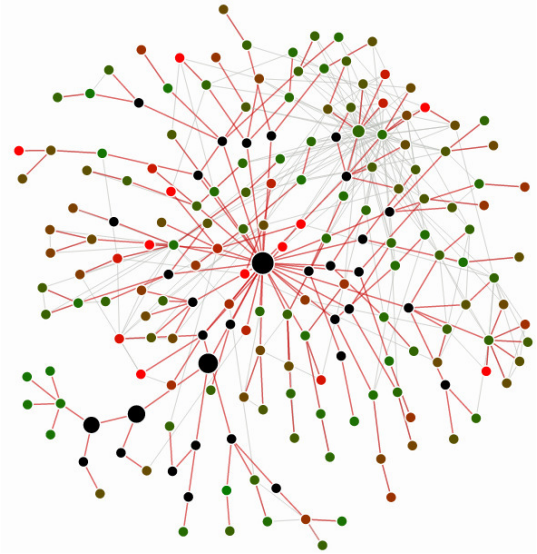


Fig. 6. The visualization of Karaf 3.0.0-RC1 produced by *Magnify*

VI. Conclusion

We follow the research on anayslis and visualisation of software and software process, and promote an approach that avoids separation between software and software process artefacts. We demonstrate that the implementation of such approach is feasible. We implement software intelligence on top of a software warehouse based on our theoretical graph-based model. We execute experiments on open-source Java programs using those tools.

In this paper we presented *Magnify* - a tool that performs static analysis and visualization of software systems. It focuses on relationships between components rather than on their internal structure.

Magnify is a general tool that can adapt other quality metrics and importance estimates. Flexibility of its design allows replacing any of its components and adding new parts. In order to support the analyses for another programming language, we have to add only an appropriate parser. All other facilities (the repository and analytic algorithms) need not be changed.

Promising ideas worth implementing in the near future include: (1) improving the vertex clustering algorithm for module repackaging, (2) gathering the information across revisions and (3) adding metadata stating how packages are split into modules and how these modules depend on each other. This kind of metadata would constitute a specification that can be matched against the source code.

References

[1] R. Dąbrowski, K. Stencel, and G. Timoszuk, "Software is a directed multigraph," in *ECSA*, ser. Lecture Notes in Computer Science, I. Crnkovic, V. Gruhn, and M. Book, Eds., vol. 6903. Springer, 2011, pp. 360–369.

[2] R. Dąbrowski, "On architecture warehouses and software intelligence," in *FGIT*, ser. Lecture Notes in Computer Science, T.-H. Kim, Y.-H. Lee, and W.-C. Fang, Eds., vol. 7709. Springer, 2012, pp. 251–262.

[3] L. J. Osterweil, "Software processes are software too," in *ICSE*, W. E. Riddle, R. M. Balzer, and K. Kishida, Eds. ACM Press, 1987, pp. 2–13.

[4] M. Wermelinger, A. Lopes, and J. L. Fiadeiro, "A graph based architectural (re)configuration language," in *ESEC / SIGSOFT FSE*, 2001, pp. 21–32.

[5] A. Tang, P. Liang, and H. van Vliet, "Software architecture documentation: The road ahead," in *WICSA*, 2011, pp. 252–255.

[6] H. P. Breivold, I. Crnkovic, and M. Larsson, "Software architecture evolution through evolvability analysis," *Journal of Systems and Software*, vol. 85, no. 11, pp. 2574–2592, 2012.

[7] J. Derrick and H. Wehrheim, "Model transformations across views," *Sci. Comput. Program.*, vol. 75, no. 3, pp. 192–210, 2010.

[8] R. Koschke, "Software visualization for reverse engineering," in *Software Visualization*, ser. Lecture Notes in Computer Science, S. Diehl, Ed., vol. 2269. Springer, 2001, pp. 138–150.

[9] J. I. Maletic, A. Marcus, and L. Feng, "Source viewer 3d (sv3d) - a framework for software visualization," in *ICSE*, L. A. Clarke, L. Dillon, and W. F. Tichy, Eds. IEEE Computer Society, 2003, pp. 812–813.

[10] C. S. Collberg, S. G. Kobourov, J. Nagra, J. Pitts, and K. Wampler, "A system for graph-based visualization of the evolution of software," in *SOFTVIS*, S. Diehl, J. T. Stasko, and S. N. Spencer, Eds. ACM, 2003, pp. 77–86, 212–213.

[11] S. P. Reiss, "Dynamic detection and visualization of software phases," *ACM SIGSOFT Software Engineering Notes*, vol. 30, no. 4, pp. 1–6, 2005.

[12] M. D'Ambros, M. Lanza, and M. Lungu, "The evolution radar: visualizing integrated logical coupling information," in *MSR*, S. Diehl, H. Gall, and A. E. Hassan, Eds. ACM, 2006, pp. 26–32.

[13] M. Ogawa and K.-L. Ma, "code_swarm: A design study in organic software visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1097–1104, 2009.

[14] K.-L. Ma, "Stargate: A unified, interactive visualization of software projects," in *PacificVis*. IEEE, 2008, pp. 191–198.

[15] P. Tabor and K. Stencel, "Stream execution of object queries," in *FGIT-GDC/CA*, 2010, pp. 167–176.

[16] R. Dąbrowski, K. Stencel, and G. Timoszuk, "Improving software quality by improving architecture management," in *CompSysTech*, 2012, pp. 208–215.

[17] F. Abreu and R. Carapuça, "Object-oriented software engineering: Measuring and controlling the development process," in *Proceedings of the 4th International Conference on Software Quality*, 1994.

[18] J. M. Roche, "Software metrics and measurement principles," *SIGSOFT Softw. Eng. Notes*, vol. 19, pp. 77–85, January 1994. [Online]. Available: http://doi.acm.org/10.1145/181610.181625

[19] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on Software Engineering*, vol. 20, pp. 476–493, June 1994. [Online]. Available: http://portal.acm.org/citation.cfm?id=630808.631131

[20] R. Dąbrowski, G. Timoszuk, and K. Stencel, "One graph to rule them all - software measurment and management," *Fundamenta Informaticae*, vol. to appear, 2013.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1477–1482

Conjunction, Sequence, and Interval Relations in Event Stream Processing

Samujjwal Bhandari
Department of Computer Science
Texas Tech University, Lubbock, U.S.A
samujjwal.bhandari@ttu.edu

Susan D. Urban
Department of Industrial Engineering
Texas Tech University, Lubbock, U.S.A
susan.urban@ttu.edu

***Abstract*—The conjunction operator can be augmented with temporal constraints to define an arbitrary pattern of events in event stream processing (ESP). However, using temporal constraints to specify patterns can be complex. This research has defined an operator hierarchy, where the top of the hierarchy defines the conjunction operator and the leaves of the hierarchy define more specific semantics associated with a sequence of events. The use of the specialized operators simplifies pattern expression and make the sequence semantics clear. Furthermore, in an experimental study, patterns using operators from the hierarchy outperform patterns expressed using the conjunction operator with temporal constraints in run time performance, further validating the usefulness of the operator hierarchy.**

***Keywords*-event operators; sequence; event processing language; operator semantics;**

I. Introduction

REAL world applications have become increasingly event-driven in nature, focusing on the occurrence or non-occurrence of several activities or their combinations to respond to a situation of interest using event processing systems such as [1], [2]. The situations of interest are encoded as complex event patterns using a specific ESP language, where complex event patterns are specified using event operators and other events. These complex event patterns are matched by the event processing system to detect complex events.

Encodings of event patterns should be able to define a situation in a unique manner. However, existing work [3], [4], [1], [2] defines patterns in an ambiguous way. For example, suppose in a health care application, situations of importance are detected if i) a high temperature is detected after nausea is detected and ii) a high temperature event is followed by a low temperature. Both *i* and *ii* can be defined as sequential occurrences of two events. In these situations, there are several possibilities that can be true of the patterns. For instance, in situation *i*, the nausea event occurs while there is a high temperature. However, in case of *ii*, a high temperature cannot occur at the same time as a low temperature. This example shows that the sequence pattern may have different interpretations. One of the possible solutions to this problem is to use the conjunction operator with relevant temporal constraints to restrict the detection of the event patterns. When an interval-based event is considered, this approach can specify all the possible patterns [5]. However, it is desirable to specify the intended semantics in an explicit way such as by using special operators. For example, rather than expressing a sequential pattern, E_1 followed by E_2 as $AND(E_1, E_2)$ **WHERE** $E_1.t_e < E_2.t_e$, where t_e represents an ending time of event occurrence, it would be intuitive to express the condition as $SEQ(E_1, E_2)$, where the **SEQ** operator explicitly defines the intended meaning.

To address the issues of specification complexity and ambiguous operator interpretation, this research defines an operator hierarchy based upon the conjunction and the sequence operators. The top of the hierarchy defines the conjunction operator. Moving down the hierarchy introduces specialized operators to express more specific situations. Though any pattern defined using the operators from the hierarchy can be expressed as a combination of the conjunction operator and appropriate temporal constraints, the use of specialized operators in defining event patterns makes the pattern specification an easier and more expressive task.

To verify the usefulness of the operators in the operator hierarchy, the operators have been implemented with reference to the conjunction operator implementation and are found to run better than their alternative versions using the conjunction operator with temporal constraints. Moreover, the work in this paper makes the following contributions:

1) Design of operators to incorporate different meanings for the sequence and the conjunction operators.
2) Design of an operator hierarchy defining the relationships pertaining to time intervals using Allen's relations [6].
3) Experimental evaluation of operators from the operator hierarchy to describe usefulness of the newly defined operator hierarchy.

II. Related Work

Past work on event processing, such as Snoop [3], Ode [7], and SAMOS [8] have collectively defined a powerful set of event operators to specify complex event patterns. However, operators such as the sequence and the repetition operators are not consistently defined in these languages. SEL [9] analyzes these event languages and identifies problems with the semantics of the negation, sequence, and repetition operators. The recent languages ([1], [2]) have adopted operators similar to past work on event processing, but have not considered the semantic inconsistency among the definition of event operators as discussed in [9]. Non-overlapping sequence defined in [10]

978-83-60810-53-8/$25.00

is considered to be an immediate sequence (i.e., an event A is immediately followed by B with no event in between), while sequence in [4] considers an arbitrary sequence (i.e., without restrictions on intervening events). These inconsistent definitions of the sequence operator as an overlapping and a non-overlapping sequence is relevant when an event is associated with an interval. The work in [5] defines a generic operator with a temporal constraint list to define all of the possible relations among intervals to remove inconsistency in the definition of event operators. However, the expression of event patterns becomes more complex.

III. Event Specification and Basic Concepts

For any two time intervals $i_1 = [t_1, t_2]$ and $i_2 = [t_3, t_4]$, there are thirteen possible relations defined as Allen's interval relations [6]. When an event e has event time $i = [t_s, t_e]$, e is said to have occurred over the interval i, where the event e started occurring at time point t_s and ended at time point t_e.

TABLE I: Situation Monitoring Event Definition

SON()	Stove is brought to ON state.
SOFF()	Stove is brought to OFF state.
LO()	A lid of a kettle is opened.
LC()	A lid of a kettle is closed.
KP()	A kettle is put on the surface.
IPOUR()	An item is put in the kettle.
T(v)	Regular event to provide temperature inside the kettle with the given value.
IP(items)	**TIMES**(IPOUR(), 3) **WITHIN** 2 MINUTES
KL()	**SEQ**(LO(), IP(items), LC()) **WHERE** IP.items.has({"water","milk","tea leaves"})
TP()	**AND**(SON(), KL(), KP(), T(v)) **WHERE** TEMP.v $\geq$ 100

Let us define a scenario to detect an activity defining the situation that the "tea has been prepared". Table I defines several events that are either observed or produced from the external environment (external events), or are a complex combination of other events (internal events). In Table I, the *T(v)* event is an external event that is generated by a thermometer. Other external events are *SON()*, *SOFF()*, *LO()*, *LC()*, *KP()*, and *IPOUR()*. An internal event is a composition of several external or other internal events. The *IP(items)* event, the *KL()*, and the *TP()* defined in Table I are internal events. The *IP(items)* event occurs when the *IPOUR()* event is detected three times within a 2 minutes window. The *KL()* event occurs with the sequential occurrence of the *LO()*, *IP(items)*, *LC()*, where *items* from the *IP(items)* event has water, milk and tea leaves in it. The *TP()* is represented as a pattern defining the occurrences of four events *SON()*, *KL()*, *KP()*, *T(v)*, where the temperature value is $\geq 100°\mathrm{C}$.

Discussion in later sections will consider events from the situation monitoring application (Table I) and the occurrences of events shown in Figure 1.

IV. Issues and Semantics of Conjunction and Sequence Operators

This section analyzes the semantics of event operators to identify the issues that must be addressed to define the

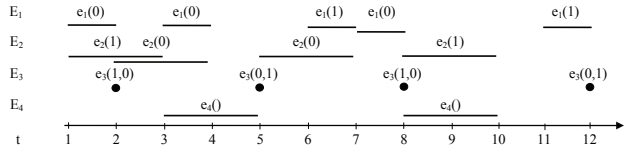


Fig. 1: Example Event Streams

semantics of operators in a clear and consistent way. Work such as that of [3], [4], [1] describes various powerful event constructs that can be categorized into conjunction, disjunction, repetition, negation, and sequence operators. Among different event operators, this work focuses on the semantics of the conjunction and sequence operators with interval-based temporal representation.

Conjunction of events E and F, denoted as *AND*(E, F), occurs when both E and F occur without temporal ordering restrictions. Detection of *AND*(E, F) starts when either E (or F) occurs and ends when F (or E) occurs. In case of interval-based semantics, all thirteen possible relations between interval are valid.

Example: The *TP()* event from Table I is a conjunction of five events *SON()*, *KL()*, *KP()*, *T(v)*, and *SOFF()*. Since the constituent events are combined using the **AND** operator, the *TP()* event occurs when all of the constituent events occur.

The complex event defined by a sequence operator has an implicit temporal constraint on events. A sequence of two events E and F, *SEQ*(E, F), detects the occurrence of an event E followed by the occurrence of the event F. Detection of *SEQ*(E, F) starts with the detection of an event E and ends with the detection of an event F. Such a requirement of event order by a sequence operator imposes temporal restrictions on event occurrences. When events are considered to be point-based, then *SEQ*(E, F) is detected if and only if $E.t < F.t$, where t is the time of event occurrence. If events are interval-based, then *SEQ*(E, F) is detected if and only if $E.[t_1, t_2] < E.[t_3, t_4]$, where $t_1 \leq t_2$ and $t_3 \leq t_4$. With these temporal conditions on event detection, we have two possible conditions: i) $t_2 < t_3$ and ii) $t_2 < t_4$. Though condition ii is included within condition i, the condition i gives the definition of a non-overlapping sequence operator, whereas the condition ii gives the definition of an overlapping sequence operator. In this section, the condition ii is used to define the sequence operation unless otherwise mentioned.

Example: The KL() event is a sequence of *LO()*, *IP(items)*, and *LC()*. The KL() event occurs when all of the constituent events occur with the constraint $LO.[t_s, t_e] < IP.[t_s, t_e] < LC.[t_s, t_e]$.

Using the definition of sequence, *SEQ*(E, F) says, E must occur before F. As discussed in this subsection, there are two possibilities for the sequence operator defining overlapping and non-overlapping sequences. Past work on event processing considers either an overlapping version of a sequence operator or a non-overlapping version. When an overlapping version of a sequence operator is used, then the sequence operator can be used to detect non-overlapping events, but it requires an

explicit temporal condition to specify that the non-overlapping sequence is intended. However, if a non-overlapping sequence is used, it cannot be used to specify an overlapping sequence. One of the solutions to this problem could be the use of the temporal filter on overlapping sequence. However, an application can demand specification of sequences of both kinds and, to make event specification more explicit, a separate operator for non-overlapping sequence may be suitable.

Examples: If only a non-overlapping version is defined, the sequence of events, such as *SEQ(SON(),SOFF())* is intuitively explicit as the stove on event precedes the stove off event. Let us encode the pattern specifying the situation that describes the condition where a kettle loading (*KL()*) process is followed by the detection of an item put (*KP()*) event. In this case, *KL()* can start before the *KP()* event and ends after the *KP()* event. This condition defines the sequence given as *SEQ*(*KP()*, *KL()*), which has the meaning of an overlapping sequence that cannot be encoded using the non-overlapping version.

V. Operator Design

This section addresses the semantic issues discussed in Section IV to design a set of event operators in a way such that each operator has a clear and consistent definition, and the operators are expressible in terms of its intended meaning. While discussing semantics of operators, only the binary operators are considered. One can extend the semantics of binary operators to their n-ary version using the binary semantics. Also, for readability, events are represented using its name only, instead of its schema.

A. Allen's Relations, Sequence, and Conjunction

Allen's 13 relations [6] define all of the possible relations between two intervals when both end points of intervals are fixed. When we have open end point relations, then the disjunction of Allen's relations to capture the desired relation can be complex. Also, the disjunction of Allen's relations or a hierarchical representation of composite relations [11] cannot express pair-wise relations among events due to non-transitivity of some operators such as overlaps [5]. Regardless of difficulty in specifying complex interval patterns, however, Allen's operators are concise constructs for capturing common interval relations. The use of these relations to express event patterns also defines the meaning of the pattern in an expressive manner. For example, the pattern defining the situation that an event E overlaps with an event F, $OVERLAPS(E, F)$ is easier to understand than *AND*(E, F) **WHERE** $E.t_s < F.t_s$ **and** $F.t_s < E.t_e$ **and** $E.t_e < F.t_e$. On the other hand, a pattern expressing the situation such as an event E ends before an event F is easier to understand as *AND*(E, F) **WHERE** $E.t_e < F.t_e$ than the encoding *OR*(*BEFORE*(E, F), $OVERLAPS(E, F)$, $MEETS(E, F)$, $STARTS(E, F)$, $DURING(E, F)$). The paragraphs that follow discuss the use of event operators and the use of temporal constraints in a suitable way to balance between understandability of expression using specific operators and the complexity of specifying the patterns.

Consider the sequence operator (**SEQ**) discussed in Section IV. To be consistent, consider that the semantics of the **SEQ** operator includes overlapping or non-overlapping occurrences of events ordered by end points. Then such a definition will include both interpretations of **SEQ** events from Section IV and this definition of **SEQ** corresponds to the definition of an overlapping sequence from Section IV. When an event pattern seeks only the overlapping sequence or only the non-overlapping sequence, then either a sequence operator with a required temporal restriction can be used to define the restricted sequence, or different operators defined for each condition can be used to specify a restricted sequence. For example, *SEQ*(E, F) **WHERE** $E.t_e < F.t_s$ is the same as Allen's *before* operator. The idea of using temporal constraints with operators or the definition of an equivalent operator defines the hierarchy of sequence operators with respect to Allen's operators. Figure 3 shows the hierarchy of a sequence operator with respect to two different forms of sequence operations along with the relation to Allen's operators. The hierarchy shown in Figure 3 depicts that the sequence operation combines *before*, *meets*, *overlaps*, *starts*, and *during* relations from Allen's relations. Section V-B further analyzes and discusses the sequence hierarchy to define operators.

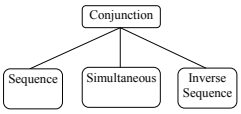


Fig. 2: Conjunction Hierarchy

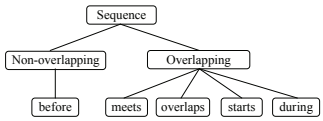


Fig. 3: Sequence Hierarchy

Conjunction (**AND**) is one of the well understood operations in event processing. The use of the conjunction operator does not define temporal restrictions on event occurrences, so the use of temporal constraints with **AND** can define every possible combination of interval relations. This idea of using temporal constraints with the conjunction operator is similar to the work done in [5], where an operator for an interval sequence *iseq*[1] is defined with the temporal constraints to define arbitrary relations on intervals. The sequence operator can be considered as a temporally restricted conjunction operator such that *SEQ*(E, F) = *AND*(E, F) **WHERE** $E.t_e < F.t_e$. Using the relations between the sequence operator and the conjunction operator, the hierarchy shown in Figure 2 can be defined. The conjunction hierarchy shown in Figure 2 defines restrictions of conjunctive combinations of events as sequential combinations, simultaneous combinations, or the inverse of sequential combinations. Section V-B further discusses the conjunction hierarchy to describe the event operators discussed in this work.

B. Operator Hierarchy

Allen's thirteen relations provide a powerful way to express relationships among interval-based events. However, there are

[1]The paper [5] defines it as ISEQ. As this work also defines an operator called **ISEQ** to denote inverse **SEQ**, we use *iseq* to denote an interval sequence operator.

2^{13}-1 (8191) total relations when Allen's relations are combined using disjunctions. When all 8191 relations are treated as operators, then the complexity of pattern specification reduces, although, the large number of event operators to specify an event pattern is undesirable from a language point of view. Further, it is not practical to define all of the operators. To cope with this situation, this section describes an operator hierarchy that defines a small set of event operators as shown in Figure 4.

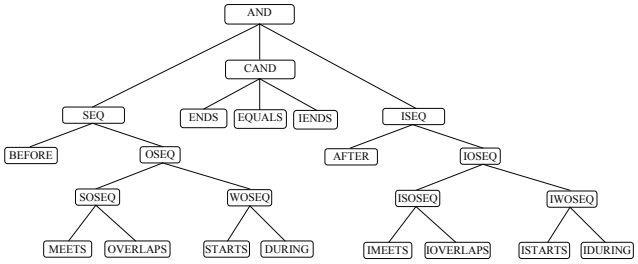


Fig. 4: Operator Hierarchy Defining Conjunction and Sequence

1) Conjunction Operators: Consider the hierarchy shown in Figure 2. The figure shows the trichotomy between two intervals i_1 and i_2 that defines sequence to describe $i_1 < i_2$, simultaneous to describe $i_1 = i_2$, and the inverse sequence to describe $i_1 > i_2$. This trichotomy between intervals considers two events as simultaneous if they end at the same time period. So, the actual relation here is described by a trichotomy between end time-points of two intervals expressed as natural numbers. In other words, if t_{e1} is the end time of the interval i_1 and t_{e2} is the end time of the interval i_2, then $i_1 < i_2$ if and only if $t_{e1} < t_{e2}$, $i_1 = i_2$ if and only if $t_{e1} = t_{e2}$, and $i_1 > i_2$ if and only if $t_{e1} > t_{e2}$. With this idea, the **AND** operator is divided into three different operators **SEQ**, **CAND**, and **ISEQ** as shown in Figure 4. Conceptually, the **AND** operator defines the relation that includes all of Allen's relations, since conjunction has no temporal constraint. The **CAND** operator is meant for concurrent conjunction and, as there are three interval relations specifying the same end time, Figure 4 defines three of Allen's operators, *ends*, *equals*, and *ended by* as **ENDS**, **EQUALS**, and **IENDS**, respectively, as specializations of a concurrent conjunction. Two other operators **SEQ** and **ISEQ** are the inverse of each other and this work does not discuss **ISEQ** in detail as its concepts can be derived from the **SEQ** operator. The hierarchy shown in Figure 4 defines the equivalent event patterns shown in Table II.

Example: In Figure 1, up to time $t = 12$, we can observe that $\mathit{CAND}(E_2, E_4)$ is detected as $\mathit{CAND}^{[8,10]}(E_2, E_4)$. Also notice that the **CAND** pattern is equivalent to the equivalence 3 in Table II, where $\mathit{AND}(E_2, E_4)$ is detected as $\mathit{AND}^{[1,5]}(E_2, E_4)$, $\mathit{AND}^{[2,5]}(E_2, E_4)$, $\mathit{AND}^{[1,10]}(E_2, E_4)$, $\mathit{AND}^{[2,10]}(E_2, E_4)$, $\mathit{AND}^{[5,10]}(E_2, E_4)$, $\mathit{AND}^{[3,7]}(E_2, E_4)$, $\mathit{AND}^{[3,10]}(E_2, E_4)$, and $\mathit{AND}^{[8,10]}(E_2, E_4)$. Similarly other equivalences can be verified.

TABLE II: Equivalent Event Patterns

1)	$\mathit{AND}(E, F)$	$\equiv$	$\mathit{OR}(\mathit{SEQ}(E, F), \mathit{CAND}(E, F), \mathit{ISEQ}(E, F))$
2)	$\mathit{SEQ}(E, F)$	$\equiv$	$\mathit{AND}(E, F)$ **WHERE** $E.t_e < F.t_e$
3)	$\mathit{CAND}(E, F)$	$\equiv$	$\mathit{AND}(E, F)$ **WHERE** $E.t_e = F.t_e$
4)	$\mathit{CAND}(E, F)$	$\equiv$	$\mathit{OR}(\mathit{ENDS}(E, F), \mathit{EQUALS}(E, F), \mathit{IENDS}(E, F))$
5)	$\mathit{ISEQ}(E, F)$	$\equiv$	$\mathit{AND}(E, F)$ **WHERE** $E.t_e > F.t_e$
6)	$\mathit{ENDS}(E, F)$	$\equiv$	$\mathit{CAND}(E, F)$ **WHERE** $E.t_s < F.t_s$
7)	$\mathit{EQUALS}(E, F)$	$\equiv$	$\mathit{CAND}(E, F)$ **WHERE** $E.t_s = F.t_s$
8)	$\mathit{IENDS}(E, F)$	$\equiv$	$\mathit{CAND}(E, F)$ **WHERE** $E.t_s > F.t_s$

2) Sequence Operators: In Figure 3, there are five Allen's relations that are clustered within the hierarchy of the sequence operator with respect to the definition discussed in the previous sections. The other five Allen's relations correspond to the inverse of sequence that can be described similarly as the sequence hierarchy is described. The remaining three Allen relations correspond to concurrent conjunction as discussed in Subsection V-B1. Figure 4 depicts that the five Allen's relations *before*, *meets*, *overlaps*, *starts*, and *during* are categorized into two different groups defining a sequence that does not overlap ($\mathit{BEFORE}(E, F)$ implied by Allen's *before* relations and an overlapping sequence ($\mathit{OSEQ}(E, F)$) implied by the other four relations. The overlapping sequence can be further sub-divided into two groups based upon the relationships between the starting time points of the intervals. For the first division, the sequence of E and F has $E.t_s < F.t_s$ and for the second division $E.t_s \geq F.t_s$. The former sub-division is given the name, strong overlapping sequence ($\mathit{SOSEQ}(E, F)$) where both the start time points and the end time points satisfy the $<$ relation. The later sub-division is understood as a weak overlapping sequence ($\mathit{WOSEQ}(E, F)$), where a start time is strictly not following the $<$ relation. Using the hierarchy shown in Figure 4, the equivalent event patterns for sequence operators can be similarly defined as in Table II, which are omitted due to space constraints. *Example:* In Figure 1, up to time $t = 12$, we can observe that $\mathit{SOSEQ}(E_1, E_2)$ is defined as $\mathit{SOSEQ}^{[1,4]}(E_1, E_2)$ and $\mathit{SOSEQ}^{[7,10]}(E_1, E_2)$ and $\mathit{WOSEQ}(E_1, E_2)$ is defined as $\mathit{WOSEQ}^{[1,3]}(E_1, E_2)$. With this, $\mathit{OSEQ}(E_1, E_2)$ is detected as one of the **SOSEQ** or the **WOSEQ** pattern is detected that verifies the equivalence: $\mathit{OSEQ}(E_1, E_2) \equiv \mathit{SOSEQ}(E_1, E_2)$ *OR* $\mathit{WOSEQ}(E_1, E_2)$.

VI. Experiments and Results

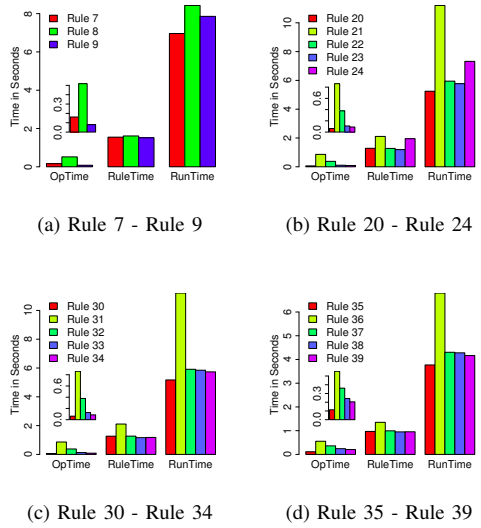
A. Experimental Setup

The experiments were conducted with 12 different pattern groups having equivalent patterns corresponding to each operator from the hierarchy with the implementation of the **AND** operator and the operators from subtrees rooted at **SEQ** and **CAND** in Figure 4. Table III shows examples of two pattern groups, where the first group has three equivalent patterns defined for the **ENDS** operator (Rule 7 - Rule 9) and the second group has five equivalent patterns defined for the **OVERLAPS** operator (Rule 35 - Rule 39). For space reasons, discussion of all the groups with equivalent patterns are omitted from this paper.

TABLE III: Examples of Equivalent Pattern Groups

No.	Pattern
7	$ENDS(E_4(), E_2())$
8	$AND(E_4(), E_2())$ **WHERE** $E_4.t_e = E_2.t_e \wedge E_4.t_s > E_2.t_s$
9	$CAND(E_4(), E_2())$ **WHERE** $E_4.t_s > E_2.t_s$
35	$OVERLAPS(E_5(), E_2())$;
36	$AND(E_5(), E_2())$ **WHERE** $E_5.t_s < E_2.t_s \wedge E_2.t_s < E_5.t_e \wedge E_5.t_e < E_2.t_e$
37	$SEQ(E_5(), E_2())$ **WHERE** $E_5.t_s < E_2.t_s \wedge E_2.t_s < E_5.t_e$
38	$OSEQ(E_5(), E_2())$ **WHERE** $E_5.t_s < E_2.t_s \wedge E_2.t_s < E_5.t_e$
38	$SOSEQ(E_5(), E_2())$ **WHERE** $E_2.t_s < E_5.t_e$

Each pattern was run 10 times for an episode of 3000 time units. For each run, the total time taken by all operators (OpTime), the total time taken by the rule processor for processing a rule after an event to be processed has been identified by event processor (RuleTime), and the total time taken by the event processor (RunTime) were recorded.



(a) Rule 7 - Rule 9 (b) Rule 20 - Rule 24

(c) Rule 30 - Rule 34 (d) Rule 35 - Rule 39

Fig. 5: Experimental Results of Run-Time Performance of Operators

B. Experimental Results

Figure 5 shows the comparisons of run-time for four different pattern groups. In each sub-figure, the first group of bars shows the OpTime, the second group of bars shows the RuleTime, and the third group of bars depicts the RunTime. Notice that each graph in Figure 5 has a graph in an inset to show the magnified form of the OPTime. In all of the graphs and all of the bar groups, the first bar shows the running time for the pattern using the operators designed in this work. Similarly, the second is associated with the equivalent pattern using the **AND** operator with temporal constraints. Other bars represent equivalent patterns, as discussed in Section V, using the parent operator with the temporal constraints, or the use of disjunction of the immediate children operators (See Figure 4).

From the experiments with run-time performance, the following results about the event operators can be observed:

a) 1: The RunTime for patterns defined using the operators from the operator hierarchy is minimum compared to all other alternatives at the higher levels of the hierarchy due to the filtering of incoming events prior to processing them (except for the **ENDS** operator), while an alternative scheme does the post-processing of incoming events.

b) 2: The RuleTime is better than other alternatives for the patterns using the operator set defined in this work, except for some patterns with the graphs shown in Figure 5.

a The graph in the sub-figure 5a shows that the pattern using the **ENDS** operator (Rule 7) takes more time to process the rule than the pattern defined using the **CAND** operator (third bar - sub-figure 5a – Rule 9). This is the direct consequence of processing the event buffer required for the **ENDS** operator and filtering the events after the buffer management. Whereas, Rule 9 detection does not maintain a buffer and events are filtered prior to the detection process.

b The RuleTime for the pattern using the **SOSEQ** operator (sub-figure 5b, first bar – Rule 20) is greater than the pattern using the **OSEQ** operator with temporal constraints (fourth bar – Rule 23). Though Rule 20 spends less time in processing event operators, Rule 20 uses expensive operations such as pattern duplication to manage partial patterns. This makes the RuleTime for Rule 20 greater than Rule 23. In a similar manner, the RuleTime for the patterns using the **MEETS** operator, represented by the first bar (Rule 30) in the sub-figure 5c is higher than the pattern using the **OSEQ** operator (fourth bar – Rule 33) with temporal constraints and the pattern using the **SOSEQ** operator (fifth bar - Rule 34) with temporal constraints. Also, Rule 35 using the **OVERLAPS** operator (first bar – sub-figure 5d) has a RuleTime greater than Rule 38 (fourth bar) using the **OSEQ** operator with temporal constraints and Rule 39 (fifth bar) using the **SOSEQ** operator with constraints.

c) 3: The OpTime is better for the patterns using the operators discussed in this work than other alternative representations for all the groups except for pattern using the **ENDS** operator (Figure 5a- Rule 7). For reasons discussed above, in case of the **ENDS** operator's buffer management and post processing filtering of events, Rule 9 runs faster than Rule 7.

d) 4: Processing with use of the new set of operators always runs faster than the use of the **AND** operator with temporal constraints for all cases of run-time comparisons.

e) 5: When a complex pattern is defined by the temporal constraints among different groups, then it is appropriate to define them using the closest upper level operator with temporal constraints or the disjunction of different operators. This result is seen from the run time comparisons of patterns shown in the graphs represented by the third and beyond bars.

As a conclusion, with the analysis of the run time results discussed above, the set of operators from the operator hierarchy are performing better than using other alternative approaches

that use parent operators from the hierarchy with additional temporal constraints or the disjunction of the children operators from the operator hierarchy in terms of total running time.

VII. Conclusions

The work in this paper has identified ambiguities in the definition of event operators in current event processing languages. The conjunction operator and its relationship with the sequence operator is used to define several possible sequential operations using the idea of Allen's interval relations and a relation hierarchy. The definition of the operator hierarchy defines how an event operator should be selected to achieve the required semantics, making the event specification semantically clear. All the operators discussed in this paper were evaluated by comparing the run-time performance. The experimental results showed that the new set of operators performs better than other alternative approaches on run-time.

There are several possible future research directions. The repetition operator is one of the powerful constructs in event pattern specification. Current event processing systems, however, define the repetition operator in an incomplete way. For example, if one specifies five occurrences of an event E, is it that we are expecting sequential repetition over the time (semantics of **SEQ**) or that the repetition does not have any temporal constraints (semantics of **AND**)? Other issues related to the definition of event operators, such as event time computation and event detection have not been addressed in this work and are left as future work.

References

[1] R. S. Barga, J. Goldstein, M. Ali, and M. Hong, "Consistent Streaming Through Time : A Vision for Event Stream Processing," in *3rd Biennial Conference on Innovative Data Systems Reseaarch (CIDR)*, 2007, pp. 363–374.

[2] F. Bry and M. Eckert, "Rule-Based Composite Event Queries: The Language XChangeEQ and Its Semantics," in *Web Reasoning and Rule Systems*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, vol. 4524, pp. 16–30. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-72982-2_2

[3] S. Chakravarthy and D. Mishra, "Snoop : An Expressive Event Specification Language For Active Databases," *Data Knowl. Eng.*, vol. 14, no. 1, pp. 1–26, 1994.

[4] R. Adaikkalavan and S. Chakravarthy, "SnoopIB: Interval-Based Event Specification and Detection for Active Databases," in *Advances in Databases and Information Systems*. Springer-Verlag Berlin Heidelberg, 2003, pp. 190–204. [Online]. Available: http://www.springerlink.com/content/d3n1vnj0bhp2cdpm

[5] M. Li, M. Mani, E. A. Rundensteiner, and T. Lin, "Complex event pattern detection over streams with interval-based temporal semantics," in *Proceedings of the 5th ACM international conference on Distributed event-based system*, ser. DEBS '11. New York, NY, USA: ACM, 2011, pp. 291–302. [Online]. Available: http://doi.acm.org/10.1145/2002259.2002297

[6] J. F. C. . Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM*, vol. 26, pp. 832–843, 1983.

[7] N. H. Gehani, H. V. Jagadish, and O. Shmueli, "Composite Event Specification in Active Databases : Model & Implementation," in *18th International Conference on Very Large Data Bases V*, 1992, pp. 327–338.

[8] S. Gatizu and K. R. Dittrich, "Events in an Active Object-Oriented Database System," pp. 1–14, 1993.

[9] D. Zhu and A. Sethi, "Sel, a new event pattern specification language for event correlation," in *Proceedings of Tenth International Conference on Computer Communications and Networks*, 2001, pp. 586–589.

[10] B. Mozafari, K. Zeng, and C. Zaniolo, "High-performance complex event processing over xml streams," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '12. New York, NY, USA: ACM, 2012, pp. 253–264. [Online]. Available: http://doi.acm.org/10.1145/2213836.2213866

[11] P.-s. Kam and A. W.-C. Fu, "Discovering temporal patterns for interval-based events," in *Proceedings of the Second International Conference on Data Warehousing and Knowledge Discovery*, ser. DaWaK 2000. London, UK, UK: Springer-Verlag, 2000, pp. 317–326. [Online]. Available: http://dl.acm.org/citation.cfm?id=646109.679272

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1483–1490

Visual Programming of MPI Applications: Debugging and Performance Analysis

Stanislav Böhm, Marek Běhálek, Ondřej Meca, Martin Šurkovský
Department of Computer Science
FEI VŠB Technical University of Ostrava
Ostrava, Czech Republic
stanislav.bohm@vsb.cz, marek.behalek@vsb.cz, ondrej.meca@vsb.cz, martin.surkovsky@vsb.cz

***Abstract*—Our research is focused on the simplification of parallel programming for distributed memory systems. Our overall goal is to build a unifying framework for creating, debugging, profiling and verifying parallel applications. The key aspect is a visual model inspired by Colored Petri Nets. In this paper, we will present how to use the visual model for debugging and profiling as well. The presented ideas are integrated into our open source tool Kaira.**

I. Introduction

PARALLEL computers with distributed memory have recently become more and more available. A lot of people can participate in developing software for them, but there are well-known difficulties of parallel programming. Therefore for many non-experts in the area of parallel computing (even if they are experienced sequential programmers), it can be difficult to make their programs run in parallel on a cluster computer. The industrial standard for programming applications in the area of distributed memory systems is *Message Passing Interface* (MPI)[1]. It represents a quite low-level interface. There are tools like *Unified Parallel C*[2] that simplify creating parallel applications, but the complexity of their development lies also in other supportive activities. Therefore, even an experienced sequential programmer can spend a lot of time learning a new set of tools for debugging, profiling, etc.

The overall goal of our research is to reduce complexity in parallel programming. We want to build a unified prototyping framework for creating, debugging, profiling and formally verifying parallel applications, where a user can implement and experiment with his/her ideas in a short time, create a real running program and verify its performance and scalability. The central role in our approach is a visual programming language (based on Petri nets) that we use for modeling developed applications. In this paper, we present how to use the same model for debugging and profiling. The presented ideas are implemented in Kaira[3], a tool that we are developing.

The work is partially supported by: GAČR P202/11/0340, the European Regional Development Fund in the IT4Innovations Center of Excellence project (CZ.1.05/1.1.00/02.0070)

[1]http://www.mpi-forum.org/
[2]http://upc.lbl.gov/
[3]http://verif.cs.vsb.cz/kaira

II. Tool Kaira

This section serves as an overview for our tool Kaira; for more details see [1], [2]. Our goal is to simplify the development of MPI parallel applications and create an environment where all activities are unified under one concept.

The key aspect of our tool is the usage of a visual model. In the first place, we have chosen the visual model to obtain an easy and clear way how to describe and expose parallel behavior of applications. The other reason is that a distributed state of the application can be shown through such visual model. The representation of an inner-state of distributed applications by a proper visual model can be more convenient than traditional ways like stack-traces of processes and memory watches. With this feature, we can provide visual simulations where the user can observe a behavior of developed applications. It can be used for incomplete applications from an early stage of the development. In a common way of developing MPI programs, it often takes a long time to get the developed application into a state where its behavior can be observed. In context of this paper, the visual model is also useful for debugging and a performance analysis as will be demonstrated later.

On the other hand, we do *not* want to create applications completely through the visual programming. Sequential parts of the developed application are written in the standard programming language (C++) and combined with the visual model that catches *parallel aspects* and *communication*. We want to avoid huge unclear visual diagrams; therefore, we visually represent only what is considered as "hard" in parallel programming. Ordinary sequential codes are written in a textual language. Moreover, this design allows for easy integration of existing C++ codes and libraries.

It is important to mention that our tool is *not* an automatic parallelization tool. Kaira does not discover parallelisms in applications. The user has to explicitly define them, but they are defined in a high-level way and the tool derives implementation details.

Semantics of our visual language is based on Coloured Petri nets (CPNs)[3]. Petri nets is a formalism for the description of parallel processes. They also provide well-established terminology, a natural visual representation for visual editing

978-83-60810-53-8/$25.00

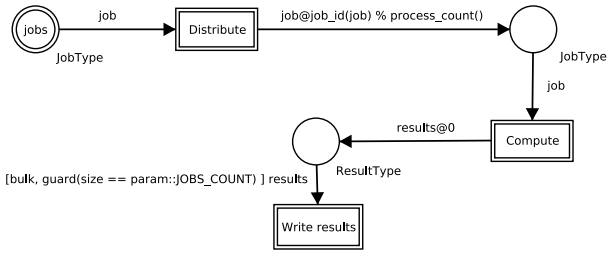


Fig. 1. The example model

of models and their simulations. Modeling tool *CPN Tools*[4] was also the great inspiration for us (especially how to visualize the model).

To demonstrate how our model works, let us consider the model in Figure 1. It presents a problem where some jobs are distributed across computing nodes and results are sent back to process 0. When all the results arrive, they are written into a file. Circles (*places* in terminology of Petri nets) represent memory spaces. Boxes (*transitions*) represent actions. Arcs run from places to transition (*input arcs*) or from transition to places (*output arcs*). The places contain values (*tokens*). Input arcs specify what tokens a transition needs to be *enabled*. An enabled transition can be executed. When a transition is executed, it takes tokens from places according to input arcs. After finishing the computation of the transition, new tokens are placed into places according to output arcs. In CPNs places store tokens as multisets, in our approach we use queues.

A double border around of a transition means that there is a C++ function inside and it is executed whenever the transition is fired. A double border of a place indicates an associated C++ function creating the place's initial content. Arcs' inscriptions use C++ enriched by several simple constructions. A computation described by this model runs on every process. Tokens can be transferred between processes by expressions after "@" symbol on output arcs.

As a more advance example, we use the heat flow problem on a cylinder. We will use a version of this problem where the body is discretized by a grid depicted in Figure 2. The implementation of this problem in Kaira is depicted in Figure 3. The transition *Compute* executes single iteration of the algorithm. It takes a process' part of the grid and two rows, one from neighbor above and one from below. It updates the grid and sends top and bottom rows to neighbors. When the limit of iterations is reached then the results are sent to process 0 where they are written. The init area (depicted as the blue rectangle) is used to set up initial values of places not only on process 0 but over specified processes (all processes in our case). The measurements of this program and a comparison to the sequential version are part of Section V.

[4]http://cpntools.org/

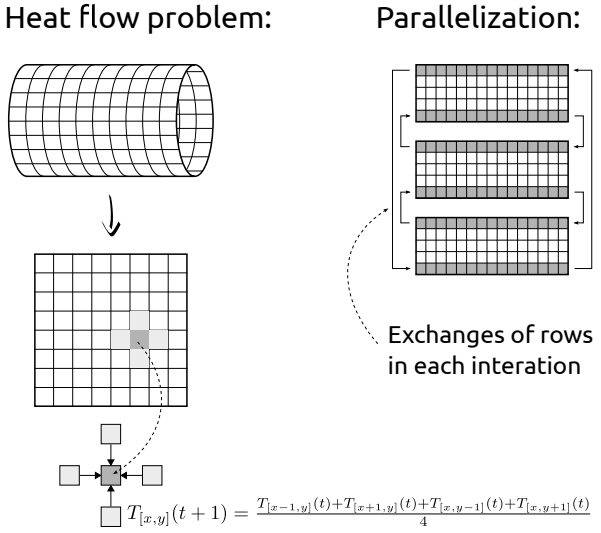


Fig. 2. The heat flow problem on a cylinder and the used method of parallelization

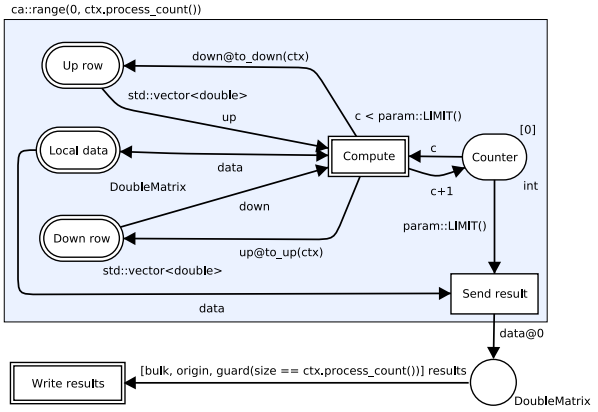


Fig. 3. The implementation of the heat flow problem in Kaira

III. Simulations and records of generated applications

In this section, we will introduce two crucial features: *simulations* and *tracing* of generated applications. Both can be used for debugging and the latter for profiling. Later we will describe two other features and we will also discuss the drawbacks of our approach.

A. Simulations

Besides generating standalone parallel applications from the model, the user can also run the developed application in the simulator. The main task of the simulator is to expose an inner state and it allows for controlling a run of the generated application. The inner state is shown in the form of labels over the original model (see Figure 4). The three types of information are depicted:

- Tokens in place (The state of memory)
- Running transitions (The state of execution)

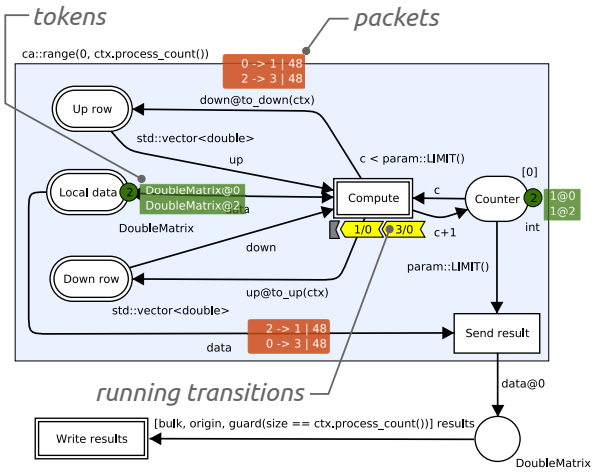


Fig. 4. The model in the simulator

- Packets transported between nodes (The state of the communication environment)

It completely describes a distributed state of the application. The user can control the behavior of the application by the three basic actions:

- Start an enabled transition
- Finish a running transition.
- Receive a packet from a network.

By executing these three types of actions, the application can be brought to any reachable state. The model naturally hides irrelevant states during sequential computations and only aspects important to parallel execution are visible and controllable.

This approach also gives us the possibility to observe the behavior of the application in a very early state of the development without any additional debugging infrastructure. For example, we can see which data are sent to another process even if there is no implementation of the receiving part.

The user has complete well-formed control of the application in the simulator; therefore, the application can be put into an interesting state (and the user can observe the consequences) even if the application rarely reaches such state.

B. Tracing

An application developed in Kaira can be generated in the tracing mode, where activities of a run of the application are recorded into a *tracelog*. When the application finishes its run, the tracelog can be loaded back into Kaira and used for the *visual replay* or for statistical summaries. Generally, issues with such post-mortem analysis can be categorized into these basic groups: *selection what to measure*, *instrumentation* and *presentation of results*. Such tracelogs can be useful both for profiling and debugging.

In the case of debugging, we usually want to collect detailed information of the run for the reconstruction of the cause of the problem. In the case of profiling, we want to discover performance issues and therefore need to measure a run with time characteristics as close as possible to real runs of the application. But the measurement itself creates an overhead that devalues the gathered information about performance. Therefore, in both cases, it is important to specify what to store in the tracelog. In common profilers, specifications of measurements are usually implemented as a list of functions that we want to measure/filter out. But it can be a non-trivial task to assemble such a list, especially in the case when we use some third-party libraries with an unclear purpose to the user. It often needs some experience to recognize what can be safely thrown away.

In Kaira, the user specifies what is measured in terms of places and transitions. It is done just by placing labels in a model (Figure 6). The tracing of transitions enables the recording of information about their execution. The tracing of places enables the recording of information about tokens that go through them. The user can easily control what to measure and it is obvious what information will be gained or lost after switching on or off each setting. Moreover, our approach also allows for simply enriching the model by more detailed tracing. Places and transitions can trace additional data. It is implemented as connecting functions to places and transitions.

The usage of this feature is demonstrated in the experiment in Section V. The experiments also demonstrate tracelog sizes so even if we trace all transitions and names of all tokens in places (that is useful for debugging), sizes of tracelogs are usually manageable. The recording of high-level information from the perspective of our visual model is far from recording every function call in the program.

The second task is the *instrumentation*, i.e. putting the measuring code inside the application. In our case, Kaira can automatically place the measuring codes during the process of generation of the parallel application. Parallel and communication parts are generated from the model, therefore we know where are interesting places where to put measuring codes. By this approach, we can obtain a traced version of an application that does not depend on the complier or computer architecture. In contrast to a standard profiler or debugger for generic applications, we do not have to deal with a machine code or manual instrumentation.

As we already said, the results are presented to the user in the form of a visual replay or as statistical summaries. In replay, the data stored in tracelog are shown in the same way as in the simulator, thus as the original model with tokens in places, running transitions and packets on the way (Figure 5). The user can jump to any state in the recorded application. Our tool also provides statistical summaries and standard charts like a normal profiler, and additionally, information is presented using the terms of the model. For example, the utilization of transitions (Figure 10), the numbers of tokens in places, etc.

C. Combination of simulation and recording

The useful feature for debugging parallel applications is a technique usually called *deterministic replay* [4]. Existing

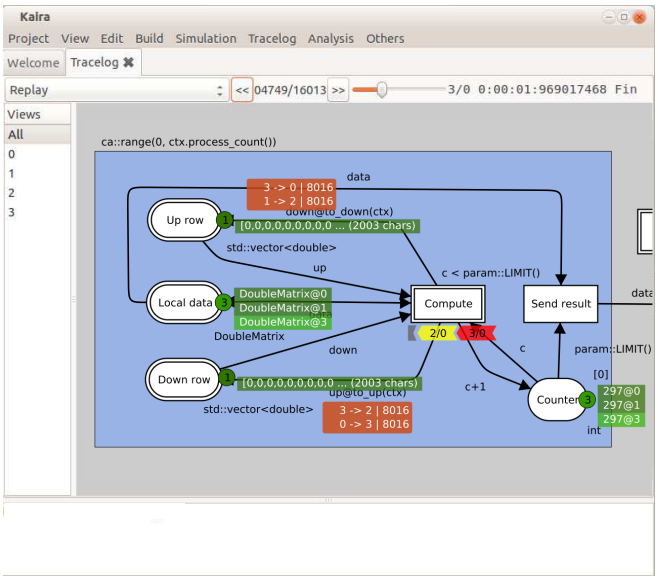


Fig. 5. The screenshot of a replay.

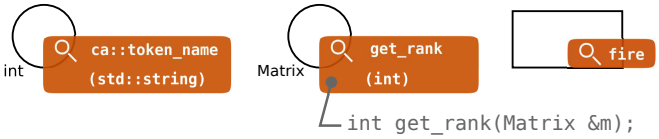


Fig. 6. Tracing labels, from left: Tracing names of tokens that arrive into the place; tracing values obtained by applying a function to each token arriving to this place; tracing transition firing.

tools use the *data-replay*, the *order-replay* or some combination of both approaches. In the data-replay approach, every communication message is recorded and a single process can be rerun with the same communication as was recorded. The advantage is that it is feasible even for instances with many processes. The disadvantage is huge tracelogs and it can be hard to discover errors that need an overall context. In the order-replay approach, we store the ordering of incoming messages. We get smaller tracelogs, but we must simulate all processes during the replay.

In Kaira we have implemented the order-replay approach in the form of *control sequences*. This feature naturally connects the infrastructure of our simulator with tracing abilities. A control sequence is a list containing actions. Each action is one of three basic types from Section III (starting and finishing transitions and receiving packets). Actions contain information about the process and the thread where the activity is executed, the transition's name (in the case of transition firing) and the source process of the message (in the case of receiving packets). When we store this information we are able to repeat the run of the application.

Sequences are generated in the simulator or they are extracted from tracelogs. The simulator can replay sequences and get the application into the desired state. Because the control sequence and the model are loosely connected, the

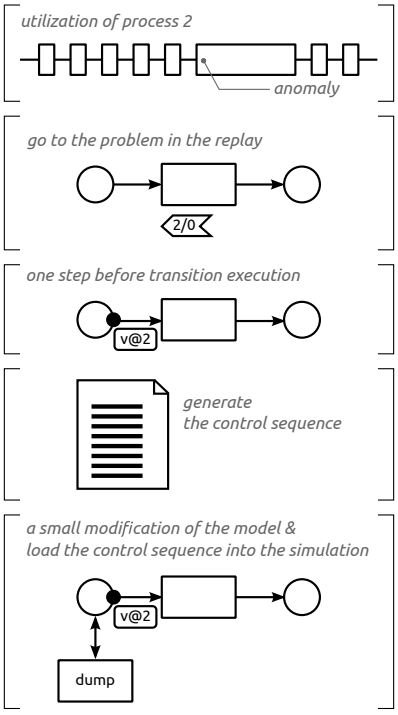


Fig. 7. The use case of control sequences

sequence remains relevant even if we make some changes into the model. The usefulness can be exposed by the following scenario: The user finds a problem by a visual replay or by summaries obtained from a tracelog. Then a sequence that brings the application exactly one step before the problem can be exported from the tracelog. Then the model can be enriched by more precise debugging outputs. For example, it can be a `printf` added into a transition's code or an extra debugging transition. Now we can get the application into the state before the problem by replaying the sequence in the simulator. In this situation, we have the possibility to obtain more information about the problem because of the modified version of the application. This scenario is captured in Figure 7.

D. Other features

Our model allows implementing two additional features that can be used for debugging or the performance analysis. Because we control how parallel aspects are generated, we can always generate the sequential application from the model. Such generated application works like the original one but it is performed exactly by one thread independently on how many processes are specified. This feature does not need any change in the model. It enables easy profiling and debugging of sequential parts of developed programs by the tools designed for sequential applications without problems caused by threads or MPI.

The other feature is the possibility to connect into a running application. We can start a generated application in a mode where the application listens on a TCP port. The application normally runs but when we connect to this port, the run is

paused and the inner state of the application is displayed in the simulator. The application can be also controlled in the same way. When the connection is closed, the application continues computing. This way we can easily debug situations when the application hangs up or we can just observe how far the computation is. But in the current implementation there are some limitations. This feature works only for applications generated with the thread backend (i.e. it does not work for MPI applications) and after the connection to the application, the control is passed to the user after finishing all current running transitions.

E. Drawbacks

Here, we want to discuss the drawbacks of our approach. The most obvious issue is that our approach does not give us any tool to debug or profile codes in places and transitions. We can say what data were on the input of the transition, what was the output. We can get the application into a state before or after execution of the transition or profile the transition as a whole. But we cannot observe, debug or profile the inner state of transition executions. This can be a serious problem and may force the user to use other tools in some situations. On the other hand the codes in transitions are sequential codes without any communication so they can be easily profiled or debugged separately. It can be further simplified by the fact that we can always generate the sequential version of the program.

Other issue is connected with our current implementation. We have focused on minimizing the performance impact of the debugging and profiling infrastructure on generated applications. On the other hand, our tool itself was not subject of optimizations, and therefore, processing a huge tracelog or a long control sequence can be time consuming and demanding on memory. Therefore, our infrastructure is not yet suitable for debugging or profiling long running applications. Some numbers to this topic are provided in Section V.

IV. Related Works

In this section, we want to compare Kaira with selected tools for profiling and debugging. For the comparison with other types of tools we refer to [2].

Different approaches have been proposed for debugging MPI applications. More about debugging in MPI environment can be found in [5], [6].

First, a MPI application runs on each computing node like a normal program; therefore, we can use standard tools like *GDB*[5] (for debugging) or *Callgrind*[6] (for profiling). This approach is sufficient to find some types of bugs or performance issues, but the major disadvantage is completely separated instances of the supportive tool for each process. It is not easy to control more debugger instances simultaneously or merge several profiler's outputs.

There are specialized debuggers and profilers to overcome this issue. For debugging there are tools: *Distributed Debugging Tool*[7] or *TotalView* [8]. They provide the same functionality like ordinary debuggers (stack traces, breakpoints, memory watches), but they allow to debug a distributed application as a single piece. Besides these tools, there are also non-interactive tools like *MPI Parallel Environment*[9]. It provides additional features over MPI, like displaying traces of MPI calls or real-time animations of communication. These tools are universal in the sense that they can debug any application. In our approach, we can only debug applications created in Kaira. On the other hand, we are able to provide the debugging infrastructure on a higher level of abstraction than source codes.

To deal with hundreds of processors, there are also automatic debugging tools. These tools usually use the static analysis of source codes to discover misuse of MPI calls (*MPI-Check*) or the analysis based on the state space exploration (*ISP* [7]).

As it was mentioned in previous sections, a potentially powerful technique for debugging of MPI applications is the deterministic replay. An example of a tool implementing this approach is MPIWiz [4].

In the case of profilers for parallel applications, one of the most successful freely available tools is *Scalasca* [8]. The big advantage is the ability to work in an environment of thousands of processors. Scalasca implements the direct instrumentation approach, it provides data summarizations at a runtime or traces for postmortem analyses. In the tracing mode, Scalasca records performance related events. Summarized performance profiles are based on functions call paths. In both cases, resulting reports can be interactively explored in the graphical browser.

The similar tool to Scalasca is *TAU* (Tuning and Analysis Utilities) [9]. It is capable of gathering performance information through instrumentation of functions, methods, basic blocks, and statements. From this perspective, both Scalasca and TAU adopt similar strategies. The main difference is in the low level measuring systems.

There are also different tools that focus mainly on the visualization of traced data like Vampir [10] and Paraver [11]. These tools are able to import tracelogs produced by others and the user is able to browse traced data. Usually, a set of filters can be specified to remove unnecessary details.

V. Experiments

This section contains two example programs. Their purpose is to demonstrate features mentioned in Section III. All programs were executed on a machine with 8 processors AMD Opteron/2500 (32 cores in total) and compiled with Intel Compiler at the optimization level `-O2`.

A. Basic measurements

As the first example, we show results for the heat flow problem introduced at the end of Section II. In this example, we

[5]http://www.gnu.org/software/gdb/
[6]http://valgrind.org/docs/manual/cl-manual.html
[7]http://www.allinea.com/products/ddt/
[8]http://www.roguewave.com/
[9]http://www.mcs.anl.gov/research/projects/perfvis/software/MPE/

show a comparison between the hand-made solution profiled by Scalasca and the version created and profiled in Kaira. Both implementations are distributed together with Kaira.

The implementations share the same computation code. It is about 380 LOC (lines of code without comments). The solution in Kaira contains 25 LOC in transitions and places and 10 LOC for binding the external types. The hand-made solution contains 100 LOC, which are not shared with the solution in Kaira. The following experiments were executed on the instance of the size 6400×32000 and 300 iterations.

All places and transitions are traced except for places *Up row* and *Down row*. The standard function writing token name for the type `std::vector<double>` stores all values from the vector into the tracelog. (6400 doubles for places *Up row* and *Down row*). In our example, we do not need such information, so we can change this writing function to store a smaller amount of data or we can just switch off the tracing (as we have done here). In the case of the hand-made solution profiled by Scalasca, 5 patterns in the filter file was used (23 functions were filtered out). These numbers are small, because of simplicity of the example. For illustration, the code generated by Kaira contains more internal functions and when it is profiled in Scalasca, we had to use 14 patterns in the filter file and 382 functions were filtered out. Without the filter file, Scalasca produces extremely huge logs (in the order of gigabytes) and it deforms runs of the application, because traced data are very often flushed on the disk.

Table I shows the comparison between the solutions generated by Kaira and the handmade solution. In both cases, the measurements were done without writing the resulting matrix into the file. Our problem scales well up to 16 processors, then it reaches limits of used computer.

In case of Scalasca we have instrumented all source files. This instrumentation adds some overhead. It can be improved by additional separation of computation code and communication, but it cannot be always possible. For example when we use an external library.

Figure 8 shows that solution produced by Kaira is comparable to handmade solution and our tracing introduces only a small overhead. For this small number of processors, the measured times are better than the handmade solution profiled by Scalasca. Scalasca is designed for thousands of processors; therefore it is not well suited for this experiment. But our goal was to show that Kaira tracing is comparable (in the scale of tens of processes) with existing mature parallel profilers and Scalasca is a well-established tool in this area.

Figure 9 shows the grow of tracelog sizes. Kaira tracelogs are bigger but still comparable. In Kaira case, they contain information for a replay, not only profile data.

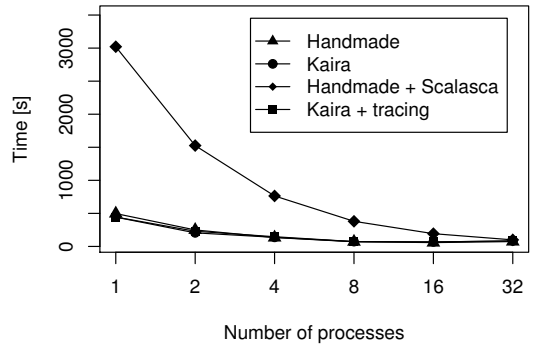


Fig. 8. The comparison of running times between the hand-made solution and the solution generated by Kaira for the heat flow example (based on Table I).

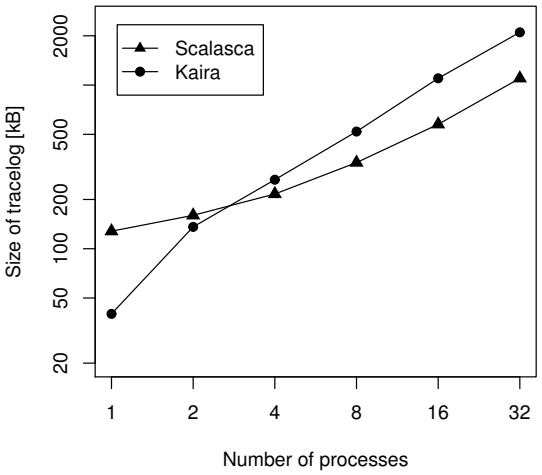


Fig. 9. The comparison of tracelog sizes between the handmade solution traced by Scalasca and Kaira's solution traced by Kaira (based on Table I).

B. Advanced measurement

In this section we will demonstrate how a tracelog can be enriched by custom data and how tool R[10] can be combined with Kaira to obtain various statistics. R is one of the most popular statistical tools. Kaira can export collected data from a tracelog in a form of a table that can be loaded into R. Each row of this table corresponds to the three basic events (explained in Section III) and their subevents (token add, token removed, packet sent). In Kaira's distribution, there is a simple script for R that provides basic operations over such table. It is often easy to extract useful information about the performance from data in this form.

As the example, we have chosen the *Ant Colony Optimization* (ACO) algorithm that is used to solve *Traveling Salesman Problem* (TSP). There are many ways to parallelize this algorithm; the presented solution is described in more detail in the paper [12]. The visual model for the solution is depicted in Figure 11. We will show how to get specific data from the application's run and present them with the help of R.

[10]http://www.r-project.org

TABLE I
MEASURED VALUES FOR THE HAND-MADE SOLUTION AND KAIRA'S SOLUTION OF THE HEAT FLOW EXAMPLE

Number of processes	**1**	**2**	**4**	**8**	**16**	**32**
Handmade solution [s]	497.39	249.58	134.39	70.98	57.88	73.38
Handmade solution + Scalasca [s]	3020.89	1525.62	763.63	380.23	193.08	99.33
Kaira solution [s]	443.5	205.57	137.75	72.95	68.04	83.09
Kaira solution with tracing [s]	444.78	229.67	147	72.98	68.14	83.06
Scalasca log size [kB]	128	160	216	336	576	1126
Kaira log size [kB]	40	136	264	520	1126	2150

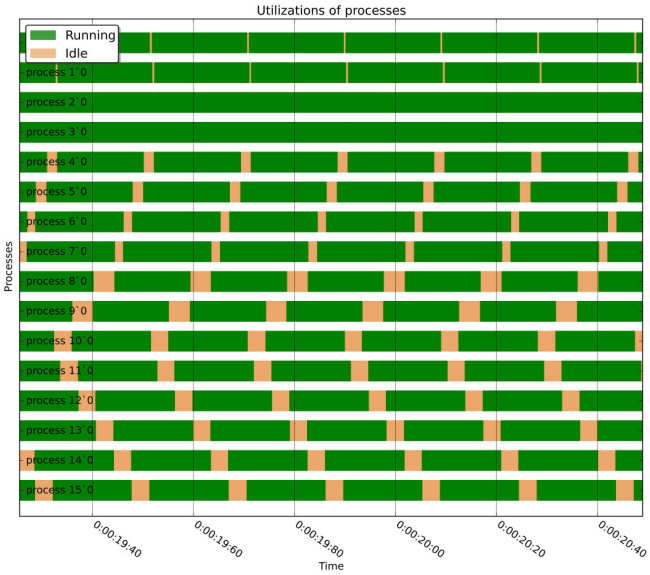


Fig. 10. The example of zoomed chart of process utilizations in the heat flow example with 16 processes.

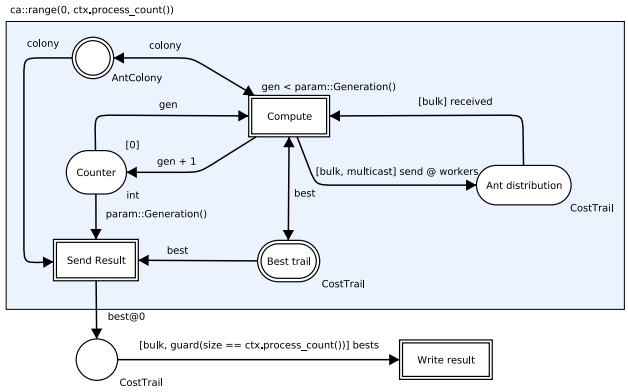


Fig. 11. The implementation of Ant Colony Optimization.

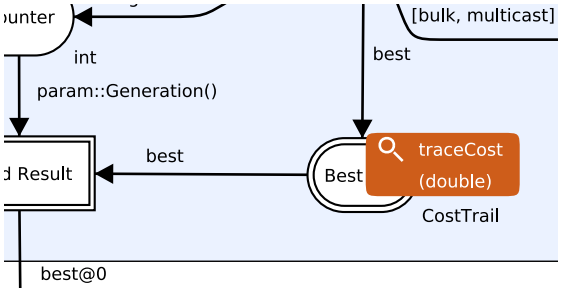


Fig. 12. Connecting a tracing function to the place *Best trail*.

For our experiment, we used the file *eil51.tsp* from TSPLIB[11].

In the used version of the ACO algorithm, ants are separated into colonies and the evolution of each colony is computed in parallel (each colony is assigned to a single MPI process). A colony is stored in the place in the top-left corner. The transition *Compute* takes a colony and computes the next generation of ants. In each iteration, every process saves the best solution to the place *Best trail*. It is distributed to other processes through the place *Ant distribution*. When the last generation is computed, *Send results* takes the best solution and it sends them to process 0, where the overall best solution is chosen.

To verify that the solution works properly, it is useful to inspect the fitness value (i.e. the quality of the solution) in time. We use the ability to connect a tracing function with a place. In our case, we connect a simple function returning a fitness value of an ant to place *Best trail* (Figure 12). When a token arrives to this place, its value is stored in the tracelog (in the scope of an event that creates this token). After exporting the tracelog table into *R*, we obtain the charts in Figures 13 and 14.

[11]It is available at http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/

When the best solution for each colony is sent to others, the convergence is the same for all processes (Figure 13), as we may expect. To check this assumption, we can disable communication, by removing the edge with the expression `[bulk, multicast] send@workers`. The fitness values for this case are shown in Figure 14.

VI. CONCLUSION

In previous papers, we have been focused on the development of MPI applications by usage of the visual model and visual programming. Our visual language is based on well-known formalism – Coloured Petri Nets. In this paper, we have presented how the same visual model and in fact the same approach was used for debugging and performance analyses. The presented ideas are implemented in our tool Kaira.

We introduced a simulator that allows the live introspection into developed programs. This simulator uses the original visual model. Thus the developer is able to inspect the de-

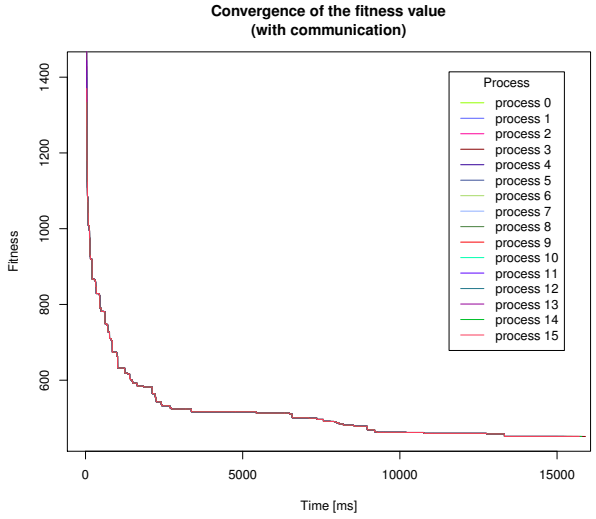


Fig. 13. Minimization of fitness values in time; colonies exchange the best solution.

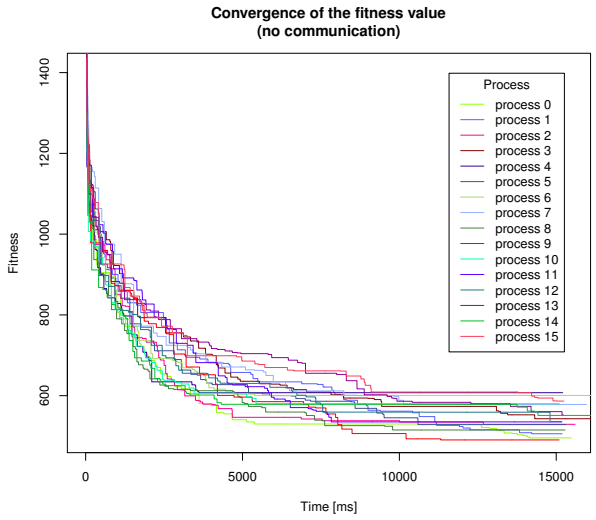


Fig. 14. Minimization of fitness values in time; without communication.

veloped application's behavior using the same visual model that he developed and that he understands. Using control sequences, we are able to capture a simulation and later it can be reproduced even on a modified visual model. They serve as basic infrastructure and they allowed us to implement deterministic replay and we want to implement more advanced features like a massive parallel replay.

Also for profiling we use a similar approach and we use the original model. We use it not only to present the obtained data (application's replay) but also to simplify the measurement specifications. This is crucial for profiling, because when we measure everything, the obtained data are usually useless and setup measurement filters in a standard tool can be hard.

We also demonstrate that presented features can be implemented with a performance that is comparable with existing mature tools. Practical experiments show that a performance of the handmade solution is comparable with the solution generated by Kaira. Measured times differences were up to 20%. The overhead introduced by tracings in Kaira is up to 3%. Our tracelogs are bigger than Scalasca's tracelogs, but their growths is similar.

We consider these features to be a successful step towards providing the unifying framework for prototyping and development of MPI applications. We are also working on more advanced features: performance prediction and verification. These parts are interconnected by our model and results from one analysis can be used in the rest of Kaira infrastructure. It can serve as another argument why to use Kaira.

REFERENCES

[1] S. Böhm and M. Běhálek, "Generating parallel applications from models based on petri nets," *Advances in Electrical and Electronic Engineering*, vol. 10, no. 1, 2012.

[2] ——, "Usage of Petri nets for high performance computing," in *Proceedings of the 1st ACM SIGPLAN workshop on Functional high-performance computing*, ser. FHPC '12. New York, NY, USA: ACM, 2012, pp. 37–48. [Online]. Available: http://doi.acm.org/10.1145/2364474.2364481

[3] K. Jensen and L. M. Kristensen, *Coloured Petri Nets - Modelling and Validation of Concurrent Systems*. Springer, 2009.

[4] R. Xue, X. Liu, M. Wu, Z. Guo, W. Chen, W. Zheng, Z. Zhang, and G. Voelker, "Mpiwiz: subgroup reproducible replay of mpi applications," in *Proceedings of the 14th ACM SIGPLAN symposium on Principles and practice of parallel programming*, ser. PPoPP '09. New York, NY, USA: ACM, 2009, pp. 251–260. [Online]. Available: http://doi.acm.org/10.1145/1504176.1504213

[5] J. M. Squyres, "Mpi debugging – can you hear me now?" *ClusterWorld Magazine, MPI Mechanic Column*, vol. 2, no. 12, pp. 32–35, December 2004. [Online]. Available: http://cw.squyres.com/

[6] ——, "Debugging in parallel (in parallel)," *ClusterWorld Magazine, MPI Mechanic Column*, vol. 3, no. 1, pp. 34–37, January 2005. [Online]. Available: http://cw.squyres.com/

[7] A. Vo, S. Vakkalanka, M. DeLisi, G. Gopalakrishnan, R. M. Kirby, and R. Thakur, "Formal verification of practical mpi programs," *SIGPLAN Not.*, vol. 44, no. 4, pp. 261–270, Feb. 2009. [Online]. Available: http://doi.acm.org/10.1145/1594835.1504214

[8] M. Geimer, F. Wolf, B. J. N. Wylie, E. Ábrahám, D. Becker, and B. Mohr, "The Scalasca performance toolset architecture," *Concurrency and Computation: Practice and Experience*, vol. 22, no. 6, pp. 702–719, Apr. 2010.

[9] S. S. Shende and A. D. Malony, "The tau parallel performance system," *Int. J. High Perform. Comput. Appl.*, vol. 20, no. 2, pp. 287–311, May 2006. [Online]. Available: http://dx.doi.org/10.1177/1094342006064482

[10] A. KnÃźpfer, H. Brunst, J. Doleschal, M. Jurenz, M. Lieber, H. Mickler, M. MÃźller, and W. Nagel, "The Vampir performance analysis tool-set," in *Tools for High Performance Computing*, M. Resch, R. Keller, V. Himmler, B. Krammer, and A. Schulz, Eds. Springer Berlin Heidelberg, 2008, pp. 139–155. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-68564-7_9

[11] V. Pillet, V. Pillet, J. Labarta, T. Cortes, T. Cortes, S. Girona, S. Girona, and D. D. D. Computadors, "Paraver: A tool to visualize and analyze parallel code," In WoTUG-18, Tech. Rep., 1995.

[12] M. Běhálek, S. Böhm, P. Krömer, M. Šurkovský, and O. Meca, "Parallelization of ant colony optimization algorithm using Kaira," in *11th International Conference on Intelligent Systems Design and Applications (ISDA 2011)*, Cordoba, Spain, Nov. 2011.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1491–1498

pLERO: Language for Grammar Refactoring Patterns

Ján Kollár, Ivan Halupka, Sergej Chodarev and Emília Pietriková
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
E-mail: {jan.kollar, ivan.halupka, sergej.chodarev, emilia.pietrikova}@tuke.sk

***Abstract*—Grammar-dependent software development and grammarware engineering have recently received considerable attention. As a significant cornerstone of grammarware engineering, grammar refactoring is, nevertheless, still weakly understood and practiced. In this paper, we address this issue by introducing pLERO, formal specification language for preserving knowledge of grammar engineers, complementing mARTINICA, the universal approach for automated refactoring of context-free grammars. With respect to other approaches, advantage of mARTINICA lies in refactoring on the basis of user-defined refactoring task, rather than of a fixed objective of the refactoring process. To understand the unified refactoring process, this paper also provides a brief insight into grammar refactoring operators, providing universal refactoring transformations for specific context-free grammars. To preserve knowledge considering refactoring process, we propose formalism based on patterns, seen as well-proven way of knowledge preservation in variety of domains, such as software architectures.**

I. Introduction

AUTOMATED grammar refactoring is the field where two or more equivalent context-free grammars may have different forms. Although two equivalent grammars generate the same language, they do not necessarily share other specific properties measurable by grammar metrics [1]. The form in which a context-free grammar is written may have a strong impact on many aspects of its future application. For instance, it may affect general performance of a parser [2], or it may influence, and in many cases limit, the choice of parser generator [2].

Since there is a close relation between the form in which a grammar is expressed and the purpose for which it is designed, different grammars become domain-specific formalizations if generating the same language. Thus, the ability to transform a grammar to another (equivalent), indeed, becomes the power to shift between domains of possible applications. Even if making each grammar more universal in its application scope, the practical benefits may be easily thwart by the difficulties. The problem is, refactoring is often a non-trivial task and if done manually, it is prone to errors, especially with large grammars. This is an issue, as in general there is no formal way to prove two context-free grammars generate the same language.

We addressed this issue in [3] by proposing mARTINICA, metrics Automated Refactoring Task-driven INcremental syntactIC Algorithm. Its main idea is to apply a sequence of simple transformation operators to a chosen context-free grammar to produce an equivalent grammar with the desired properties. Each refactoring operator transforms arbitrary context-free grammar to an equivalent context-free grammar which may have different form than the original. Properties the grammar should possess are defined by so called objective function. That is, the purpose of mARTINICA is to find a sequence of refactoring operator instances transforming particular context-free grammar to an equivalent with a form satisfying user-defined requirements. Current state of the algorithm development requires grammar production rules to be expressed in the BNF notation as it in general, unlike EBNF, expresses elementary properties, e.g. left/right recursion or iteration.

With respect to diversity of possible requirements on the qualitative properties, refactoring operators provide relatively universal grammar transformations. Although the relative universality of refactoring operators contributes to versatility of the algorithm, it also may lead to high computational complexity and, in specific cases, to inability to fulfill the refactoring task. Within the current research, we propose a solution of these issues based on patterns which, in this context, we consider to be a problem-specific refactoring operators.

In general, we consider a pattern to be a problem-solution pair in given context [4] [5]. Alexander argues each pattern can be understood as an element of reality, and of language [4]. As an element of reality, pattern reflects a relation between specific context, certain system of forces recurring in given context, and certain spatial configuration leading to balance in a given system of forces [4]. As an element of language, pattern reflects an instruction showing how certain spatial configuration can be repeatedly used to balance certain system of forces wherever specific context makes it relevant [4].

As such, patterns are tools for documenting existing, well proven design knowledge, supporting construction of systems with predictable properties and quality attributes [5]. Thence, the role of patterns in the field of grammar refactoring is:

1) To preserve knowledge of language engineers about when and how to refactor context-free grammars, and
2) To support process of grammar refactoring by providing this knowledge.

To incorporate patterns in automated grammar refactoring, we have coined a new term: *grammar refactoring patterns*. Each grammar refactoring pattern describes a way in which a context-free grammar can be transformed preserving generated language, and a specific situation of this to be possible. Description of the situation, in which transformation provided by

978-83-60810-53-8/$25.00

a pattern can be applied, defines refactoring problem addressed by the pattern, while grammar transformation defines solution of the refactoring problem.

II. Motivation

Grammarware engineering as an up-and-rising discipline aims at solving grammar development issues, promising an overall rise in grammar quality, and development productivity [6]. Grammar refactoring may occur in many fields, e.g. grammar recovery, evolution and customization [6]. In fact, it is one of five core processes of grammar evolution, alongside the extension, restriction, error correction, and recovery [7]. However, unlike a well-proven practice of program refactoring, grammar refactoring is little understood and practised [6].

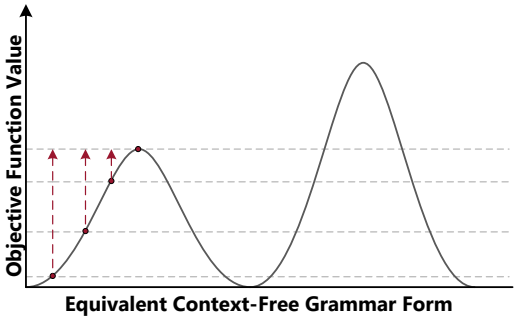


Fig. 1. Previous research approach

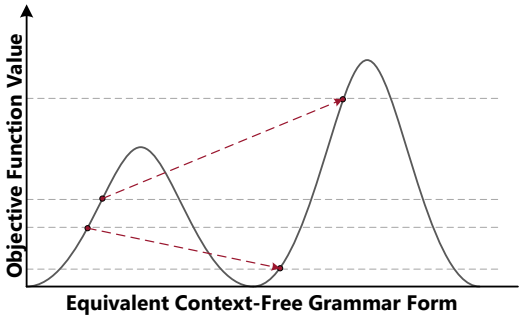


Fig. 2. Current research approach

Fig. 1 and 2 both reflect the objective function value at particular grammar forms. Horizontal axis, indeed, denotes a hypothetic area (not a dimension) of all the equivalent context-free grammar forms, and vertical axis denotes the objective function values of this grammar.

The points marked in the graphs represent forms of a context-free grammar. All the points originate from a single point, corresponding to the initial grammar form and its objective function value.

Our previous approach involved an improvement of the objective function through the application of refactoring operators [3]. Its potential to improve is expressed by the up arrows in Fig. 1. The issue lied in local extrema: If populations reached them, further slide over the function became uncertain.

The algorithm of mARTINICA solved this issue by enabling the populations to regress, but merely in a certain number of steps [8], which is one of the few possible heuristics. Two potential ways of the algorithm are to enable a progress to a certain value or a certain number of steps. However, neither of them is ideal and cannot work universally. Further, the issue lied in a negative impact on the computational complexity as well.

Consequently, the current approach considers *refactoring patterns*. If applied to a grammar, at the corresponding objective function it is possible to skip the local extremus (Fig. 2), what is their primary feature. Certainly, various patterns concern various objective functions. That is, this solution is not universal, however, it is ideal for domain-specific tasks, such as left recursion removal.

Since this approach is not heuristic and it always works, it is considered to be progressive according to the previous one.

Generally, the main idea behind our research lies in grammar modifications according to their objective functions, which is supplemented by the current research dedicated to creation of a tool for grammar modifications according to properties of refactoring operators.

III. Related Work

Unfortunately, it was possible to find very little reported research in the field of automated grammar refactoring. The small amount of the published work is mostly concerned with refactoring context-free grammars achieving some fixed domain-specific objective.

Kraft, Duffy and Malloy developed a semi-automated grammar refactoring approach to replace iterative production rules with left-recursive rules [9]. They present a three-step procedure consisting of grammar metrics computation, metrics analysis to identify candidate nonterminals, and transformation of the candidate nonterminals. The first and third step are fully automated, while the process of identifying nonterminals, to be transformed by replacing iteration with left recursion, is done manually. Since grammar metrics are calculated automatically, this approach is called metrics-guided refactoring. However, the resulting values must be interpreted by human, using them as a basis for making the decisions necessary for resuming the refactoring procedure. The work also provides an exemplary illustration of the grammar refactoring benefits, since left-recursive grammars are more useful for some aspects of the grammar application [10], and are also more useful to human users [11] than iterative grammars.

In the field of compiler design, the procedure of left-recursion removal is a well-known practice. Louden reports an algorithm for automated removal of direct and indirect left recursion [12]. This approach is further extended by Lohmann, Riedewald and Stoy [11], presenting a technique for removing left-recursion in attribute grammars and semantic preservation while executing this procedure.

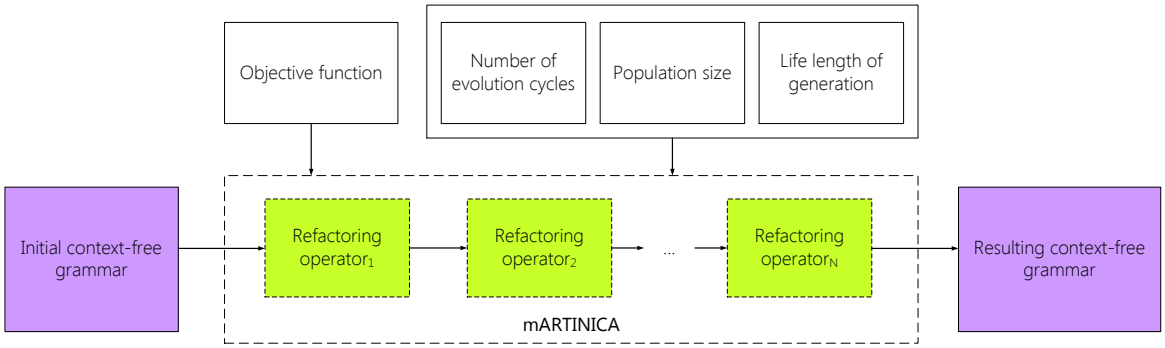


Fig. 3. Black-box view of mARTINICA

Lämmel presented suite of fifteen grammar transformation operators, four considering grammar construction, five considering grammar destruction and six considering grammar refactoring [13]. These operators are in large degree tailored for solving issues of two specific problem domains e.g. grammar adaptation and grammar recovery. Paper [13] also introduced the idea of incremental grammar refactoring through the sequence of simple transformations deriving from application of refactoring operators, however no specific automated refactoring approach, such as mARTINICA [3] was introduced.

Lämmel and Zaytsev introduced suite of four refactoring operators, specifically aimed for tackling refactoring tasks occurring in the process of grammar extraction from multiple diverse sources of information [14].

IV. Background

This section discusses refactoring operators as a basis for understanding grammar refactoring patterns and the core idea of the approach. This section also briefly introduces a method of describing a context-free grammar properties through the formalism of an objective function, used as a specification of the refactoring objective.

A. Refactoring Operators

Formally, grammar refactoring operator is a function taking a context-free grammar $G = (N, T, R, S)$ and using it as a basis for creating new grammar $G' = (N', T', R', S')$ equivalent to G. At this stage of development, the experiments were performed on the basis of eight operators: Unfold, Fold, Remove, Pack, Extend, Reduce, Split and Nop. The first three have been adopted from Lämmel's paper on grammar adaptation [13], while the others are proposed by us [8].

Grammar refactoring patterns are proposed as an addition to the base of refactoring operators. However, in this context, the key difference between refactoring operators and patterns is that the growth in the number of patterns (in the base of operators) does not have significant negative impact on the algorithm complexity, and the opposite is often true. This is caused by their domain-specific orientation and quite narrow scope of refactoring tasks to which individual patterns are applicable.

B. Objective Function

We adopt a modified understanding and notation of objective functions from mathematical optimization. An objective function describes properties of a context-free grammar to be achieved by refactoring. However, it does not describe the way in which the this should be performed, and the condition in which desired context-free grammar properties are achieved.

In our view, the objective function consists of two parts: *objective* and *state function*. Our refactoring algorithm works with only two kinds of objectives, which are minimization and maximization of a state function. We define a state function as an arithmetic expression whose only variables are the grammar metrics [1] calculable for any context-free grammar. As such, a state function is a tool for qualitative comparison of two or more equivalent context-free grammars.

Exemplary objective function prescribing minimization objective under state function consisted of count of nonterminals (var) and count of production rules ($prod$) is (1).

$$f(G) = minimize\ 2 * var + prod \tag{1}$$

C. mARTINICA: Refactoring Algorithm

The main idea behind mARTINICA (Fig. 3) lies in applying a sequence of grammar refactoring operators to a context-free grammar, to produce an equivalent grammar with a lower value of the objective function if the objective is minimization, or a higher if the objective is maximization. On the other hand, pLERO allows specifying one operator of such a sequence.

Since mARTINICA is an evolutionary algorithm, it also requires other input parameters, in addition to the *initial grammar* and the *objective function*, in order to be executed. It requires three other input parameters: *number of evolution cycles*, *population size* and *length of a generation life*. The first two are typical for algorithms of a similar type, while the third parameter is our own.

As shown in Fig. 4, presenting a white-box view, the algorithm starts with creation of an initial population of grammars. Each population member is then created in the basis of the initial grammar, transformed by semi-random sequence of refactoring precesses. After the initial phase, the algorithm iterates for count of evolution cycles through:

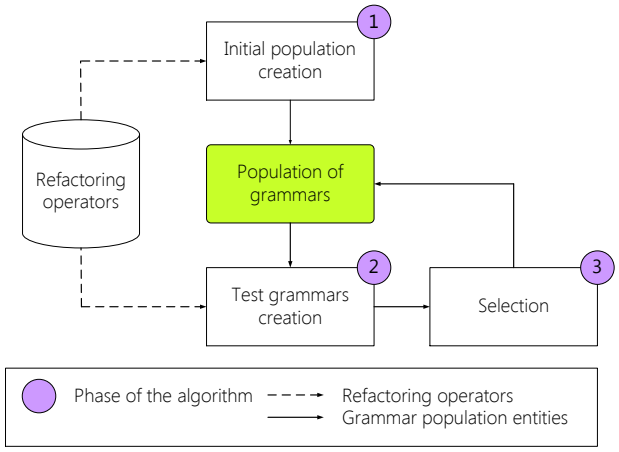


Fig. 4. White-box view of mARTINICA

1) *Test grammars creation* in which candidate population members are created. For each grammar in each generation three test grammars are created:
   a) Self-test grammar that attempts to redouble transformation that led to improvement in value of objective function in past generations of this population member.
   b) Foreign-test grammar that attempts to incorporate transformation that led to progress in value of objective function in past generations of some other population member.
   c) Random-test grammar that attempts to transform grammar towards optimization of value of objective function on the basis of random sequence of refactoring operators.
2) *Selection* in which population members are substituted by candidates with best value of objective function.

The resulting grammar reflects a population member of the last generation with the best value of the objective function.

Detailed description concerning proposal and implementation of mARTINICA algorithm along with some experimental results can be found in [3] [8].

V. Grammar Refactoring Patterns

In our view, each grammar refactoring pattern provides an equivalent transformation to context-free grammars. In this sense, the concept of grammar refactoring patterns is closely related to the concept of refactoring operators. However, there are several key differences between grammar refactoring patterns and refactoring operators.

First of all, refactoring operators provide problem-independent transformations, while grammar refactoring patterns provide problem-specific transformations. This means refactoring operators provide general transformations, with usage not bound by any specific class of refactoring tasks, while grammar refactoring patterns provide domain-specific transformations, intended for tackling the issues of particular class of refactoring problems.

Secondly, each of the refactoring operators can be applied to an arbitrary context-free grammar, including the situation of particular grammar form not allowing occurance of a specific transformation. In this case, the original grammar form is returned as a result of the transformation. On the other hand, each grammar refactoring pattern prescribes some specific preconditions a context-free grammar must fulfill in order to be transformable by a particular refactoring pattern.

In our approach, each pattern is represented as a specification consisting of a set of transformation rules, while transformation rule provides transformation on some subset of grammar's production rules that exhibit specific structural properties. In this notion of refactoring patterns, each instance of refactoring operator is actually a refactoring pattern which lacks explicit specification of required structural properties of grammar's production rules, and each refactoring pattern is in fact non-parametric refactoring operator.

VI. Core

For the purposes of patterns expression, we propose pLERO, pattern Language of Extended Refactoring Operators.

pLERO is currently being developed in two distinct dialects e.g. imperative [15] and functional. Refactoring patterns written in imperative dialect of pLERO are more process-centric, meaning that they are intended for specification of particular steps of a refactoring process, while refactoring patterns written in functional dialect are more result-centric and facilitate understanding of a grammar's structural changes. In this paper, we present the functional dialect of pLERO, while detailed description of the imperative dialect of pLERO can be found in [15].

A. pLERO

Through pLERO it is possible to define patterns for grammar refactoring or other transformations, applicable to grammars expressed in BNF. That is, pLERO is a language for defining parameterless operators of a problem class.

Pattern description consists of a set of transformation rules, while each rule comprises predicate describing the shape of a grammar's production rules, and transformation defining how production rules matched against predicate should be changed.

Predicate and transformation are expressed in similar fashion by formalism of meta-production rules. Each meta-production rule defines structure of some subset of a grammar's production rules. Predicate is specified by exactly one meta-production rule matched against grammar's production rules, while transformation is described by set of meta-production rules defining shape of production rules to be included in grammar during refactoring process.

Each meta-production rule can be divided in two parts, namely, left side of meta-production rule and right side of meta-production rule. Left side of meta-production rule specifies nonterminal on the left side of a grammar's production rule, while right side of meta-production rule specifies sequence of symbols that can be found on the right side of a grammar's production rules. Left side of meta-production rule

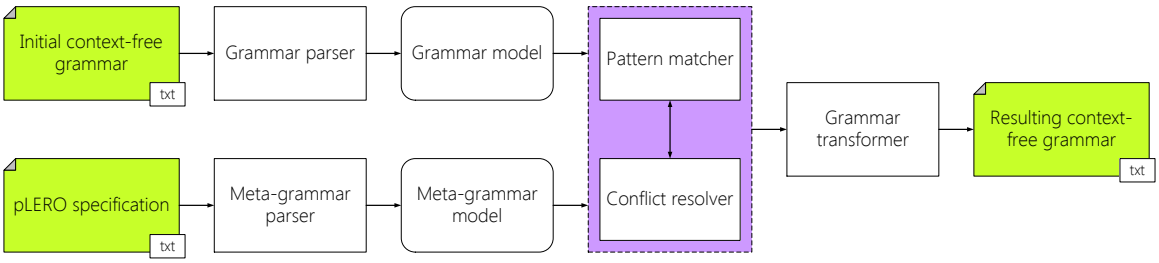


Fig. 5. Architecture of pattern application system

is some pattern variable, while right side of meta-production rule is concatenation of pattern variables.

Pattern variable specifies homogenous sequence of symbols in a grammar's production rules, consisting of variable prefix and name. Variable prefix describes possible matched symbols and their number, while variable name is identifier of this sequence. The prefix can be "t" for terminals, "n" for nonterminals, and "s" for both terminals and nonterminals. The letter specifying the symbol type can be followed by the asterisk "*" denoting the variable can match a sequence of symbols instead of a single symbol. For instance, the most generic variable type has prefix "s*" that can match any sequence of symbols. Variable prefix and name are separated by dot ".". After the dot, the variable name follows, e.g. "s*.symbols".

Pattern variable on the left side of meta-production rule may only have prefix "n" not followed by asterisk, denoting exactly one nonterminal. Each pattern variable on the right side of meta-production rule can have arbitrary valid prefix.

Each specification of refactoring pattern in pLERO must comply with the same template (Fig. 6) which allows specification of global pattern variables denoting same symbol sequences in all transformation rules of a pattern during entire refactoring process.

```
pattern [pattern_name] {
   variables:
        [prefix1].[variable_name1],
        [prefix2].[variable_name2],
           ...
        [prefixn].[variable_namen];
   new symbols:
        [prefixI].[variable_nameI],
        [prefixII].[variable_nameII],
           ...
        [prefixN].[variable_nameN];
   [transformation_rule1];
   [transformation_rule2];
      ...
   [transformation_rulem];
}
```

Fig. 6. Template of a pattern notation

The template also enables to specify new nonterminal symbols, which need to be generated for the use in production of new production rules. Notion of individual transformation rules must also follow specific template shown in Fig. 7.

```
predicate -> transformation
```

Fig. 7. Transformation rule decomposition

While variables may consist of all the possible prefixes, new variables may not; more specifically they cannot consist of partially deterministic constructs such as "*" or "s", and in current version of pLERO only "n" is allowed. Reason for this is that these constructs do not specify unambiguous concatenation of symbols and though it is not possible to generate definite sequence of symbols on their basis. Moreover new variables may be used only in meta-production rules contained within transformation part of transformation rule, and their use in predicate is prohibited. Reason for this is that new variables correspond with sequences of symbols that occur only in refactored grammar, and not in the original grammar.

1) Pattern Matching:

In order to apply transformation provided by refactoring pattern on some context-free grammar it is first necessary to match this grammar against this pattern. Process of pattern matching has two purposes:

- Determining if a grammar is transformable by a pattern
- Determining which pattern variable represents which sequence of symbols within production rules of a grammar

To each assignment of specific sequence of symbols to particular pattern variable we refer as to variable binding and to each variable representing definite sequence of symbols we refer as to bound variable.

Variables are bound during the matching of the rule and used in the replacement construction process. Global variables keep their value after they are bound during the first successful match. Other variables (to which we refer as to local variables) are bound only during the application of a rule and cleaned before the next matching.

The matching of a predicate against a grammar production is successful if all the pattern variables can be bound to a

part of the production and no unmatched symbols are left. Variables can match only some type of symbols, based on their prefix. Simple variables must match exactly one symbol of a specified type, while sequence variables can match any number of symbols (including zero).

For instance, the predicate "`n.1 ::= n.2 s*.1 t.1`" would match production "`A ::= B 'c' 'd' 'e'`", resulting in bindings "`n.1`" = "`A`", "`n.2`" = "`B`", "`s*.`" = "`'c' 'd'`", "`t.1`" = "`'e'`", and also "`B ::= D 'f'`", "`n.1`" = "`B`", "`n.2`" = "`D`", "`s*.`" is empty, "`t.1`" = "`'f'`". Since it does not start with a nonterminal, it would not match "`C ::= 'd' 'e' 'f'`".

If the pattern "`n.1 ::= s*.1 n.2 s*.2`" is matched against the production "`A ::= B C D`", the resulting binding would be "`n.1`" = "`A`", "`s*.1`" is empty, "`n.2`" = "`B`", and "`s*.2`" = "`C D`".

Variable prefix specifies only structure of some sequence of symbols, and it does not define particular symbols of a specific grammar. On the other hand, variable name is an identifier of a specific variable binding established during a particular pattern matching process. For instance, the predicate "`n.1 ::= n.1 n.2`" would match production "`A ::= A B`", however it would not match production "`A ::= B C`" since in this case variable `n.1` would be bound to two different nonterminals (`B` and `C`).

The matching of sequences is non-greedy. This means that short sequences are performed first during the matching. The process continues while the entire production is matched.

However, there are some cases in which conflicts in matching of predicate against production can arise, e.g. conflict always occurs if predicate contains two consecutive sequences of arbitrary symbols ("`s*.A s*.B`"). In this case, we have adopted first-match found resolution strategy.

2) Pattern Application:

Each transformation rule of refactoring pattern describes structure of some production rules and specifies new production rules that should replace this production rule. Predicate is a concatenation of pattern variables, which can match a sequence of production rule symbols and then represent these symbols in the transformation.

If a variable of the same type and name is present in a transformation rule, it will represent the same sequence of symbols in all its occurrences. In the transformation, meta-production rules have to consist only of the variables occurring on the predicate side of the transformation rule or in global variables. After the predicate matches any grammar production, its variables are bound to parts of the production and the replacement productions are constructed on the basis of transformation patterns.

If applied to a grammar, all transformation rules of a pattern are traversed in the order of their specification. Predicate is then matched against all the unprocessed productions of the original grammar. If the match is successful, replacement production is constructed and the production is replaced in the grammar.

Order of specification of transformation rules within a pattern is important, for it serves as conflict resolution mechanism in case when there are multiple predicates that can be matched against one production rule.

On the other hand, multiple production rules can be matched against one predicate, but only if all global variables of a predicate are bound to a same sequence of symbols in each production, and in that case replacement productions are constructed for each such rule.

B. Implementation

To be able to perform experiments and to demonstrate the correctness of the approach, automated pattern application system (Fig. 5) has been implemented, in which pLERO plays a central role.

The system takes the initial grammar and the pLERO pattern specification from the two different text files, and after the refactoring it creates new text file containing the resulting grammar.

The first text file is parsed by grammar parser which creates its representation in the form of grammar model, while the second is parsed by meta-grammar parser which creates meta-grammar model.

The core of the system is divided in two coexisting entities:

1) *Pattern matcher* – The purpose is matching of grammar model against meta-grammar model
2) *Grammar transformer* – The purpose is construction of replacement productions and generating of refactored grammar.

To resolve various conflicts occurring during the process of pattern matching, various resolution strategies are implemented in a separate module to which we refer to as a *conflict resolver*.

VII. Experimental Results

As an example, see Fig. 8 and 9 containing fragment of Algol 60 grammar [16] and pattern for immediate left-recursion removal (not direct). Then, Fig. 10 and 11 reflect equivalent grammar fragments produced after two sequential pattern applications.

After the first application of the pattern immediate left-recursion concerning nonterminal "`term`" was removed.

After the second application of the pattern immediate left-recursion concerning nonterminal "`factor`" was removed.

VIII. Conclusion

The most significant contribution, that we expect based on the results presented in this paper, is the contribution to automated grammar evolution. As such, our refactoring approach presents an appropriate basis for creation of new theory concerning automated task-driven grammar refactoring, while the provided experimental results as well as the other experiments [3] [8] explicitly demonstrate correctness and effectiveness of this approach.

However, achievement of this goal also requires deeper understanding and intensified research in refactoring operators,

```
term ::= factor
term ::= term multiplying_operator factor
multiplying_operator ::= 'x'
multiplying_operator ::= '/'
multiplying_operator ::= '÷'
factor ::= primary
factor ::= factor '↑' primary
primary ::= unsigned_number
primary ::= variable
primary ::= function_designator
primary ::= '('arithmetic_expression')'
```

Fig. 8. Fragment of Algol 60 grammar [16]

```
pattern LeftRecursionRemoval {
   variables: n.A;
   new symbols: n.A1;
   n.A ::= n.A s*.x ->
           n.A1 ::= , n.A1 ::= s*.x n.A1;
   n.A ::= s*.x ->
           n.A ::= s*.x n.A1;
}
```

Fig. 9. Example of a pattern for immediate left-recursion removal

```
term ::= factor N4
N4' ::=
N4' ::= multiplying_operator factor N4
multiplying_operator ::= 'x'
multiplying_operator ::= '/'
multiplying_operator ::= '÷'
factor ::= primary
factor ::= factor ↑ primary
primary ::= unsigned_number
primary ::= variable
primary ::= function_designator
primary ::= '('arithmetic_expression')'
```

Fig. 10. Resulting grammar after first application of refactoring pattern

as well as quality-based grammar metrics. Crucial part of this research are refactoring patterns, since they operate with knowledge derived from experience of language engineers, and thus they present an appropriate tool for converging of state-of-art and state-of-practice in the field of grammar refactoring.

In the future, we would like to focus on achieving greater abstraction power of the pLERO language, so it would formalize other knowledge considering refactoring problems and context of their occurrence, such as consequences of pattern's application on grammar's quality attributes. We would also like to adopt our approach to EBNF notation, which is structurally richer and would cause pattern matching to be more deterministic.

```
term ::= factor N4
N4 ::=
N4 ::= multiplying_operator factor N4
multiplying_operator ::= 'x'
multiplying_operator ::= '/'
multiplying_operator ::= '÷'
factor ::= primary N20
N20 ::=
N20 ::= '↑' primary N20
primary ::= unsigned_number
primary ::= variable
primary ::= function_designator
primary ::= '('arithmetic_expression')'
```

Fig. 11. Resulting grammar after second application of refactoring pattern

However, our vision goes even further, since mARTINICA and pLERO currently cover only one aspect of grammar adaptation, e.g. grammar refactoring, while the ultimate goal is to create universal approach covering other processes concerning grammarware engineering, e.g. grammar construction and destruction.

In case of interest, it is possible to download automated pattern application system from:

`http://plero.fei.tuke.sk`

ACKNOWLEDGMENT

This work was supported by project VEGA 1/0341/13 *Principles and methods of automated abstraction of computer languages and software development based on the semantic enrichment caused by communication*.

REFERENCES

[1] J. Cervelle, M. Crepinsek, R. Forax, T. Kosar, M. Mernik, and G. Roussel, "On defining quality based grammar metrics," in *Proceedings of International Multiconference (IMCSIT '09)*. Los Alamitos, USA: IEEE Computer Society Press, 2009, pp. 651–658.

[2] T. Mogensen, *Basics of Compiler Design*. Copenhagen, DK: University of Copenhagen, 2007.

[3] I. Halupka and J. Kollár, "Evolutionary algorithm for automated task-driven grammar refactoring," in *Proceedings of International Scientific Conference on Computer Science and Engineering (CSE'2012)*. Slovakia: Technical University of Košice, 2012, pp. 47–54.

[4] C. Alexander, *The Timeless Way of Building*. New York, USA: Oxford University Press, 1979.

[5] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, and M. Stal, *Pattern-Oriented Software Architecture Volume 1: A System of Patterns*. New York, USA: John Wiley & Sons, 1996.

[6] P. Klint, R. Lämmel, and C. Verhoef, "Toward an engineering discipline for grammarware," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 14, no. 3, pp. 331–380, 2005.

[7] T. Alves and J. Visser, "A case study in grammar engineering," in *Proceedings of 1st International Conference on Software Language Engineering (SLE' 2008)*. Berlin-Heidelberg: Springer-Verlag, 2008, pp. 285–304.

[8] I. Halupka, J. Kollár, and E. Pietriková, "A task-driven grammar refactoring algorithm," *Acta Polytechnica*, vol. 52, no. 5, pp. 51–57, 2012.

[9] N. Kraft, E. Duffy, and B. Malloy, "Grammar recovery from parse trees and metrics-guided grammar," *IEEE Transactions on Software Engineering*, vol. 35, no. 6, pp. 780–794, 2009.

[10] R. Lämmel and C. Verhoef, "Semi-automatic grammar recovery," *Software: Practice and Experience*, vol. 31, no. 15, pp. 1395–1438, 2001.

[11] W. Lohmann, G. Riedewald, and M. Stoy, "Semantics-preserving migration of semantic rules during left recursion removal in attribute grammars," *Electronic Notes in Theoretical Computer Science (ENTCS)*, vol. 110, pp. 133–148, 2004.

[12] K. Louden, *Compiler Construction: Principles and Practice*. Boston, USA: PWS Publishing, 1997.

[13] R. Lämmel, "Grammar adaptation," in *Proceedings of the International Symposium of Formal Methods Europe on Formal Methods for Increasing Software Productivity (FME '01)*. London, UK: Springer-Verlag, 2001, pp. 550–570.

[14] R. Lämmel and V. Zaytsev, "An introduction to grammar convergence," in *Proceedings of the 7th International Conference on Integrated Formal Methods*. London, UK: Springer-Verlag, 2009, pp. 246 – 260.

[15] J. Kollár and I. Halupka, "Role of patterns in automated task-driven grammar refactoring," in *2nd Symposium on Languages, Applications and Technologies (SLATE'13)*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013, pp. 171–186.

[16] R. L. Sites, *Algol-60 Version 5 Reference Manual*. Control Data Corporation (CDC), 1979. [Online]. Available: http://www.computinghistory.org.uk/det/7244/Algol-60-Version-5-Reference-Manual/

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1499–1506

Incremental JIT Compiler for Implicitly Parallel Functional Language

Petr Krajča

Dept. Computer Science, Palacky University, Olomouc
17. listopadu 12, CZ–77146 Olomouc, Czech Republic
petr.krajca@upol.cz

***Abstract*—We present a novel method for automatic parallelization of functional programs which combines interpretation and just-in-time compilation. We propose an execution model for a Lisp-based programming language which involves a runtime environment which is able to identify portions of code worth running in parallel and is able to spawn new threads of execution. Furthermore, in order to achieve better performance, runtime environment dynamically identifies expressions worth compiling and compiles them into a native code.**

I. Introduction

With advent of multi-core processor we are experiencing growing demand for programs which are able to utilize multiple processor cores to increase their efficiency. This trend is likely to continue as the hardware manufacturers have switched their focus from increasing performance by scaling clock speed to developing processors that are able to process multiple tasks simultaneously. Traditionally, programmer has to explicitly define which parts of a program will run in parallel and eventually has to use special means, e.g., locks, semaphores, critical sections, to synchronize parallel branches of execution in order to ensure correct outputs of the program. Currently, this approach to parallel programming is the prevailing one, on the other hand, it is also a very demanding approach from the programmer's point of view.

Another approach, we call it *implicit parallelism*, is to provide a language which is used the same way as any sequential language but its execution can be implicitly parallelized by a compiler or an interpreter of the language. Basically, compiler or runtime environment identifies portions of code which can be run in parallel and takes care of their proper synchronization, and thus, program returns always the same results, no matter which parts of the program were executed simultaneously. If this is achieved, the programmer can completely forget about the issues with parallel execution which is the desirable effect. Nowadays, implicit parallelism is relatively rare and has limited use.

In [14] we outline an evaluation model of the Lisp-based language called Schemik. The execution model of the language is based on a pushdown automaton which allows to describe operational semantics of the language in terms of transitions of the automaton. This model allows for further extensions. The most important extension detects expressions which are worth evaluating in a separate thread of execution and spawns new threads computing values of expressions in parallel. The model has shown its practical merits. (1) Implementation of the model shows that programs can be implicitly parallelized. (2) Using the automaton it can be proved that program returns always the same results no matter which parts of the program run in parallel. (3) One can incorporate software transactional memory to isolate side-effects performed in each thread of execution, hence efficiently parallelize even programs with side effects as discussed in [15]. On the other hand, the model itself is proposed for an interpreted language and this brings significant overhead. This paper presents a new extension for the evaluation model proposed in [14] which incorporates methods of just-in-time (JIT) compilation and preserves all mentioned features of the execution model, especially the ability to run programs in parallel without the need to use dedicated language constructs.

The paper is organized as follows. First, we provide a brief description of the language and of the formal model of the execution, including the methods which are used to automatically parallelize programs. Afterwards, we describe new extension which adds support for JIT compilation. The paper is concluded with a section focusing on implementation details and experimental evaluation.

Supported by grant no. 202/12/P167 of the Czech Science Foundation.

II. Interpreter

A. Evaluation Model: An Overview

In this section, we briefly describe a sequential evaluator which is essential for understanding our approach to parallel computation. We assume that readers are acquainted with some dialect of Lisp or Scheme, and therefore, in the following text we tacitly use the usual terminology which can be found in [4], [9], [12].

Our stack-based evaluator is a deterministic pushdown automaton with two stacks—*execution stack* (denoted E) containing *stack operations* controlling the evaluation, and *result stack* (denoted R) containing Schemik objects playing the roles of operands and intermediate results of operations.

The input (first expression to be evaluated) is encoded on the execution stack and the output (result of evaluation) is pushed on the result stack at the end of the computation. The content of both stacks is what completely describes the current state of the evaluator. Therefore, a setting of the stacks shall be called

978-83-60810-53-8/$25.00

a *configuration* of the evaluator. The result stack contains first-class objects of Schemik in their internal representation which are basically the same objects as in Scheme [12] (i.e., numbers, symbols, functions, special operators, etc). The execution stack can be seen as a stack for pending operations. Unlike the result stack, it contains stack operations which are not Schemik objects. The stack operations are represented by tuples of the form $\langle$operation-name, $arg, \mathcal{E}$, flag$\rangle$ where operation-name is a name for a step of evaluation (e.g., EVAL, FEVAL, FUNCALL, IF, INSPECT, etc.); *arg* is an object representing argument for stack operation; $\mathcal{E}$ is an environment (i.e., a table describing bindings of lexical variables) associated to the stack operation; flag is an indicator of the tail-recursion optimization [12]. For brevity, attributes *arg*, $\mathcal{E}$, and flag can be omitted from descriptions of a stack operations if they are not used in that particular operation.

During the computation, the automaton changes its configuration. A change from one configuration to another will be called a *transition*. Each transition is determined by stack operation which resides on the top of the execution stack. The computation halts if the execution stack is empty and the object on the top of the result stack is a *result* of the computation. Each transition of the automaton may be depicted by two pairs of stacks—configuration *before* and *after* the transition—and it may be written as follows:

E: *stack with operations before transition* ⟧
R: *stack with results before transition* ⟧
E: *stack with operations after transition* ⟧
R: *stack with results after transition* ⟧

The first pair of stacks represents the configuration *before* the transition. The second pair of stacks, drawn below the first one, represents the configuration *after* the transition. Bottom of all stacks is on the right and is denoted by symbol ⟧.

The *start configuration* of the automaton contains a single stack operation $\langle$EVAL, $expr, \mathcal{E}_t$, N$\rangle$ on the execution stack and an empty result stack, meaning that symbolic expression $expr$ will be evaluated in the top-level environment $\mathcal{E}_t$ (environment containing initial bindings of lexical variables), N says that *expr* does not appear in a tail position, see [12].

The operation $\langle$EVAL, $object, \mathcal{E}$, f$\rangle$ initiates evaluation of $object$ (e.g., a symbolic expression) in environment $\mathcal{E}$. If this operation is on the top of the execution stack, an appropriate transition is made depending on the type of *object*. Three situations may occur: (i) *object* is a self-evaluating object (i.e., neither a symbol nor a non-empty list), in which case *object* is pushed on the result stack; (ii) *object* is a symbol (name of a lexical variable), its value in environment $\mathcal{E}$ is pushed on the result stack; (iii) *object* is a list $(head\ arg_1 \cdots arg_n)$ then the top of the execution stack is replaced as follows:

E: $\langle$EVAL, $(head\ arg_1 \cdots arg_n), \mathcal{E}$, f$\rangle \cdots$ ⟧
R: $\cdots$ ⟧
E: $\langle$EVAL, $head, \mathcal{E}$, N$\rangle$,$\langle$INSPECT, $(arg_1 \cdots arg_n), \mathcal{E}$, f$\rangle \cdots$ ⟧
R: $\cdots$ ⟧

The previous transition has prepared the $head$ of the original list for evaluation. The role of INSPECT is to distinguish between different evaluation rules based on the value of the *head*, i.e. the first element in the list. The *head* may evaluate to a function, a special operator, or a macro. For instance, if INSPECT is on the top of the execution stack and the top of the result stack contains a *function*, INSPECT will prepare application of the function, i.e., it will prepare all arguments for evaluation and then it prepares a function call using FUNCALL:

E: $\langle$INSPECT, $(arg_1 \cdots arg_n), \mathcal{E}$, f$\rangle \cdots$ ⟧
R: $function \cdots$ ⟧
E: $\langle$EVAL, $arg_1, \mathcal{E}$, N$\rangle, \ldots, \langle$EVAL, $arg_n, \mathcal{E}$, N$\rangle$,
$\langle$FUNCALL, $n, \mathcal{E}$, f$\rangle \cdots$ ⟧
R: $function \cdots$ ⟧

The application of functions is performed by FUNCALL which is used in the form $\langle$FUNCALL, $n, \mathcal{E}$, f$\rangle$, where n is the number of arguments. Arguments and the function itself are stored on the result stack. In general, $\langle$FUNCALL, $n, \mathcal{E}$, f$\rangle$ takes $n+1$ objects from the result stack where first n objects represent arguments and the last object taken is the desired function. If the function is a primitive function (i.e., a built-in function), the arguments are passed to the function and the result is pushed on the result stack. If the function is a user-defined function, a body of the function is evaluated in a local environment where formal parameters of the function are bound to arguments obtained from the stack. Therefore, if the body of the function is B, the corresponding transition is the following:

E: $\langle$FUNCALL, $n, \mathcal{E}$, f$\rangle \cdots$ ⟧
R: arg_1, $arg_2, \ldots,$ arg_n, *user-defined fn. with body* $B \cdots$ ⟧
E: $\langle$EVAL, $B, \mathcal{E}'$, T$\rangle \cdots$ ⟧
R: $\cdots$ ⟧

Note that $\mathcal{E}'$ denotes the local environment. According to the value of the flag f, $\mathcal{E}'$ is a fresh new environment (if f equals N) or a reused old environment (if f equals T), which corresponds to a proper tail call, see [12]. For each of the operators, INSPECT has a separate rule of evaluation. For instance, for the special operator `if` the INSPECT operation enforces that only one branch of execution is performed as follows.

Typically, the operator `if` is used in the form: `(if` *cond then-branch else-branch*`)`. If *cond* evaluates to anything but `#f`, the special form returns the value of *then-branch*; otherwise the value of *else-branch* is returned or `#void` is returned in case *else-branch* is not present. Since the result is given by the value of *cond*, INSPECT places the two branches on the result stack, evaluates condition with the operation EVAL, and a new stack operation IF is used to select branch for further evaluation according to the value of *cond*:

E: $\langle$INSPECT, $(cond$ *then-branch else-branch*$), \mathcal{E}$, flag$\rangle \cdots$ ⟧
R: *special operator* `if` $\cdots$ ⟧
E: $\langle$EVAL, $cond, \mathcal{E}$, N$\rangle$,$\langle$IF, $\mathcal{E}$, flag$\rangle \cdots$ ⟧
R: *then-branch*, *else-branch* $\cdots$ ⟧

Now it remains to clarify the newly introduced stack operation IF. The stack operation pops three values from the result stack—an object O representing the value of the *cond*, *then-branch*, and *else-branch*. If O is not equal to `#f`, *then-branch* is pushed back to the execution stack for evaluation:

E: $\langle \mathsf{IF}, \mathcal{E}, \mathsf{flag} \rangle \cdots]\!]$
R: `#f`, *then-branch*, *else-branch* $\cdots]\!]$
E: $\langle \mathsf{EVAL}, \textit{else-branch}, \mathcal{E}, \mathsf{flag} \rangle \cdots]\!]$
R: $\cdots]\!]$

Otherwise, *else-branch* is pushed back to the execution stack for evaluation.

B. Parallel Evaluation Model

In this section, we introduce the implicit parallel evaluation model which is an extension of the evaluation model proposed in the previous section. We outline only the parallel evaluator of a purely functional subset of the language, i.e., subset of Schemik without assignment operations and mutators of pairs (`set!`, `set-car!`, `set-cdr!`), I/O functions (`display`, etc.), and continuations (`call/cc`). Description of the language in its full extent can be found in [16].

Our parallel evaluator of purely functional subset of the language is based on the idea that each operation EVAL may be processed in an independent evaluator as far as the resulting value is dependent only on given arguments and environment. We introduce a new stack operation FEVAL (abbreviation for "fork eval") representing evaluation in an *independent evaluator*, i.e., a thread. Basically, all the independent evaluators form a tree structure with a single (main) evaluator at the top of the hierarchy. Each evaluator has its own independent pair of stacks and environments are shared among all evaluators. We can share environments without any inconsistency issues because we are considering programs without side effects only. This means, once an environment is established, it does not change bindings of symbols (lexical variables).

C. Stack Operation FEVAL

Operation FEVAL is similar to EVAL; it prepares an expression for evaluation in an independent evaluator, i.e., using a new pair of stacks. Operation FEVAL is specific in that it is not a result of any transition caused by another operation. The appearance of FEVAL on any execution stack is caused by external entity called *scheduler* which converts a suitable EVAL operation on an execution stacks to the FEVAL operation and creates a new evaluator containing the original EVAL on top of its execution stack. This way a parallel branch of the evaluation is created. The transition created by scheduler may be depicted as follows:

EV_1: E: $\cdots \langle \mathsf{EVAL}, \textit{object}, \mathcal{E}, \mathsf{f} \rangle \cdots]\!]$
R: $\cdots]\!]$
$\vdots$
EV_1: E: $\cdots \langle \mathsf{FEVAL}, \langle EV_2, \textit{object} \rangle, \mathcal{E}, \mathsf{f} \rangle \cdots]\!]$
R: $\cdots]\!]$
EV_2: E: $\langle \mathsf{EVAL}, \textit{object}, \mathcal{E}, \mathsf{N} \rangle]\!]$
R: $]\!]$

An invocation of FEVAL represents merging of two branches of the evaluation. If $\langle \mathsf{FEVAL}, \langle EV_i, \textit{object} \rangle, \mathcal{E}, \mathsf{f} \rangle$ appears on the top of the execution stack, the evaluation in the referred evaluator EV_i is stopped and its stacks are appended to the corresponding stacks of the current evaluator processing the FEVAL operation.

This behavior of FEVAL is correct because the referenced evaluator EV_i has performed right the same transitions that should be otherwise performed by the evaluator containing the FEVAL operation. We can consider the creation of this operation as a request for parallel evaluation to other evaluator and processing of FEVAL operation as a request for obtaining (partial) results from the referenced evaluator.

D. Scheduling Strategies

An important issue in implicit parallel evaluation is how to decide which expression is worth evaluating in a parallel branch. In theory, an arbitrary expression may be parallelized but it is desirable to focus on particular expressions that require larger amounts of transitions.

The decision on which expression will be evaluated in a parallel branch is done by the *scheduler*, associated to each evaluator, regularly seeking through the execution stack for operations EVAL that may be converted into the FEVAL operations. A stack operation $\langle \mathsf{EVAL}, \textit{object}, \mathcal{E}, \mathsf{f} \rangle$ is considered as a good candidate for parallel evaluation if $\textit{object}$ is a list or a list fulfilling further criteria. For instance, $\textit{object}$ should be a list containing other list, or a list with a user-defined function as its first element.

In particular, we use strategy assigning limited number of independent evaluators to candidate expressions which has its origin in the organization of the execution stack. This strategy assumes that more complex expressions are at the bottom of the stack and their evaluation will need more transitions, and thus, their evaluation in a parallel branch will be more beneficial. Therefore, scheduler walks through the execution stack from the bottom and looks for candidate expressions.

III. Just-in-Time Compilation

The ability of the execution model to automatically parallelize programs comes from the fact that one can easily determine the structure of the program execution by inspection of the automaton. Particularly, by inspection of the *execution stack*. However, if the program, or its portion, is compiled, we are losing the ability to analyze the program structure since scheduler is unable to identify expression for parallelization. In order to combine JIT compilation with the ability to parallelize programs, we propose an extension which compiles into a native code only expressions which are insignificant from the point of parallelization, and thus, compiler does not interfere with the mechanism for automatic parallelization. Such typical insignificant expression is an atom (symbol or self-evaluating object) or function call where arguments are atomic expressions. Apparently, their evaluation is very fast and their evaluation in parallel branches would lead to fine-grained parallelism and to a related overhead caused by

scheduling and spawning too many threads of execution. These expressions which are insignificant from the point of view of parallelization and which are, on the other hand, worth compiling shall be called *compilable*. Formally, an expression E is *compilable* if it satisfies one of the following conditions:

(1) Expression E is either atom (i.e., symbol, number, string, etc.),
(2) or E is an expression of a form $(E_1\ E_2\ \ldots E_n)$ where E_1 is a primitive function or a special operator and $E_2, \ldots, E_n$ are all compilable expressions.

Remark 3.1: Notice that the definition of compilable expression is recursive. Later, we use this fact to incrementally compile complex expression using the less complex ones.

Remark 3.2: It might be impossible to decide whether expression is compilable or not, strictly speaking, generally, it is an undecidable problem. Therefore, in a case when we are unable to determine if an expression is compilable, we assume that the expression is not compilable at all. When determining compilability of a non-atomic expression, it is crucial to determine the type of the first expression. For this purpose we use the following approximation. If the first expression E_1 is a symbol defined only in the top-level environment (w.r.t. the environment in which E_1 is evaluated) and if a primitive function or a special operator is bounded to this symbol, we assume that E_1 evaluates to a primitive function or a special operator. Note that this condition can be checked in a constant time and covers among others common function calls where symbol is used to directly represent a primitive function. For instance, `(+ a 1)` is an example of such expression.

A. Compilation: An Overview

The compilation process consists of several independent steps.

Before the program execution starts, all lists in the source code consisting solely of atoms are marked with a flag indicating that the given expression is a good candidate for compilation. For this purpose all source code expressions are equipped with a data structure similar to property-lists known from Common Lisp [9] which contains flags and further information produced by compiler, e.g., native code. Notice that this flag does not mean that expression is compilable, rather it is an indicator that compiler should try to compile this expression.

Every time the operation EVAL is on the top of the execution stack, its argument (evaluated expression) is checked if it has associated native code generated by the compiler, if so, the given code is executed and returned value is pushed on the result stack as if the evaluation was performed in a usual way. Otherwise, if the given expression is marked as a good candidate for a compilation, it is inserted into a queue of expressions waiting for compilation. In any case, if the expression has not attached any native code, it is evaluated in a standard way as described in Section II.

All expressions waiting for compilation in a queue are processed one by one by a compiler. For each expression, compiler checks whether is compilable, and if not, flag representing that an expression is a good candidate for compilation is removed. Otherwise, expression is compiled and the results of the compilation process (native code and intermediate representation of the code) are attached to the expression. Furthermore, expression containing this expression is marked as a good candidate for compilation.

Remark 3.3: Marking a parent expression as a candidate expression for compilation is a necessary step allowing us to incrementally compile complex expressions. For example, initially entire expression `(+ a (+ b 1))` is not marked as a good candidate for compilation since the second argument is not an atom and one can not decide if `(+ b 1)` is compilable. However, subexpression `(+ b 1)` is marked as a good candidate. If this subexpression is compiled, an entire expression `(+ a (+ b 1))` becomes also a good candidate and may be compiled into a native code. This corresponds with the recursive definition of *compilable* expression.

The key part of the compilation process is a transformation of a given expression into a native code, this transformation and related implementation aspects are discussed in the next section.

B. Compilation: Code Generation

The main task of the compiler is to take an expression and transform it into a native code which under the same conditions evaluates to same value as it would be evaluated usually. For instance, having an expression `(+ a 1)` we need to generate function with a prototype like this:

```
value *plus_a_1(environment *)
```

which gets value of `a` in a given environment and returns its value increased by one.

The process of expression compilation we propose can be divided into three steps. At first, expression is compiled into a high-level intermediate representation (HIR), this representation by its nature is very close to three-address code used by many compilers. In the next step optimization techniques are applied. Afterwards, HIR is transformed into a native code. Optionally, this step may consist of additional transformation. For instance, we transform HIR into a corresponding low-level intermediate representation (LIR) which is used to perform low-level optimizations, and after that, the native code is generated.

The high-level intermediate representation we propose consists of operations having up to three arguments. These arguments may be constants (regular Schemik objects), registers playing the role of (temporary) variables, or lists of operations of HIR playing the role of code blocks. We assume that registers are enumerable and shall be denoted R_i where i is an integer. List of essential operations used by HIR along with their semantics is presented in Table I.

The recursive way we define compilable expressions has been reflected in a way we compile expressions into a HIR. Entire process of transformation of the expression to HIR is described by the recursive procedure COMPILEHIR procedure

TABLE I
LIST OF OPERATIONS USED IN HIGH-LEVEL INTERMEDIATE REPRESENTATION

operation	meaning
`set` R_i, *source*	assigns value of *source* to the register R_i; *source* may be either a constant or register
`eval-symbol` R_i, S	evaluates symbol S w.r.t. current environment and stores the result into the register R_i
`prepare` *count*	an auxiliary operation introducing each function call, its purpose is to declare number of arguments
`putarg` i, *source*	passes value of *source* as the i-th argument of the function; *source* may be either a constant or register
`funcall` R_i, *fun*	calls function *fun* and stores the returned value into the register R_i
`add` R_i, R_j, *value*	adds R_j and *value*, which may be a constant or register, and stores the result into R_i; note that there are also operations for subtraction, multiplication, comparisons, etc.
`car` R_i, R_j	extracts the *car*-part of a dotted pair stored in the register R_j and stores the result into the register R_i
`cdr` R_i, R_j	analogous operation to `car` extracting the *cdr*-part of a dotted pair
`if` R_i, BRANCH$_1$, BRANCH$_2$	if value in the register R_i is other then `#f`, code in BRANCH-1 is performed, otherwise, BRANCH-2 is executed; note that BRANCH-1, BRANCH-2 represents an entire list of operations in HIR, i.e., correspond to code blocks
`rslt-push` R_i	pushes value in the register R_i on top of the result stack
`exct-push` *expression*	pushes value of *expression* on the execution stack; *expression* may be either a constant or register

which is presented in Algorithm 1. This procedure has two arguments—the given expression E and register number $base$. This number identifies the first register that can be used to compute the value of the expression E. The crucial feature of this procedure is that COMPILEHIR does not generate code directly, but returns a procedure which generates the final HIR for the expression. This procedure has one argument i and allows to shift used registers by offset i. In other words, COMPILEHIR does not generate final code, but rather a template of the code. This behavior is illustrated in Figure 1 (top) containing the result of COMPILEHIR procedure for expression `(foo a 1)` and argument $base$ equal to 10. As one can see this "template" is very close to three address code and the main difference is that registers does not have fixed numbers and may be shifted by offset i. Nonetheless, without loss of generality we may apply optimization techniques proposed for three-address code, e.g., constant and copy propagation, or dead-code elimination.

We now turn our attention to the HIR(i) procedure which is a result of the COMPILEHIR(E, $base$) procedure, as described in Algorithm 1. We assume that the generated code satisfies

Algorithm 1 Procedure COMPILEHIR(E, $base$)

```
1:  return procedure HIR(i) such that:
2:  if E is a constant (e.g., number) then
3:      emit operation set R_{base+i}, E
4:  if E is a symbol then
5:      emit operation eval-symbol R_{base+i}, E
6:  if E has attached HIR code then
7:      invoke HIR(base + i)
8:  if E is an expression (fun E_2 E_3 ... E_n) where fun
    is a primitive function then
9:      for all E_j where j ∈ {2, ..., n} do
10:         invoke COMPILEHIR(E_j, base + i + j − 1)
11:     if operation exct-push was emitted then
12:         abort compilation
13:     if fun is primitive function + then
14:         invoke COMPILEADDITION(base + i, n)
15:     else if fun is primitive function car then
16:         emit operation car R_{base+i}, R_{base+i+1}
17:     else
18:         invoke COMPILEFUNCALL(base + i, n, fun)
19: if E is expression (if E_cond E_1 E_2) and E_cond is
    compilable then
20:     invoke COMPILEIF(base + i, E_cond, E_1, E_2)
21: if E is quotation (quote val) then
22:     emit operation set R_{base+i}, val
23: otherwise abort compilation
```

the following conditions:

(i) The result of the computation is stored in the register R_{base+i} or generated code modifies content of the evaluator's stacks directly and in that case no value is returned.
(ii) No register with an index lesser than $base + i$ will be used to compute the result.

If the expression is an atom (see Algorithm 1, lines 2–5), procedure HIR generates operation which gets the value of the atom and stores the result into the register R_{base+i}. In case we encounter an expression which has attached code in HIR, see lines 6–7, we use the fact that intermediate code is already available and we use it to compute value of the expression. Notice that HIR is in fact a procedure and is invoked with an argument $base+i$. This ensures that once generated HIR code may use different set of registers which does not interfere with the registers already used. Furthermore, from the assumption (i) follows that the result, if returned, will be in the register R_{base+i}.

Code generation for non-atomic expressions consists of multiple sub-actions, see lines 8–22. Initially, the value of the first element is determined, i.e., the called function is determined. Subsequently, code that computes values of all arguments is generated, assuming that all arguments are evaluated in the left-to-right order and arguments are stored in registers $R_{base+i+1}$, $R_{base+i+2}$, …, $R_{base+i+n}$. To generate code evaluating arguments, we recursively invoke the COM-

Procedure HIR(i):
emit operations:
`eval-symbol` R_{11+i}, a
`set` R_{12+i}, 1
`prepare` 2
`putarg` 1, R_{11+i}
`putarg` 2, R_{12+i}
`funcall` R_{10+i}, *foo*

Procedure HIR(i):
emit operations:
`eval-symbol` R_{11+i}, a
`if` R_{11+i}, {
`set` R_{11+i}, 0
`rslt-push` R_{11+i}
}, {
`exct-push (foo b)`
}

Fig. 1. HIR of `(foo a 1)` (top) and HIR of `(if a 0 (foo b))` (bottom)

PILEHIR procedure at lines 9–10. Due to the assumption (ii), it is guaranteed that the registers used to compute arguments will not overlap in an undesirable way. This means, values of already computed arguments which are stored in the registers will not be overwritten during the computation of remaining arguments. Note that if any argument is compiled expression which does not return a value and rather directly manipulates with stacks, it is a sign that we can not obtain a value of this argument and entire expression is not compilable and compilation process has to be aborted, see lines 11–12. Situation when the generated code directly tempers with the execution stack is discussed later.

If the called function is a primitive function and if the compiler has mean to compile this function into a native code, appropriate code is generated. This is the case of frequently used functions, for instance, function for addition, subtraction, `car`, `cdr`, etc. Procedure presented in Algorithm 1 shows how this rule applies for addition and function `car`, see lines 13–16 and related Algorithm 2. Otherwise, code for generic function call is generated, see line 18 in Algorithm 1 and Algorithm 3.

Remark 3.4: In case of frequently called functions (e.g., addition) it may be efficient to generate code directly instead of performing generic function call, since calling a function in native code has overhead related to argument passing.

Compilation of special operators requires different approach. In this paper we discuss only two operators—`if` and `quote` since our experiments have shown, these operators affect the program performance the most. Compilation of the `(quote val)` expression is straightforward, it suffices to assign value of `val` to the register R_{base+i}. In case of the `if` operator is situation much more complex.

Let us consider the following two snippets of source code `(if (> a 0) 1 (+ a 1))` and `(if (> a 0) 1 (foo a))` and let us assume that `(> a 0)`, `(+ a 1)` are compilable expressions and

Algorithm 2 Procedure COMPILEADDITION(i, n)

1: **emit** operations:
`set` R_i, 0
`add` R_i, R_i, R_{i+1}
...
`add` R_i, R_i, R_{i+n}

`(foo a)` is not compilable. In the first case we have an expression which can be safely compiled because all subexpressions are compilable. In the second case we have limited information on `(foo a)` and we have to assume that this expression may perform any sequence of operations, and thus, is significant from the parallelization point of view. In conclusion, we may assume that only `(> a 0)` is compilable. Unfortunately, the second type of conditions appears in real-world programs more often than the first one, hence we have to take this fact into account.

The compilation process of conditional expression is presented in the COMPILEIF procedure (described in Algorithm 4) which is invoked from COMPILEHIR. While compiling conditional expression, we always assume that the condition is compilable expression and its HIR is available, see lines 1–4 in Algorithm 4. For expressions representing branches two situations may occur. (1) Both branches are compilable and HIR is available, (2) or at least one expression is not compilable or HIR is not available for this expression. In the first case, we generate a code which computes value of the condition and if the value is true or false, it computes value of the first or the second branch, respectively, and stores the result into a given register, see lines 5–10 and 15 in Algorithm 4. Notice that this case corresponds to the first snippet of the code discussed in the previous paragraph. In the second case, which corresponds to the second code snippet from the previous paragraph, we are unable to determine the value of the expression entirely and at some point we have to pass control back to the evaluator. Basically, generated code has to compute value of the condition and accordingly perform one branch of the execution. Code generated for each branch depends on the nature of the expression in a given branch. If HIR of the expression is available, this code is returned along with the operation `rslt-push` which pushes result on the *result stack* and passes execution back to the evaluator, see lines 11–14 in Algorithm 4. Otherwise, operation `exct-push` is performed, this operation pushes given expression on the execution stack, in other words, it evaluates expression by means of the evaluator. An illustrative example of compiled conditional expression `(if a 0 (foo b))`, which is generated by the COMPILEIF procedure, is presented in Figure 1 (bottom).

C. Native Code Generation

Using the COMPILEHIR procedure we have obtained high-level intermediate representation of the expression. This representation by its nature allows multiple optimizations, especially, those based on data-flow analysis. Since the HIR is very

Algorithm 3 Procedure COMPILEFUNCALL(i, n, *fun*)

emit operations:
`prepare` n
`putarg` 1, R_{i+1}
`putarg` 2, R_{i+2}
...
`putarg` n, R_{i+n}
`funcall` R_i, *fun*

Algorithm 4 Procedure COMPILEIF(i, E_{cond}, E_1, E_2)

if E_{cond} has attached HIR code without the `exct-push` operation **then**
    invoke HIR(i)
else
    abort compilation
for all E_j where $j \in \{1, 2\}$ **do**
    // create code block BRANCH$_j$ *such that:*
    if E_j has attached HIR code **then**
        BRANCH$_j$ $\leftarrow$ HIR(i)
    else
        BRANCH$_j$ $\leftarrow$ operation `exct-push` E_j
if E_1 has attached HIR code **and** BRANCH$_2$ contains `exct-push` **then**
    append to BRANCH$_1$ operation `rslt-push` R_1.
if E_2 has attached HIR code **and** BRANCH$_1$ contains `exct-push` **then**
    append to BRANCH$_2$ operation `rslt-push` R_2
emit operation `if` R_i, BRANCH$_1$, BRANCH$_2$

close to three-address code, we may directly apply optimization techniques. Namely we use copy propagation, constant propagation, constant folding, and dead code elimination. All these techniques significantly simplify the resulting code. These techniques are well-known and thoroughly studied, for instance, in [18] or [2], therefore we omit discussion on this topic and refer readers to these books.

Apparently, each operation of the HIR may be transformed into an assembly language. Usually, it requires less than ten instructions of assembly language to express these operations, because operations of HIR are in fact very simple. However, in order to increase portability of the compiler, our implementation converts HIR into a low-level intermediate representation (LIR) which is afterwards compiled to a native code, by a library. Our implementation is based on the MyJIT library [1], therefore our low-level intermediate representation is equivalent to the instruction set of this library. This library generates machine code for numeric operations, conditional and unconditional jumps, function calls, memory access operations, and takes care of register allocation. Since the LIR is tightly coupled with implementation aspects of the interpreter/compiler, we do not describe transformation from the HIR to the LIR and code generation in this paper. Nonetheless, it could be possible to use any other similar library to generate native code, for instance, LLVM.

Major part of our JIT compiler is written in Schemik itself, i.e., the JIT compiler is self-hosted and its code is a part of the standard library and all intermediate representations are objects (lists) in Schemik language as well. This way we benefit from the expressiveness of the Schemik language, especially, from the ability to conveniently transform one list of values to another. Every time the evaluator encounters compilable expression, it invokes the `jit:compile` function which is an ordinary function written in Schemik taking care of the transformation of an expression into the HIR and, subsequently, into a native code. For this purpose, Schemik has bindings to the MyJIT library. Function `jit:compile` returns representation of the expression in the HIR and a native code and an evaluator assigns these results to a given expression. The `jit:compile` function is always evaluated in a separate thread, however, it is evaluated with a regular evaluator as discussed in Sections II and III, hence benefits from the parallel evaluation of expressions and even from the JIT compilation of expressions. Since the compilation of the user supplied source code is mixed together with the compilation of the compiler itself, it usually takes some minimal amount of time for compiler to take effect, because compiler tends to compile itself first. However, if compiled version of an expression is not available, regular evaluation process is used, hence compilation has minimal impact on the execution time, if compared with the regular evaluator.

IV. EXPERIMENTS

In order to evaluate impact of the JIT compiler we performed set of experiments on a computer equipped with two Intel Xeons E5642 (eight CPU-cores in total) using standard benchmarks. Results of these experiments are presented in Table II containing times needed to compute given task for various configurations of the interpreter. First two columns represent results for an interpreter/compiler running only with one thread. The next two columns represent results when interpreter/compiler was allowed to use up to eight threads to compute results. The last column shows the results for the Guile Scheme (1.8.8). Experiments shows typical performance improvement by factor 2 or 3 depending on the type of benchmark and data size. It also shows that the results are fully comparable with Guile.

Important feature of parallel systems is their scalability. Our experiments show that the compiler slightly decreases scalability, see Figure 2. However, this is mostly only an effect of the Amdahl's law. Nevertheless, this shortcoming is adequately compensated since the compiler significantly improves performance, as it is apparent from Table II.

V. RELATED WORKS

Basically, two approaches to impose implicit parallelism into programming languages can be distinguished. The first is static code analysis which hard-wires parallel execution into a program during the compilation. This technique proved to be useful and has become part of the commercial compilers.

TABLE II
RESULTS OF BENCHMARKS (IN SECONDS) FOR VARIOUS CONFIGURATION OF THE EVALUATOR

	1 thread		8 threads		
	No JIT	JIT	No JIT	JIT	Guile
bubblesort	5.77	3.27	5.81	3.29	1.35
combinations	2.74	1.62	1.33	0.94	1.84
cpstak	8.77	5.19	2.08	1.40	2.79
fib30	1.23	0.49	0.43	0.30	0.31
fib33	5.24	2.03	1.69	0.87	1.27
fib35	13.62	5.19	4.49	2.42	3.32
mazefun	7.62	3.55	1.69	1.15	2.24
mergesort	6.53	3.70	2.99	1.69	0.14
nqueens	3.68	1.80	1.30	1.08	0.64
powerset	2.41	1.53	1.78	0.95	1.97
primes	8.65	3.63	4.11	3.48	1.86
quicksort	6.33	2.24	3.79	1.60	3.70
sum	8.08	2.78	3.33	2.84	2.31
tak	5.72	1.49	1.31	0.81	1.73

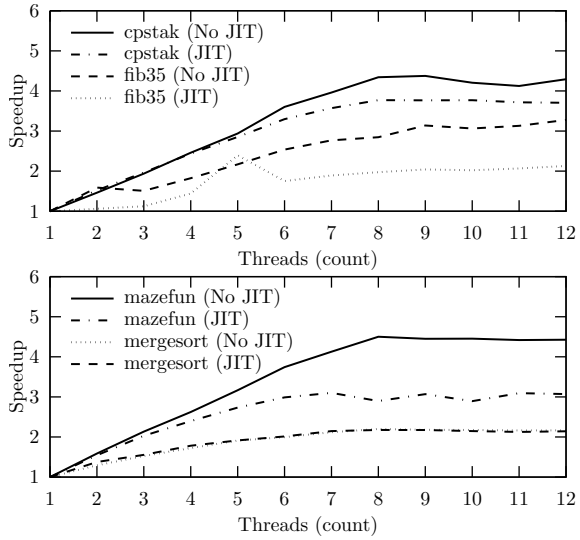


Fig. 2. Scalability of standard benchmarks

However, static code analysis usually focus only on loop parallelization, see [5], [13], what limits its use.

The second approach utilizes run-time analysis and dynamic parallelization. From this category most influential programming languages are Fortress [3] and Concurrent Haskell [19], [6]. However their execution model is tied to specifics of given languages and therefore it is hardly possible to transfer their model into other languages, e.g., into Scheme. Idea to parallelize Lisp-programs is not novel and many approaches to parallel execution of programs in Lisp [17] and Scheme [8], [10] appeared. Nevertheless, majority of this approaches were focusing on explicit parallelism, survey of these approaches can be found in [11]. Relatively, new approach to automatic parallelization of Scheme programs is proposed in [7] which is an interpretation model based on a speculative execution model. To the best our knowledge there is no other execution model like ours which combines run-time analysis and dynamic parallelization along with compilation.

VI. CONCLUSIONS

We have proposed an extension of the execution model for an implicitly parallel programing language. Our extension introduces means of JIT compilation into the execution model which was originally intended for interpreted programming language. Our experiments show that JIT compiler can efficiently decrease execution time of programs while preserving all features of the underlying model language, especially the ability to automatically parallelize programs. In future we intend to focus on further optimizations which are able to decrease program execution time even more, e.g., function inlining or specialization.

REFERENCES

[1] Myjit. http://myjit.sourceforge.net.
[2] Alfred V. Aho, Monica S. Lam, Ravi Sethi, and Jeffrey D. Ullman. *Compilers: Principles, Techniques, and Tools (2nd Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2006.
[3] Eric Allen, David Chase, Joe Hallett, Victor Luchangco, Jan-Willem Maessen, Sukyoung Ryu, Guy L. Steele Jr., and Sam Tobin-Hochstadt. The Fortress Language Specification Version 1.0, March 2008.
[4] John Allen. *Anatomy of LISP*. McGraw-Hill, Inc., New York, NY, USA, 1978.
[5] Pierre Boulet, Alain Darte, Georges-Andr Silber, and Frdric Vivien. Loop parallelization algorithms: From parallelism extraction to code generation. *Parallel Computing*, pages 421–444, 1998.
[6] Tim Harris, Simon Marlow, Simon L. Peyton Jones, and Maurice Herlihy. Composable memory transactions. In Keshav Pingali, Katherine A. Yelick, and Andrew S. Grimshaw, editors, *Proceedings of the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (10th PPOPP'2005), ACM SIGPLAN Notices*, pages 48–60, Chicago, IL, USA, June 2005.
[7] Charlotte Herzeel and Pascal Costanza. Dynamic parallelization of recursive code: part 1: managing control flow interactions with the continuator. In *Proceedings of the ACM international conference on Object oriented programming systems languages and applications*, OOPSLA '10, pages 377–396, New York, NY, USA, 2010. ACM.
[8] Guy L. Steele Jr. Lambda: The Ultimate Declarative. AI Memo 379, MIT AI laboratory, November 1976.
[9] Guy L. Steele Jr. *Common LISP the Language*. Digital Press, 2nd edition, 1990.
[10] Guy L. Steele Jr. and Gerald J. Sussman. Lambda: The Ultimate Imperative. AI Memo 353, MIT AI laboratory, March 1976.
[11] Robert H. Halstead Jr. New Ideas in Parallel Lisp: Language Design, Implementation, and Programming Tools. In Takayasu Ito and Robert H. Halstead Jr., editors, *Workshop on Parallel Lisp*, volume 441 of *Lecture Notes in Computer Science*, pages 2–57. Springer, 1989.
[12] Richard Kelsey, William Clinger, and Jonathan Rees (Editors). Revised^5 Report on the Algorithmic Language Scheme. *ACM SIGPLAN Notices*, 33(9):26–76, 1998.
[13] Ken Kennedy and John R. Allen. *Optimizing compilers for modern architectures: a dependence-based approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
[14] Petr Krajca and Vilém Vychodil. Data parallel dialect of scheme: outline of the formal model, implementation, performance. In Sung Y. Shin and Sascha Ossowski, editors, *SAC*, pages 1938–1939. ACM, 2009.
[15] Petr Krajca and Vilém Vychodil. Software transactional memory for implicitly parallel functional language. In Sung Y. Shin, Sascha Ossowski, Michael Schumacher, Mathew J. Palakal, and Chih-Cheng Hung, editors, *SAC*, pages 2123–2130. ACM, 2010.
[16] Petr Krajca and Vilém Vychodil. Stack-based model of implicit parallel execution of functional programs. In preparation.
[17] John McCarthy. Recursive functions of symbolic expressions and their computation by machine, Part I. *Communications of the ACM*, 3(4):184–195, 1960.
[18] Steven S. Muchnick. *Advanced Compiler Design and Implementation*. Morgan Kaufmann, 1997.
[19] Simon Peyton Jones, Andrew Gordon, and Sigbjorn Finne. Concurrent haskell. In *Proceedings of the 23rd ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, POPL '96, pages 295–308, New York, NY, USA, 1996. ACM.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1507–1514

Reconstruction of Instruction Idioms in a Retargetable Decompiler

Jakub Křoustek, Fridolín Pokorný
Faculty of Information Technology,
Brno University of Technology,
Božetěchova 1/2, 612 66 Brno, Czech Republic
Email: ikroustek@fit.vutbr.cz, xpokor32@stud.fit.vutbr.cz

***Abstract*—Machine-code decompilation is a reverse-engineering discipline focused on reverse compilation. It performs an application recovery from binary executable files back into the high level language (HLL) representation. One of its critical tasks is to produce an accurate and well-readable code. However, this is a challenging task since the executable code may be produced by one of the modern compilers that use advanced optimizations. One type of such an optimization is usage of so-called instruction idioms. These idioms are used to produce faster or even smaller executable files. On the other hand, decompilation of instruction idioms without any advanced analysis produces almost unreadable HLL code that may confuse the user of a decompiler. In this paper, we present a method of instruction-idioms detection and reconstruction back into a readable form with the same meaning. This approach is adapted in an existing retargetable decompiler developed within the Lissom project. The implementation has been tested on several modern compilers and target architectures. According to our experimental results, the proposed solution is highly accurate on the RISC (Reduced Instruction Set Computer) processor families, but it should be further improved on the CISC (Complex Instruction Set Computer) architectures.**

***Keywords*—compiler optimizations, reverse engineering, decompiler, Lissom, instruction idioms, bit twiddling hacks**

I. Introduction

REVERSE engineering is a process analyzing existing objects to discover knowledge about their functionality. Within the computer and information security, reverse engineering is often used for analysis of binary executable files. This is useful for vulnerability detection, malware analysis, compiler verification, code migration, etc. In present, this analysis is commonly done by using low-level tools such as disassemblers and dumpers.

Machine-code decompilers (i.e. reverse compilers) are more effective but not so wide-spread. Their task is to recover HLL representation (e.g. C source code) from executable files. In contrast to compilation, the process of decompilation is much more difficult because the decompiler must deal with incomplete information on its input (e.g. information used by the compiler but not stored within the executable file). Furthermore, the input machine code is often heavily optimized by one of the modern compilers (e.g. GCC, LLVM, MSVC); this makes decompilation even more challenging.

This work was supported by the BUT grant FEKT/FIT-J-13-2000 Validation of Executable Code for Industrial Automation Devices using Decompilation and BUT FIT grant FIT-S-11-2.

Code de-optimization is one of the necessary transformations used within decompilers. Its task is to properly detect the used optimization and to recover the original HLL code representation from the hard-to-read machine code. One example of this optimization type is the usage of *instruction idioms* [1]. An instruction idiom is a sequence of machine-code instructions representing a small HLL construction (e.g. arithmetic expression, assignment statement) that is highly-optimized for its execution speed and/or small size.

The instructions in such sequences are assembled together by using Boolean algebra, arbitrary-precision arithmetic, floating-point algebra, bitwise operations, etc. Therefore, the meaning of such sequence is usually hard to understand at the first sight. A notoriously known example is the usage of an exclusive or to clear the register content (i.e. `xor reg, reg`) instead of an instruction assigning zero to this register (i.e. `mov reg, 0`).

The goal of this paper is to present an approach how to deal with instruction-idioms detection and their reconstruction during decompilation. This approach has been successfully adapted within an existing decompiler developed within the Lissom project [2, 3]. Moreover, this decompiler is developed to be retargetable (i.e. independent on a particular target platform, operating system, file format, or a used compiler); therefore, the proposed analysis has to be retargetable too.

This paper is organized as follows. In Section II, we give an introduction to instruction idioms and their usage within compiler optimizations. Then, we briefly describe the retargetable decompiler developed within the Lissom project in Section III. Afterwards, we present our own approach in Section IV. The most common instruction idioms employed in the modern compilers are presented and illustrated in the same section. Section V discusses the related work of instruction idioms reconstruction. Experimental results are given in Section VI. Section VII closes the paper by discussing future research.

II. Instruction Idioms used in Compilers

In present, the modern compilers use dozens of optimization methods for generating fast and small executable files. Different optimizations are used based on the optimization level selected by the user. For example, the GNU GCC compiler supports these optimization levels[1]:

- `O0` – without optimizations;

[1] See http://gcc.gnu.org/onlinedocs/gcc/Optimize-Options.html for details.

978-83-60810-53-8/$25.00

- `O1` – basic level of speed optimizations;
- `O2` – the common level of optimizations (the ones contained in `O1` together with basic function inlining, peephole optimizations, etc.);
- `O3` – the most aggressive level of optimizations;
- `Os` – optimize for size rather than speed.

In the nowadays compilers, the emission of instruction idioms cannot be explicitly turned on by some switch or command line option. Instead, these compilers use selected sets of idioms within different optimization levels. Each set may have different purpose, but multiple sets may share the same (universal) idioms. The main reasons why to use instruction idioms are:

- The most straightforward reason is to exchange slower instructions with the faster ones. These optimizations are commonly used even on the lower optimization levels.
- The floating-point unit (FPU) might be missing, but a programmer still wants to use floating-point numbers and arithmetic. Compilers solve this task via floating-point emulation routines (also known as *software floating point* or *soft-float* in short). Such routines are generated instead of hardware floating-point instructions and they perform the same operation by using available (integer) instructions.
- Compilers often support an optimization-for-size option. This optimization is useful when the target machine is an embedded system with a limited memory size. An executable produced by a compiler should be as small as possible. In this case, the compiler substitutes a sequence of instructions encoded in more bits with a sequence of instructions encoded in less bits in general. This can save some space in instruction cache too.

Another type of optimization classification is to distinguish them based on the target architecture. Some of them depend on a particular target architecture. If a compiler uses platform-specific information about the generated instructions, these instructions can be classified as platform-specific. Otherwise, they are classified as platform-independent.

As an example of platform-independent idiom, we can mention the `div` instruction representing a fixed-point division. The fixed-point division (signed or unsigned) is one of the most expensive instruction in general. Optimizing the division leads to a platform-independent optimization.

On the other hand, clearing the content of the register by using the `xor` instruction (mentioned in the introduction) is a highly platform-specific optimization. Different platforms can use different approaches to clear the register content. As an example, consider the zero register on MIPS (`$zero` or `$0`), which always contains the value of 0. Using this register as a source of zero bits looks like a faster solution than using a `xor` instruction.

Furthermore, different compilers use different instruction idioms to fit their optimization strategies. For example, GNU GCC uses an interesting optimization when making signed comparison of a variable. When a number is represented on 32-bits and bit number 31 is representing the sign, logically shifting the variable right by 31 bits causes to set the zeroth bit equal to the original sign bit. The C programming language classifies 1 as *true* and 0 as a *false*, which is the expected result of the given less-than-zero comparison. This idiom is shown in Fig. 1. The Fig. 1a represents a part of a source code with this construction. The result of its compilation with optimizations enabled is depicted in Fig. 1b. We illustrate the generated code on the C level rather than machine-code level for better readability.

The compiler used the before-mentioned instructions idiom—replacing the comparison by the shift operation. The not-standardized `lshr()` function is used in the output listed in Fig. 1b. The C standard does not specify whether operator ">>" means logical or arithmetical right shift. Compilers deal with it in an implementation-defined manner. Usually, if the left-hand-side number used in the shift operation is signed, arithmetical right shift is used. Analogically, logical right shift is used for unsigned numbers.

III. Lissom Project Retargetable Decompiler

In this section, we briefly describe the concept of an automatically generated retargetable decompiler developed within the Lissom project [2]. This decompiler aims to be independent on any particular target architecture, operating system, object file format, or originally used compiler. The concept of the decompiler is depicted in Fig. 2. Its detailed description can be found in [3]. Currently, the decompiler supports decompilation of MIPS, ARM, and Intel x86 executable files stored in different file formats.

The input binary executable file is preprocessed at first. The preprocessing part tries to detect the used file format, compiler, and (optional) packer, see [4] for details. Afterwards, it unpacks and converts the examined platform-dependent application into an internal uniform Common-Object-File-Format (COFF)-based representation. Currently, we support conversions from UNIX ELF, Windows Portable Executable (WinPE), Apple Mach-O, Symbian E32, and Android DEX file formats. The conversion is done via our plugin-based converter described in [5, 6]. Afterwards, such COFF-file is processed in the decompilation core that consists of three basic parts—a *front-end*, a *middle-end*, and a *back-end*. The last two of them are built on top of the LLVM Compiler Infrastructure [7]. LLVM Intermediate Representation (LLVM IR) [8] is used as an internal code representation of the decompiled applications in all particular decompilation phases.

```
int main(void)
{
  int a, b;

  /* ... */

  b = a < 0;

  /* ... */
}
```

(a) Input.

```
int main(void)
{
  int a, b;

  /* ... */

  b = lshr(a, 31);

  /*... */
}
```

(b) Output (for better readability in C).

Fig. 1: Example of an instruction idiom (C code).

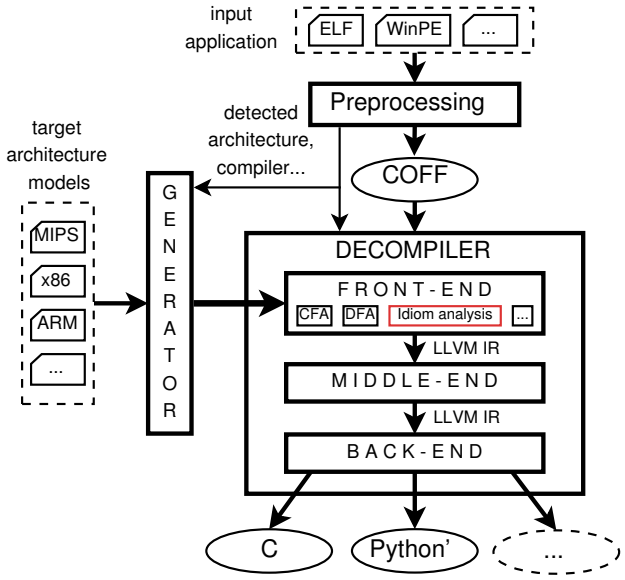


Fig. 2: The concept of the Lissom project retargetable decompiler.

After that, the unified COFF files are processed by the front-end part.Within this part, we use the ISAC architecture description language [9] for an automatic generation of the *instruction decoder*. The decoder translates the machine-code instructions into sequences of LLVM IR instructions. The resulting LLVM IR sequence characterizes behaviour of the original instruction independently on the target platform. This intermediate program representation is further analysed and transformed in the static-analysis phase of the front-end. This part is responsible for eliminating statically linked code, detecting the used ABI, recovery of functions, etc. [3]. When debugging information (e.g. DWARF, Microsoft PDB) or symbols are present in the input application, we may utilize them to get more accurate results, see [10].

The output of the front-end part (i.e. LLVM IR code representing input application) is sizable. The main reason is because it reflects a complete behavior of each machine-code instruction, which may not be necessary. For example, each side-effect of an instruction (e.g. setting a register flag based on instruction operands) is represented via the LLVM IR code, but results of these side-effects may not be used anywhere. Therefore, the front-end output is further processed within the middle-end phase, which is built on the top of the LLVM `opt` tool. This phase is responsible for reduction and optimization of this code by using many built-in optimizations available in LLVM as well as our own passes (e.g. optimizations of loops, constant propagation, control-flow graph simplifications).

Finally, the back-end part converts the optimized intermediate representation into the target high-level language (HLL). Currently, we support C and a Python-like language. The latter is very similar to Python, except a few differences—whenever there is no support in Python for a specific construction, we use C-like constructs. During the back-end conversion, high-level control-flow constructs, such as loops and conditional statements, are identified, reconstructed, and further optimized. Finally, it is emitted in the form of the target HLL.

The decompiler is also able to produce the call graph of the decompiled application, control-flow graphs for all functions, and an assembly representation of the application.

```
%a = load i32* @regs0
%b = xor i32 %a, %a
store i32 %b, i32* @regs0
```

(a) Optimized form of an instruction idiom in LLVM IR.

```
store i32 0, i32* @regs0
```

(b) De-optimized form of an instruction idiom in LLVM IR.

Fig. 3: Example of the bit-clear `xor` instruction-idiom transformation.

IV. Idiom Analysis and Reconstruction in the Retargetable Decompiler

The aim of the decompiler presented in the previous section is to allow retargetable decompilation independently on the particular target platform or the used compiler. Therefore, the methods of instruction idiom detection and reconstruction have to be retargetable too. For this reason, we implement these methods within the front-end phase because it uses the unified code representation in the LLVM IR format.

Detection of instruction idioms is based on a detection algorithm operating on the LLVM IR code. LLVM IR is a set of low-level instructions similar to assembly instructions. Moreover, LLVM IR is platform-independent and strongly typed, which meets our requirements. Therefore, machine instructions from different architectures can be easily mapped to sequences of LLVM IR instructions. This brings an ability to implement platform-independent instruction-idiom analysis.

The detection algorithm is similar to a peephole technique used in optimizing compilers [11]. It operates on a basic-block level, where each basic block contains a continuous sequence of instructions described via LLVM IR operations. A particular idiom is detected only if a basic block contains predefined LLVM IR instructions stored in a proper order and they must contain expected operand values (e.g. constant, register number). While examining instructions in an idiom sequence, we may found unrelated instructions (e.g. inserted by a code-motion compiler optimization). The detection algorithm may skip these instructions and continue the search on the following ones, but only if they do not modify the operands of already detected instructions. Otherwise, the search continues behind the current instruction from the beginning. Whenever an instruction idiom is detected, it is substituted by its more readable de-optimized version, once again in the LLVM IR form. The skipped instructions are left untouched on their original positions.

It should be noted that this algorithm does not search for a particular idiom over multiple basic blocks in present. The detection enhancement via usage of a control-flow analysis represents a future research. However, according to our experimental tests, the nowadays common compilers rarely scatter instructions related to idioms over multiple basic blocks. Therefore, the impact of this enhancement may be low.

An example demonstrating this substitution on the LLVM IR level is shown in Fig. 3. Fig. 3a represents the already mentioned `xor` bit-clear instruction idiom. To use register

TABLE I: Shortened list of instruction idioms found in compilers.

Instruction idiom	GNU GCC	Visual Studio C++	Intel C/C++ Compiler	Open Watcom	Borland C Compiler
Less than zero test	✓	×	✓	×	×
Greater equal zero test	✓	×	×	×	×
Bit clear by using `xor`	✓	✓	✓	✓	✓
Bit shift multiplication	✓	✓	✓	✓	✓
Bit shift division	✓	✓	✓	✓	✓
Division by `-2`	✓	×	×	×	×
Expression `-x - 1`	✓	✓	×	×	×
Modulo power of two	✓	✓	✓	×	×
Negation of a float	✓	×	×	×	×
Assign `-1` by using `and`	×	✓	✓	×	×
Multiplication by an invariant	✓	✓	✓	✓	×
Signed modulo by an invariant	✓	×	✓	×	×
Unsigned modulo by an invariant	✓	✓	✓	×	×
Signed division by an invariant	✓	✓	✓	×	×
Unsigned division by an invariant	✓	✓	✓	×	×
Substitution by `copysignf()`	✓	×	×	×	×
Substitution by `fabsf()`	✓	×	×	×	×

content, a register value has to be loaded into a typed variable `%a`. Using the `xor` instruction, all bits are zeroed and the result (in variable `%b`) can be stored back into the same register. To transform this idiom into its de-optimized form, a proper zero assignment has to be done. This de-optimized LLVM IR code is shown in Fig. 3b. In this case, the typed variable `%b` holds zero, which can be directly stored in the register.

In Table I, we can see a shortened list of instruction idioms used in common compilers. This list was retrieved by studying the source codes responsible for code generation (this applies to open-source compilers—GNU GCC 4.7.1 and Open Watcom 1.9) and via reverse engineering of executable files generated by these compilers (this method was used for other compilers—Microsoft Visual Studio C++ Compiler 16 and 17, Borland C++ 5.5.1, and Intel C/C++ Compiler XE13). Some of these instruction idioms are widespread among modern compilers. We have also found out that actively developed compilers, such as GNU GCC, Visual Studio C++, and Intel C/C++ Compiler, are using these optimizations heavily. For example, they generate the `idiv` instruction (fixed signed division) only in rare cases on the Intel x86 architecture; they generate optimized division by using magic number multiplication instead. The decompiler currently supports all of these idioms, among others.

A decompilation of an executable file is a time consuming process. The decompilation time highly depends on the executable size. A good approach how to optimize instruction-idioms analysis is to use any available information to save decompilation time. This is especially important when we support many instruction idioms. Some of them are specific for a particular compiler and therefore, they can be omitted from the detection phase whenever another compiler is detected. On the other hand, detection of the used compiler (as described in [4]) may be inaccurate in some cases and the algorithm will not detect any used compiler. In that case, the idiom analysis tries to detect all the supported idioms. Another optimization approach is to detect only the platform-specific idioms based on the target architecture and omit idioms for other architectures.

Transformation of instruction idioms by using LLVM IR is a quite straightforward task and it is entirely platform independent. On the other hand, the detection of an instruction idiom is more challenging. For example, the expected operand values (e.g. values used for magic number multiplication) may not be stored as clearly as in a original HLL source code. For example, the original HLL constant may not be stored directly as a number (i.e. immediate value), but it may be computed through several machine-code instructions. These instructions fold the original value at run-time based on different resources (e.g. register value, memory content). For example, the MIPS instruction set does not allow direct load of 32-bit immediate value and it has to be done using more instructions (e.g. `lui` and `ori`). Therefore, the operand value is not stored directly within one instruction but it is assembled by an instruction sequence. This is quite complicated because the idiom-detection phase (as well as the rest of the decompiler) is done statically and run-time information is unavailable. To deal with this problem, we utilize a static-code interpreter, originally used for function reconstruction—see [3] for a detailed description of the static-code interpreter.

Using an interpreter to statically compute a value stored in a register is quite common task in instruction-idiom analysis. An example is shown in Fig. 4. An interpreter has to be run to statically compute a number stored in register `@regs3` by using the backtracking of previously used operations and their operands. The obtained result is in this case `680390859`. This number is used in optimized division by number `101` performed by the magic-number multiplication on ARM and the GNU GCC compiler. Another similar issue is accessing the data segment to load constants; the interpreter can solve this issue as well.

The last implementation issue is not related to idiom detection or transformation, but to the processing within the middle-end phase of the decompiler. This phase is responsible for optimization of the LLVM IR code generated by the front-end phase. We exploit the LLVM `opt` and its optimizations for this purpose. For example, it serves to reduce the code size, which has a positive impact on the decompilation results (as

```
%a = add i32 679477248, 0
store i32 %a, i32* @regs3

%b = load i32* @regs3
%b_1 = add i32 913408, 0
%b_2 = add i32 %b_1, %b
store i32 %b_2, i32* @regs3

%c = load i32* @regs3
%c_1 = add i32 203, 0
%c_2 = add i32 %c_1, %c
store i32 %c_2, i32* @regs3

; @regs3 contains value 680390859
; = 203 + 913408 + 679477248
```

Fig. 4: An example of a constant computation in LLVM IR.

has been discussed in Section III). Besides our decompilation project, `opt` is normally used as an optimization part of the LLVM compiler toolchain. However, LLVM is a modern compiler toolchain and it also uses instruction idioms for code optimizations. Therefore, the `opt` tool has tendencies to bring back idioms instead of the de-optimized code. Therefore, we had to disable these optimization passes used in `opt`.

In Fig. 5, we demonstrate reconstruction of another idiom. In this figure, we can compare decompilation results with and without the instruction-idiom analysis. Fig. 5a illustrates a simple C program containing the division idiom. Decompilation result obtained without instruction-idiom analysis is depicted in Fig. 5c. It contains three shift operations and one multiplication by a magic value. Without the knowledge of fundamentals of this idiom, it is almost impossible to understand the resulting code. On the other hand, the decompilation result with instruction-idiom analysis enabled is well readable and a user can focus on a program sense, not on deciphering optimizations done by a compiler, see Fig. 5b.

```
int main(void)
{
  int a;

  /* ... */

  a = a / 10;

  /* ... */
}
```

(a) Input.

```
int main(void)
{
  int a;

  /* ... */

  a = a / 10;

  /* ... */
}
```

(b) Output with idiom analysis enabled.

```
int main(void)
{
  int a;

  /* ... */

  a = (lshr(a * 1717986919, 32) >> 2) -
      (a >> 31);

  /* ... */
}
```

(c) Output with idiom analysis disabled.

Fig. 5: C code example of decompilation with and without the idiom analysis.

```
#include <stdio.h>
int main(void)
{
  int a;

  /* ... */

  printf("1. Multiply: %d\n", a * 4);
  printf("2. Divide: %d\n", a / 8);
  printf("3. >= 0 idiom: %d\n", a >= 0);
  printf("4. Magic sign-div: %d\n", a / 10);
  printf("5. XOR by -1: %d\n", -a - 1);
  return a;
}
```

Fig. 6: C source code used for decompilers' testing.

V. Related Work

The fundamentals of instruction idioms and their usage within compiler optimizations are well documented, see [1, 12–16]. From these publications, we can gain insights into the principles behind instruction idioms as well as how and when to use them to obtain a more effective machine code.

Contrariwise, the detection and reconstruction of instruction idioms from a machine code is mostly an untouched area of machine-code decompilation. This topic is only briefly mentioned in [17–19]. Nevertheless, some of the existing (non-retargetable) decompilers support this feature. In order to observe the state of the art, we look closely on their approaches.

We used a test containing five idioms from a larger list listed in Table I. These idioms are the most common ones (e.g. multiplication via left shift) and the support of idiom reconstruction within a tested decompiler should be easily discovered via these idioms. A source code of this test is listed in Fig. 6. Each expression of the `printf` function represents one instruction idiom, whose meaning is described in Section IV. This source code was compiled for different target platforms (i.e. processor architecture, operating system, and file format) based on their support in each decompiler. Finally, each decompiler was tested by using this executable file and we analysed the decompiled results afterwards.

Boomerang is the only existing open-source machine-code decompiler [20]. However, it is no longer developed. According to our tests, it was able to reconstruct only the first instruction idiom.

REC Studio (also known as REC Decompiler) is freeware, but not an open-source decompiler. It has been actively developed for more than 25 years [21]. None of the instruction idioms was successfully reconstructed. We only noticed that REC Studio can reconstruct the register cleaning idiom (via the `xor` instruction) described in Section I.

SmartDec decompiler is another closed-source decompiler specialising on decompilation of C++ code, see [22] for details.

However, SmartDec was unable to reconstruct any instruction idiom from the machine-code.

Hex-Rays decompiler [23] achieved the best results—three successfully reconstructed idioms from five (it succeeded in the 1st, 2nd, and 4th test). Therefore, we have chosen this decompiler for a deeper comparison with our own solution as described in Section VI.

There are two other interesting projects. The *dcc* decompiler was the first of its kind, but it is unusable for modern real-world decompilation because it is no longer developed [17, 24]. On the other hand, the *Decompile-it.com* project looks promising, but the public beta version [25] is probably still in an early version of development and it cannot handle any of these instruction idioms.

In conclusion, we cannot compare our idiom-detection algorithm with approaches used in other tools because of two reasons. (1) They are either not distributed as open-source. (2) The open-source solutions do not support idiom recovery at all or they support only a very limited number of idioms. On the other hand, we can compare our results with the Hex-Rays Decompiler.

VI. Experimental Results

This section contains an evaluation of the proposed method of instruction-idiom analysis and reconstruction. The decompiled results are compared with the nowadays decompilation "standard"—the Hex-Rays Decompiler [23] that is a plugin to the IDA disassembler [26]. We used the latest versions of these tools, i.e. Hex-Rays Decompiler v1.8.0.130306 and IDA disassembler v6.4.130306. The Hex-Rays Decompiler is not an automatically generated retargetable decompiler, such as our solution, and it supports the Intel x86 and ARM target architectures. Our solution also supports the MIPS architecture at the moment.

All the three mentioned architectures are described as instruction-accurate models in the ISAC language in order to automatically generate our retargetable decompiler. MIPS is a 32-bit processor architecture, which belongs to the RISC processor family. The processor description is based on the MIPS32 Release 2 specification [27]. ARM is also a 32-bit RISC architecture. The ISAC model is based on the ARMv7-A specification with the ARM instruction set [28]. The last architecture used for the comparison is Intel x86 (also known as IA-32) that belongs in the CISC processor family. The model is based on the 32-bit processor core specified in [29] without extensions (e.g. x86-64).

We created 21 test applications in the C language. Each test is focused on a detection and reconstruction of a different instruction idiom. The Minimalist PSPSDK compiler (version 4.3.5) [30] was used for compiling MIPS binaries into the ELF file format, the GNU ARM toolchain (version 4.1.1) [31] for ARM-ELF binaries, and the GNU compiler GCC version 4.7.2 [32] for x86-ELF executables (the 32-bit mode was forced by the `-m32` option).

As can be observed, we used the ELF file format in each test case; however, the same results can be achieved by using the WinPE file format [33, 34]. All three compilers are based on GNU GCC. The reason for its selection is the fact that it allows retargetable compilation to all three target architectures and it also supports most of the idioms specified in Sections II and IV.

Different optimization levels were used in each particular test case. Because of different optimization strategies used in compilers, not every combination of source code, compiler, and its optimization level leads to the production of an instruction idiom within the generated executable file. Therefore, we count only the tests that contain instruction idioms. Furthermore, it is tricky to create a minimal test containing an instruction idiom without its removal by compiler during compilation.

An example of this problem is depicted by using a C code with multiplication idiom in Fig. 7a. The result of this code can be computed during compilation; therefore, the compiler emits directly the result without the code representing its computation (see the example in Fig. 7b). Therefore, we use functions from the standard C library for initialization of variables used in idioms. For example, this can be done by using statements `a = rand();` or `scanf("%d", &a);`. Example of an enhanced test is depicted in Fig. 7c. Such code cannot be eliminated during compilation and the instruction idiom is successfully generated in the executable file, see Fig. 7d.

The testing was performed on Intel Core i5 (3.3 GHz), 16 GB RAM running a Linux-based 64-bit operating system. The GCC compiler (v4.7.2) with optimizations enabled (`O2`) was used for creation of the decompiler.

Finally, we enabled the emission of debugging information in the DWARF standard [35] by using the `g` option because both decompilers exploit this information to produce a more accurate code, see [10] for details. The debugging information help to eliminate inaccuracy of decompilation (e.g. entry-point detection, function reconstruction) that may influence testing. However, the debugging information does not contain information about usage of idioms and therefore, its usage does not affect the idiom-detection accuracy.

All test cases are listed in Table II. The first column represents description of a particular idiom used within the test. The maximal number of points for each test on each architecture is

```c
int main(void)
{
  int a = 1;
  a = a * 8;
  return a;
}
```

(a) Test C code.

```c
int main(void)
{
  return 8;
}
```

(b) Compiler optimized code without instruction idiom.

```c
#include <stdlib.h>
int main(void)
{
  int a = rand();
  a = a * 8;
  return a;
}
```

(c) Enhanced test C code.

```c
#include <stdlib.h>
int main(void)
{
  int a = rand();
  a = a << 3;
  return a;
}
```

(d) Compiler optimized code with an instruction idiom.

Fig. 7: Problem of idiom removal by compiler.

TABLE II: Experimental results—number of successfully detected and reconstructed instruction idioms. Note: several tests differ only in the used numeric constant; however, different instruction idioms are emitted based on this value.

Tested instruction idiom	MIPS			ARM					Intel x86				
		Lissom			Lissom		Hex-Rays			Lissom		Hex-Rays	
	tests	✓	(%)	tests	✓	(%)	✓	(%)	tests	✓	(%)	✓	(%)
`intA = intB < 0`	5	5	100.0	5	5	100.0	0	0.0	5	0	0.0	0	0.0
`intA = intB >= 0`	5	5	100.0	5	5	100.0	0	0.0	5	5	100.0	0	0.0
`intA = 0`	0	-	-	0	-	-	-	-	1	1	100.0	1	100.0
`intA = intB * 4`	5	5	100.0	5	5	100.0	5	100.0	5	5	100.0	5	100.0
`intA = intB / -2`	0	-	-	5	5	100.0	5	100.0	4	0	0.0	4	0.0
`intA = intB / 4`	1	0	0.0	5	5	100.0	5	100.0	4	0	0.0	4	0.0
`intA = intB / 10`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`intA = intB / 120`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`uintA = uintB / 7`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`uintA = uintB / 9`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`intA = -intB - 1`	5	5	100.0	5	5	100.0	0	0.0	5	5	100.0	0	100.0
`intA = intB % 2`	0	-	-	5	4	80.0	2	40.0	4	0	0.0	0	0.0
`intA = intB % 3`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`intA = intB % 5`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`intA = intB % 8`	0	-	-	4	4	100.0	3	75.0	4	0	0.0	1	25.0
`uintA = uintB % 3`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`uintA = uintB % 5`	0	-	-	4	4	100.0	4	100.0	4	0	0.0	4	100.0
`uintA = uintB % 8`	5	5	100.0	5	5	100.0	1	20.0	5	0	0.0	0	0.0
`floatA = -floatB`	5	5	100.0	5	5	100.0	0	0.0	0	-	-	-	-
`floatA = copysign(floatB, floatC)`	5	5	100.0	5	5	100.0	0	0.0	0	-	-	-	-
`floatA = fabs(floatB)`	5	5	100.0	5	5	100.0	0	0.0	0	-	-	-	-
Total	**41**	**40**	**97.6**	**91**	**90**	**98.9**	**53**	**58.2**	**74**	**16**	**21.6**	**47**	**63.5**

five (i.e. one point for each optimization level – `O0`, `O1`, `O2`, `O3`, `Os`). Some idioms are not used by compilers based on the optimization level or target architecture; therefore, the number of total points can be lower than five. For example the MIPS and ARM architectures lack a floating-point unit (FPU) and the essential FPU operations are emulated via *soft-float* idioms. On the other hand, the Intel x86 architecture implements these operations via the x87 floating-point instruction extension; therefore, the instruction idioms are not used in this case.

The decompilation results are depicted in Fig. 8. We can observe four facts based on the results. (1) The Hex-Rays decompiler does not support the MIPS architecture; therefore, we are unable to compare our results on this architecture. (2) Results of the Hex-Rays decompiler on ARM and Intel x86 are very similar (approximately 60%). Its authors covered the most common idioms for both architectures (multiplication via bit shift, division by using magic-number multiplication, etc.). However, the non-traditional idioms are covered only partially or not at all (e.g. integer comparison to zero, floating-point idioms). (3) Our solution achieved almost perfect results on MIPS and ARM; only one test for each architecture failed.

(4) The concept of idiom detection within the front-end phase reaches its limits on the Intel x86 architecture, where the accuracy drops to 20%. The difference between the same tests for ARM (or MIPS) and x86 lies in the complexity of the instruction-semantics description. In general, RISC instructions have only a few side effects (modification of registers, flags, or memory) and their behavioural description in LLVM IR is compact. Therefore, detection of instruction idioms on such smaller pieces of code is quite easy. Contrariwise, almost every CISC instruction has several side-effects and its description in LLVM IR is much more longer. In such long code sequences of LLVM IR, we are unable to detect idioms with higher accuracy. The solution of this problem represents a future research and it is described in Section VII.

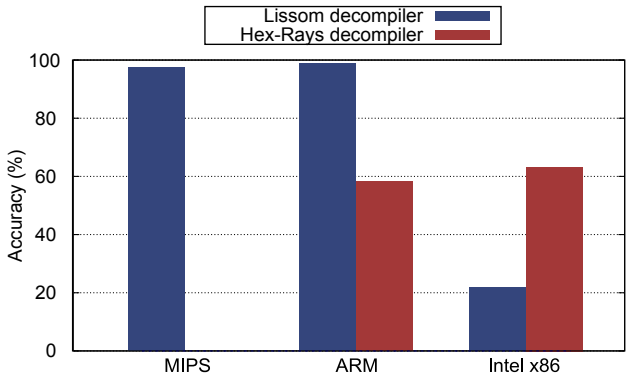


Fig. 8: Results of idiom analysis and reconstruction architecture.

VII. Conclusion

In this paper, we have focused on the problem of instruction idiom detection and reconstruction during the decompilation process of an existing retargetable decompiler. We proposed a new concept of this analysis that has been successfully tested on the MIPS, ARM, and x86 architectures within the Lissom project [2] retargetable decompiler.

In conclusion of the experimental results, our solution is capable to detect and reconstruct instruction idioms for the common RISC architectures with a very good accuracy (i.e. more than 97%), which is better than existing non-retargetable decompilers (some of them lacks this analysis as we demonstrated in Section V). However, the accuracy drops down significantly with the increasing instruction-set complexity. This issue has to be solved in the future research.

The major problem is the code complexity of CISC instructions during the early phases of decompilation (i.e. the front-end phase). On this level, it is problematic to properly detect the idiom without an increase of the false-positive ratio (that

is kept to be zero). However, it should be possible to perform this analysis after the optimization phase (i.e. the middle-end phase) that is based on the LLVM `opt` tool. This optimization phase will simplify the analysed code and it should be possible to detect idioms more easily. Moreover, changing order of this phase is supposed to have a minimal effect on the other phases.

The second propose of the future research lies in further testing of the retargetable idiom detection and reconstruction by using executables created by different compilers and for different target architectures. There is always a room for improvement by adding new instruction idioms into our database of supported idioms.

Finally, usage of control-flow analysis for instruction-idiom detection may be useful when dealing with more aggressive optimizations.

REFERENCES

[1] H. S. Warren, *Hacker's Delight*. Boston, US-MA: Addison-Wesley, 2003.

[2] Lissom, http://www.fit.vutbr.cz/research/groups/lissom/, 2013.

[3] L. Ďurfina, J. Křoustek, P. Zemek, and B. Kábele, "Detection and recovery of functions and their arguments in a retargetable decompiler," in *19th Working Conference on Reverse Engineering (WCRE'12)*. Kingston, ON, CA: IEEE Computer Society, 2012, pp. 51–60.

[4] J. Křoustek and D. Kolář, "Preprocessing of binary executable files towards retargetable decompilation," in *8th International Multi-Conference on Computing in the Global Information Technology (ICCGI'13)*. Nice, FR: International Academy, Research, and Industry Association (IARIA), 2013, pp. 1–6.

[5] J. Křoustek, P. Matula, and L. Ďurfina, "Generic plugin-based convertor of executable file formats and its usage in retargetable decompilation," in *6th International Scientific and Technical Conference (CSIT'11)*, 2011, pp. 127–130.

[6] J. Křoustek and D. Kolář, "Object-file-format description language and its usage in retargetable decompilation," in *AIP Conference Proceedings (SCLIT'12)*, vol. 1479. American Institute of Physics (AIP), 2012, pp. 466–469.

[7] The LLVM Compiler Infrastructure, http://llvm.org/, 2013.

[8] LLVM Assembly Language Reference Manual, http://llvm.org/docs/LangRef.html, 2013.

[9] K. Masařík, *System for Hardware-Software Co-Design*, 1st ed., ser. VUTIUM. Brno, CZ: Brno University of Technology, Faculty of Information Technology, 2008.

[10] J. Křoustek, P. Matula, J. Končický, and D. Kolář, "Accurate retargetable decompilation using additional debugging information," in *6th International Conference on Emerging Security Information, Systems and Technologies (SECURWARE'12)*. International Academy, Research, and Industry Association (IARIA), 2012, pp. 79–84.

[11] J. W. Davidson and D. B. Whalley, "Quick compilers using peephole optimization," *Software: Practice and Experience*, vol. 19, no. 1, pp. 79–97, 1989.

[12] M. Beeler, R. W. Gosper, and R. Schroeppel, *HAKMEM*. Massechusetts Institute Of Technology, 1972.

[13] R. M. Stallman and the GCC Developer Community, "GNU Compiler Collection Internals," http://gcc.gnu.org/onlinedocs/gccint.pdf, 2010.

[14] W. von Hagen, *The Definitive Guide to GCC*. Apress, 2006.

[15] R. Hyde, *The Art of Assembly Language*. San Francisco, US-CA: No Starch Press, 2003.

[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, UK: Cambridge University Press, 2007.

[17] C. Cifuentes, "Reverse compilation techniques," Ph.D. dissertation, School of Computing Science, Queensland University of Technology, Brisbane, QLD, AU, 1994.

[18] M. J. V. Emmerik, "Static single assignment for decompilation," Ph.D. dissertation, University of Queensland, Brisbane, QLD, AU, 2007.

[19] L. Ďurfina, J. Křoustek, P. Zemek, D. Kolář, T. Hruška, K. Masařík, and A. Meduna, "Advanced static analysis for decompilation using scattered context grammars," in *Applied Computing Conference (ACC'11)*. World Scientific and Engineering Academy and Society (WSEAS), 2011, pp. 164–169.

[20] Boomerang, http://boomerang.sourceforge.net/, 2013.

[21] Reverse Engineering Compiler (REC), http://www.backerstreet.com/rec/rec.htm, 2013.

[22] SmartDec, http://decompilation.info/, 2013.

[23] Hex-Rays Decompiler, www.hex-rays.com/products/decompiler/, 2013.

[24] The dcc Decompiler, http://itee.uq.edu.au/~cristina/dcc.html, 2013.

[25] Decompile-It.com – Online C Decompiler, http://decompile-it.com/, 2013.

[26] IDA Disassembler, www.hex-rays.com/products/ida/, 2013.

[27] MIPS Technologies Inc., *MIPS32 Architecture for Programmers Volume II-A: The MIPS32 Instruction Set*, MIPS MD00086 ed., 2010, https://www.mips.com/products/architectures/mips32/.

[28] ARM Limited, *ARM Architecture Reference Manual: ARMv7-A and ARMv7-R edition*, ARM DDI 0406C ed., 2011, https://silver.arm.com/download/download.tm?pv=1199569.

[29] Intel Corporation, "Intel 64 and IA-32 architectures software developer's manual volume 1: Basic architecture," 2013, http://download.intel.com/products/processor/manual/253665.pdf.

[30] Minimalist PSPSDK, http://sourceforge.net/projects/minpspw/, 2013.

[31] GNU ARM Toolchain, http://www.gnuarm.com/, 2012.

[32] GCC: the GNU Compiler Collection, http://gcc.gnu.org/, 2013.

[33] TIS Committee, "Tool Interface Standard (TIS) Executable and Linking Format (ELF) Specification," 1995, http://refspecs.freestandards.org/elf/elf.pdf.

[34] Microsoft Corporation, "Microsoft portable executable and common object file format specification," http://www.microsoft.com/whdc/system/platform/firmware/PECOFF.mspx, 2013, version 8.3.

[35] *DWARF Debugging Information Format*, 4th ed., DWARF Debugging Information Committee, 2010, http://www.dwarfstd.org/doc/DWARF4.pdf.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1515–1522

Declarative Specification of References in DSLs

Dominik Lakatoš*, Jaroslav Porubän†, Michaela Bačíková‡
Technical University of Košice
Letná 9, Košice, Slovakia
Email: {dominik.lakatos*, jaroslav.poruban,† michaela.bacikova‡}@tuke.sk

***Abstract*—The occurrence of identifiers and references in computer languages is a common issue. The same applies for domain specific languages, whose popularity is increasing and there is a need for aid in their design process. This paper analyses the problem of identifiers and references in computer languages. Current methods use an imperative approach for supporting references in languages; therefore a language designer is required to manually write reference resolving. The method proposed in this paper perceives references and identifiers as language patterns, which can be specified in a declarative manner with much less knowledge about the problem of resolving references in computer languages.**

I. Introduction

REFERENCES in languages are common. We use simple identifiers in our everyday life to identify objects, activities, abstract designs, etc. Even every one of us has an identifier, which we call "name". If we want to reference to somebody else, we use the name of that person. The problem arises when somebody uses short name *John* and there is more than one person with that short name known to him. How do we know, which *John* is he or she referring to? In that case we have to know more information and understand the mutual scope of names between the communicating parties to be able to properly identify the correct person.

The situation is not different in the area of computer languages. Majority of computer languages use references in a form of variable names, function names or any other named structures. How can we decide if the identifier used in our code refers to this particular structure, if we have more than one structure declared with the same name? In the theory of computer languages, the process of searching for these identifiers is called *reference resolving*. The basic solution of reference resolving is routine: A language designer creates a table of identifiers for storing every identifier declared in the code. Then he/she defines a lookup method for searching in the table for a suitable identifier based on the given textual reference in the code. Every language needs to solve the lookup process with its own function even when such a function is well-known and similar in most languages. Defining the lookup function is a common and routine task, but the implementation is poorly automatized.

In the last few years we can notice the increase in the language-oriented development [1]. Developers create special small or medium sized languages, called *domain specific languages (DSLs)* for many different areas. The development of DSLs is difficult [2] and identifiers and references are common in DSLs, which doesn't make it easier without proper support. The problem of a reference resolving is particularly acute in the area of external DSLs, as they need to function autonomically without any existing general purpose language (GPL). In this paper we aim to simplify the specification of DSLs with references by proposing declarative manner of specifying language constructs as identifiers and references.

II. Declaring References in Languages

Every GPL uses some sort of references. Even the oldest ones use some a form of named variables and they need a method to resolve the variable identifiers on their places of usage. It is possible to identify two types of statements concerning variables:

- *declaration* - a place where a name (and, optionally, a type) is assigned to a variable
- *usage* - a place where we use or change the value contained in a named variable by referring to its name

The declaration of one variable can occur only once in a language sentence scope but usage of the variable can occur on multiple places. A typical example of a declaration is:

```
float a;
```

It is a declaration of a variable `a` restricted to contain a decimal value. A typical example of usage is an assignment or reading of a value:

```
a = 10; //assignment
factor(a); //reading of a value
```

Supporting tools for processing and executing sentences from a language can be created in two different ways. A language designer can design a language and manually implement all the tools needed for language processing. The second option is to write a language specification using any of the existing compiler generators and then generate the processing tools. Use of a compiler generators is preferable because it provides better automation and is frequently used in the area of specification of DSLs.

Compiler generator tools help with specifying lexical units, syntax, semantic actions [3] and sometimes they even explicitly define language concepts by means of abstract syntax. There are many compiler generator tools, we can mention the most common known Yacc [4], or even more sophisticated ones such as ANTLR [5], Beaver [6], LISA [7], [8], JastAdd [9] or Xtext framework [10], [11]. If we look on the problem of resolving textual references to places of its declaration, usually it can be addressed within the semantic

978-83-60810-53-8/$25.00

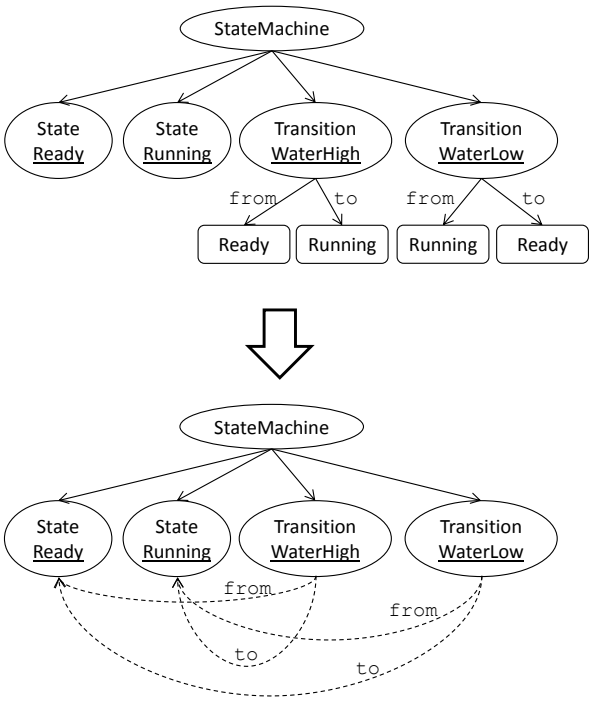


Fig. 1. Transformation from the abstract syntax tree to the abstract syntax graph

actions of the designed language. Some of the current compiler generators use attribute grammars in order to simplify the specification of semantic actions, for example in LISA. JastAdd uses an expanded version of attribute grammars [12] described by Hedin [13], [14] as reference attributed grammars (RAGs). RAGs aim to simplify attribute grammars for references by allowing attributes to be of reference type to other nodes in the syntax tree. RAGs only expand theoretical type model of attribute types and the actual process of discovering nodes in the tree needs to be manually implemented in the semantic actions of an attribute grammar. Xtext framework offers different approach with simple definition of referenced node types, scoping needs to be programmed with provided API.

We would like to show the usual process of declaring and resolving references on an example of a simple language of state machines. For the purposes of better comparison with the existing research we will use the example of the state machine language used in the JastAdd tutorial paper [15].

Listing 1. EBNF context-free grammar for a state machine language [15]

```
<statemachine> ::= <declaration>*
<declaration>  ::= <state> | <transition>
<state>        ::= "state" ID ";"
<transition>   ::= "trans" ID":" ID"->" ID ";"
ID = [a-zA-Z][a-zA-Z0-9]*
```

In the listing 1 is specified the grammar of the state machine language, which consists of *declarations*. A declaration can be a *state* or a *transition*. A non-terminal for a *state* contains only the state name represented by the named token *ID*. A *transition* contains the name (*ID*) of the transition, the name of the starting *state* and the name of the ending *state* of the transition. The abstract syntax of the state machine language is displayed in the listing 2. For every non-terminal there is one concept in the abstract syntax and instead of using just textual terminals for references (as it is common during the language desing phase) we used declarations of actual references.

Listing 2. Abstract syntax for the state machine language

```
concept StateMachine
  AS: declarations: list of Declaration

concept Declaration

concept State : Declaration
  AS: name: string

concept Transition : Declaration
  AS: name: string, from: State, to: State
```

In the abstract syntax and the grammar, it is possible to identify the point where there is a need for reference resolving during the language processing. Every *transition* has a textual name reference to the existing *state* concept. A sentence in the state machine language can be represented by a simple example of identifiers and references (see listing 3).

Listing 3. State machine language example sentence

```
state Ready;
state Running;
trans WaterHigh: Ready -> Running;
trans WaterLow: Running -> Ready;
```

Each *state* has a name, which acts as an identifier as well as each *transition* has a name serving as an identifier, but in this example we actually need only the state identifier because there will be no reference to the transition names. In the listing 3 we can see two states: *Ready* and *Running* and there are two transitions, each using the already defined state names as starting and ending states. During the language processing, the names represented as strings have to be connected to the actual state declarations. The process of interconnecting the referenced language concepts in any computer language can be perceived as a transformation from abstract syntax tree (AST) to abstract syntax graph (ASG). AST is a tree structure of language concepts created directly after parsing of the language textual form. Tree structure of textual forms of languages does not allow creation of direct references to existing declaration.Therefore we need textual placeholder for identifier to make references possible in later phase of language processing. ASG is a structure transformed from AST, which allows references from any node to any other node, so it means that a graph structure is created by extending the tree structure. The AST and ASG of our state machine code are displayed in fig. 1.

The transformation from AST to ASG in current compiler generators is accomplished manually within semantic actions. An example of semantic actions defined via attributed grammar written in JastAdd for a simple name analysis of the state machine language is shown in listing 4. Attributes are specified

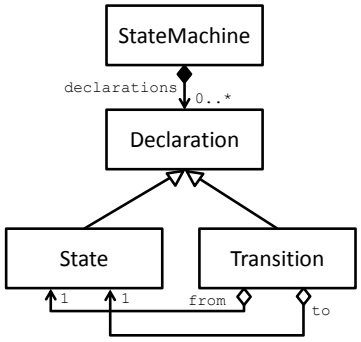


Fig. 2. Model of the state machine language

for *from Transition* named *source* and for *to Transition* named *target*. The *lookup* attribute is used for searching for a specific named *State* in the already specified list of *States*. Understanding of the provided example requires more knowledge about the specification notation used in JastAdd [15], but it shows the common amount of code needed to solve references in a language sentence and this problem is repeating in most of existing languages.

Listing 4. Code for name analysis for the state machine language specification in JastAdd [15]

```
aspect NameAnalysis {
 syn State Transition.source() =
       lookup(getSourceLabel());
 syn State Transition.target() =
       lookup(getTargetLabel());
 inh State Declaration
               .lookup(String label);

 eq StateMachine.getDeclaration(int i)
       .lookup(String label) {
   for (Declaration d :
              getDeclarationList()) {
     State match = d.localLookup(label);
     if (match != null) return match;
   }
   return null;
 }
 syn State Declaration
    .localLookup(String label) = null;
 eq State.localLookup(String label) =
   (label.equals(getLabel()))?this:null;
}
```

III. Language specification with annotated classes

There are different methods of specifying languages. The most common is in form of *grammars*. In our approach we use our own tool for language specification, which allows to model languages as classes in object-oriented programming paradigm [16]. This way the reuse of UML modeling tools for modeling languages [17] is possible.

In order to use our approach, the model of the state machine language has to be described by classes as described in Fig. 2. Each class in the model represents exactly one language concept. The connections between the language concepts are represented by the relationships in the model (inheritance, composition). Each UML composition relationship describes a required connection between the concepts, while the aggregation relationship means a connection created with identifiers. Generalization is used to represent alternation for the parent concept.

Modeling a language in a *graphical form* is great for general overview, but as we enrich a model with additional information such as concrete syntax and semantics, it becomes hard to understand. Execution of a plain model without any specified actions or supporting tool is usually not possible. In order to solve these problems and for practical purposes we define the language concepts in a textual form of classes of object oriented languages. The first class could be created for the root concept (*StateMachine*) of the state machine language.

Listing 5. StateMachine concept specification

```
class StateMachine {
  List<Declaration> declarations;

  StateMachine(List<Declaration> decls) {
    this.declarations = decls;
  }
}
```

Each class diagram is a representation of the abstract syntax with the recognized language concepts and their connections, as it is possible to notice in Fig. 2. Abstract syntax of *StateMachine* in listing 5 is represented by the definition of the class fields, in this case we have only one field named *declarations* for storing the list of *Declaration* concepts. Concrete syntax is represented by the class constructor. In this case there is only one concrete syntax representation (hence one constructor) which consists only of the list of *Declaration* concepts. It is a simple example for defining language constructs in a form recognizable to any programmer with a knowledge of object-oriented programming. In later examples we will discuss additional forms of concrete syntax specification. More details about this representation can be found in [16].

The next specification is a declaration of an abstract language concept *Declaration*, which in grammar form serves as a non-terminal for choosing between *State* and *Transition* non-terminals. It is possible to achieve the same effect in our language specification using inheritance and abstract classes. Therefore, the *Declaration* concept can be specified as a simple empty abstract class or an interface. We have chosen to use abstract class as this representation is closer to our model situation.

Listing 6. Declaration concept specification

```
abstract class Declaration {}
```

A. Declaring identifiers

The presence of identifiers and references is so common in computer languages, that we can consider it a language pattern. This pattern was already used in languages, but we have identified and distilled it as one of the building blocks of languages and therefore it can be extracted and later used

in any language in form of declaration. Our example of the state machine language uses a special form of unique named identifiers for the *State* concepts. And this is where our method comes forward and provides the capability to simply annotate a field which should be used as a special identifier for encapsulating the language concept. In our example we have the *State* language concept with the *label* property, which serves as an identifier of the *State* instance (see listing 7). In order to declare the *label* property to be used as identifier, we only need to annotate the property with the *@Identifier* annotation. The optional options available for the *@Identifier* annotation will be discussed later in this paper.

Listing 7. State concept specification

```
class State extends Declaration {
  @Identifier
  String label;

  @Before("state")
  @After(";")
  State(@Token("ID") String id) {
    this.label = id;
  }
}
```

We have used also other annotations, which are not new in our language specification. The *@Before* and *@After* annotations serve for defining terminal symbols in textual concrete syntax of *State*. The *@Token* annotation has a similar purpose, it can be used to concretize the lexical symbol used for checking and extracting data. If we compare this concrete syntax specification with the grammar defined for the state machine language in listing 1, it is simple to find connections between both specifications. The token *ID* is already defined in a special language configuration file in our tool and it serves for accepting names.

B. Declaring references

For the identifiers to have a logical meaning in the language, it is needed to reference to them from another place in the language sentence. The places used for referencing to existing identifiers are marked with the *@References* annotation. In our example we marked the constructor parameters *sourceLabel* and *targetLabel* as referencing names to the *State* concept. We need to define the concrete syntax according to the grammar from listing 1, therefore we are using the *@Before*, *@After* and *@Token* annotations as well.

Listing 8. Transition concept specification

```
class Transition extends Declaration {
  String label;
  State source;
  State target;

  @Before("trans")
  @After(";")
  Transition(
    @Token("ID")
    String label,

    @Before(":")
    @Token("ID")
    @References(State.class, field = "source")
    String sourceLabel,

    @Before("->")
    @Token("ID")
    @References(State.class, field = "target")
    String targetLabel
  ) {
    this.label = label;
  }
}
```

The usage of the *@References* annotation has an impact on some programming techniques. We are using *@References* to annotate constructor parameters and this way we are assigning special behavior to those parameters. Our language specification implementation allows us not to store the value of the annotated constructor parameter in any traditional programming way, actually this value is automatically stored for us during the language processing phase. Therefore, in the listing 8 it is possible to leave the constructor body with only one statement for storing the *label* of the transition. The requirement for the usage of the *@References* annotation is that it should be used on a *String* parameter and it is required to specify the referencing class (in the *value* parameter of the annotation), in our example it is the *State* class. The next parameter of *@References* is *field*, which allows us to define the name of the referenced class field (sometimes called property) used to store the specified language concept (object of the specified class). Both parameters are used later to filter all identifiers according to their corresponding language concepts, and they are used for injecting [18] appropriate object into the class field using reflection [19]. In our example of the *Transition* concept the objects of the *State* class are injected into the *source* and *target* fields, which are also checked for type consistency.

IV. Scope of References

References in computer languages are seldom created within one universal scope of availability. Even if there is a universal scope, called global scope, languages are not using only this one scope. Every variable in the language can be declared in a different scope and the rules for discovering a proper defined variable can be very difficult to define. An example of different scopes for the *count* variable is shown in listing 9. The *count* variable is declared in the global scope and at the same time it is also declared in the local scope of the *foo* function (D2), therefore the output of *foo* is `0` as we are using the variable declared closer to the place of its usage. The output of *goo* is `10` as the *count* variable name is referencing to the global variable declaration (D1) because there is no other closer variable declaration with this name.

Listing 9. Simple scope example with the *count* variable

```
int count = 10; // D1

void foo() {
  int count = 0; // D2
  print count;
}
```

```
void goo() {
  print count;
}
```

Support for using identifiers and references in scopes is very different in most compiler generator tools. Most of the tools are delegating this work to the designer by means of custom implementation (Yacc [4], Flex [20], JavaCC [21], Beaver [6]), they offer attribute grammars to help with the issue as it is done in JastAdd [9] and Lisa [7] or Xtext framework provides API for implementing scopes [11]. Still the language designer needs to know how to implement the scoping of references for his/her new language. ANTLR [22] provides a special support for defining scopes of variables for non-terminals and it allows easier access in order to find the declared variable in the tree of scoped variables, but still the user is required to specify proper variables and implement the use of scopes in the lookup functions. Therefore we can summarize, that all popular compiler generator tools require the user to write a custom solution for almost every issue concerning the reference resolving of identifiers within any scope.

In our method we support automatic reference resolution with scoping rules. Result of parsing language sentence is provided in a form of AST. References with scopes are allowed by selecting valid nodes in the required scope. In order to select nodes in AST it is useful to use a tree quering method. Our current solution uses the XPath querying language [23] as it is specially designed for efficient filtering of XML nodes [24] and use simple syntax. AST is usually not represented in XML, but XML is also a tree structure, therefore, the transformation from AST to XML is straighforward.

Each language is different and each reference in the language can be defined in a different scope. In order to characterize scoping, we have divided references to three levels of scoping:

- Global scope
- Simple local scope
- Complex scope

A. Global scope

The state machine language used in the previous sections uses references without scope or we can say that it uses global scope. Every *State* language concept is uniquely named within the full scope of a language sentence. Therefore it is a language within the first level of scoping and there is no need to declare it with any additional parameters. Previous listings (5, 6, 7 and 8) of class specifications of the state machine language are sufficient for our compiler generator with automatic resolution of references.

B. Simple local scope

Declaration of a referenceable language concept in the scope of parent concept is an example of a simple local scope level. Any scoping with a simple XPath query is considered to be a simple local, scope level. We can illustrate the problem on an example of a language for description of departments with cars and employees. Every car has a name, a fuel type and a year of acquisition. Every employee has a name and he/she can have a name of a car, which can be used by him/her. Employees can have a permission to drive the car owned by the same department, but it is not possible to use cars from different departments. Name of the car in the employee specification line is actually a reference to a declared car name in the department. Textual representation of this language is displayed in listing 10.

Listing 10. Department language with cars and employees

```
department: Accounting
  car: Audi A4; diesel; 2008
  car: Ford Focus; gas; 2012
  empl: John Smith {Audi A4}
  empl: Emma Fisher {Ford Focus}
  empl: Jim Parker

department: Technical services
  car: VW Golf; diesel; 2010
  car: Ford Focus; gas; 2005
  empl: Steve Cosby {Ford Focus}
```

Specification of the *Employee* language concept in our class form is displayed in listing 11. Scoping in references is allowed with the *path* parameter in the *@References* annotation. The value of the *path* parameter is an XPath query for accessing all relevant *Car* concepts. In our example we have used the XPath `parent::Department/Car`, which means that we are expecting to traverse the AST from the actual *Employee* node to the parent node named *Department* and then to find all child nodes named *Car*. This way, we have specified scoping for referencing to the *Car* concept within the same *Department* concept. The method of finding the *Car* node with a proper identifier is included in our method and it is not necessary to specify it in the XPath query, the user only needs to define the path for all relevant language concepts.

Listing 11. Specification of the *Employee* concept in the department language

```
class Employee {
  String name;
  Car car;

  @Before("empl:")
  Employee(
    String name,
    @Before("{")
    @After("}")
    @References(value=Car.class,
        path="parent::Department/Car")
    String car
  ) {
      this.name = name;
  }

  @Before("empl:")
  Employee(String name) {
    this.name = name;
  }
}
```

In the context of our approach, the *simple local scope* can be perceived as any local scope, which can easily be defined with an XPath query inside the *@References* parameter *path*. The complexity of a possible query is depending on the user's knowledge of XPath. Although, we are not restricted only to XPath, it is possible to use any other language for traversing the tree structures, however it needs to be included in our language tool.

C. Complex scope

In our research we aim for tool support of DSLs, which are usually not as complex as general purpose languages and therefore do not need to use complex scopes. Although, sometimes it is useful to have complex scope rules. In this section we will discuss the power of XPath expressions in our language design method and we provide an overview for advanced scoping.

A usual behavior of reference resolving using an XPath expression in the *path* parameter of the *@References* annotation consists of adding an XPath predicate with a test of the name of an identifier to the end of the original expression. For example an expression for finding a *Car* with the name `Audi A4` from listing 11 is converted to the following XPath expression.

```
parent::Department/Car[_id="Audi A4"]
```

In some XPath queries it is useful to specify a place for filtering nodes before getting all nodes. In order to write an XPath expression for searching for appropriate variable `count` declaration from listing 9 we can use an XPath expression

```
(ancestor::*[Variable/_id="count"])[last()]/
    Variable[_id="count"]
```

In order to parameterize this XPath expression we can use `##cmp##` as a placeholder for `_id = "NAME"`, therefore our extended XPath expression can be written as follows:

```
(ancestor::*[Variable/##cmp##])[last()]/
    Variable[##cmp##]
```

This XPath expression first chooses the appropriate ancestor with expected named variable declaration.

There is always a possibility to completely forget our declarative way of solving references and to implement it on our own. Our language method allows it without any problems, but the user would lose a great advantage. Still there can be some very specific scoping rules, which cannot be properly described with XPath and in this case we recommend using the traditional manual implementation of referencing to identifiers. The user has to be aware that he needs to manage his own storage of identifiers as well as the method for connecting places of identifier usage to the language concepts with the declared identifiers. Manual solution is possible and sometimes needed, although we are sure it is not necessary in most DSLs and our identifiers and references patterns are sufficient in such scenarios.

V. Process of Reference Resolving

Our method for reference resolving consists of three main steps:

- Unique identifier
- Reference searching and injecting
- Creation of undefined identifiers

A. Unique identifier

The first step in our reference resolving method is to check for uniqueness of identifiers. Every identifier for one language concept needs to be unique in its own scale. If identifiers are not unique it would cause a problem in reference resolving process, as one reference needs to reference only one language concept. This part of our method is providing functionality, which is not commonly used in other tools and where the user needs to implement it on his own using a table of identifiers or any other similar method. The default behavior of the *@Identifier* annotation without parameters is that the concept name is unique in the global scale of a language sentence and for language concept. Standard setting for uniqueness of identifiers used for specifying the *state* concept in the state machine language (see listing 7) is using the *@Identifier* annotation without parameters, therefore during language processing it is tested if every *state* has a unique name.

In order to specify other scale of identifier uniqueness, we are using XPath expressions as it was described in the section IV-B. It is possible to define a query for nodes of AST in which there is only one occurrence of an identifier of the specified type. This way we can define automatic checking for uniqueness of identifiers of *car* in *department* in our department language example (listing 10).

Listing 12. Specification of the *Car* concept in the department language

```
class Car {
 @Identifier(unique="parent::Department")
  String name;
  String fuelType;
  int year;

 @Before("car:")
  Car(
   @Token("NAME")
    String name,

   @Before(";")
   @Token("NAME")
    String fuelType,

   @Before(";")
   @Token("YEAR")
    int year) {
      this.name = name;
      this.fuelType = fuelType;
      this.year = year;
  }
}
```

Listing 12 specifies the language concept *car*, but our main focus is on the *@Identifier* annotation with the *unique* parameter. With this parameter we have specified that every *car*

has a unique name only in the scope of its parent *department* language concept. Therefore it is possible to have cars with the same name in different departments and the parsing process will finish without any error. Otherwise, in the case of not specifying the *unique* parameter, the uniqueness of *car* name would be tested in the global scale, producing an error even if different departments would have a car with the same name.

Scoping of references defined in the section IV-B and scoping of identifier uniqueness is based on similar XPath expressions. It is recommended to scope uniqueness of an identifier when we are using scopes for referencing. Identifier XPath scopes are usually simpler then XPath expressions for finding proper identifiers in *@References*. As an example, consider uniqueness scope of identifier in a function for listing 9, which can be defined as a simple XPath for variable identifier: `parent::Function` and if we compare it with XPath expression for finding proper variable reference shown in the section IV-C, it is obvious that the XPath expression for uniqueness scope is much simpler to write and understand.

B. Reference searching and injecting

The second and the most important part of our method is actual resolving of textual references. Details about textual reference resolving were already described in this paper in section III and scoping details in section IV. In our method we are using declarative specification of references with the *@References* annotation, which marks the textual parameters containing names of referencing language concepts defined in the fields marked with the *@Identifier* annotation. Referencing can be constrained with the scope rules defined by XPath. Our method discoveres referenced language concepts and inject them into fields specified in the *field* (for specifying the field name of class) and *value* (used to define the referencing type) parameters of the *@References* annotation.

Reference resolving can be carried out in two ways:

- after parsing an entire input sentence (explicit)
- during the parsing process (implicit)

The explicit approach is about an explicit execution of reference resolvining at the end of the parsing process. After execution we get an exact information about any inconsistencies in identifiers and references. Any error can be detected right after explicit execution call and propagated to the user. On the other hand it needs this one explicit execution of reference resolving and it makes this solution less modular.

The implicit approach is about resolving references and identifiers after each language concept creation (usually during the parsing process). For each parsed language concept it is trying to resolve the unresolved references and to match them with identifiers. The advantage of such solution is elimination of explicit execution of reference resolving, on the other hand it is hard to check for inconsistencies as it cannot tell if the registration of the new language concepts has finished or not. It is possible only to check for the actual state of reference resolution and to get information about the non-resolved references. At the end of the process of language parsing we should not have any non-resolved references. It is optional to check for non-resolved references, when we are using the implicit approach to resolution of references.

Both presented approaches to reference resolution have advantages and disadvantages and it is possible to use either of them. A language designer should decide which approach would be preferable to his new language. The implicit approach does not need any action for resolution of the references, but cannot guarantee that all references have been resolved properly without any optional check. The explicit approach is better in performance as the resolution of references is done only once and it can propagate the error in case of unresolved references, but it is required to execute this method explicitly and therefore there is a direct dependency between the parsing process and the reference resolution process.

C. Creation of undefined identifiers

The third part of our method is focused on the solution of one specific problem concerning references and identifiers within languages. In languages it is possible to have a reference to an identifier without the previous declaration of the identifier. It is a common occurrence in untyped or dynamically typed languages to use a variable without previous declaration (see listing 13). It is safe to claim, that a declaration of an identifier is required only if the declaration contains additional information about the language concept except the identifier name.

Listing 13. Example of a language without variable declaration

```
x = 10; // it is possible to use this

var x;  // instead of this
x = 10;
```

The previous example represents a GPL sentence fragment, but the same situation can be found when designing a DSL language. If we would look on the sentence written in the state machine language in listing 3 it is clear that the declaration of *state* does not contain any additional information except the name of a *state*. Taking that information into account, we can write the same state machine without any state declaration as it is shown in the following listing 14.

Listing 14. State machine language example without state declaration

```
trans WaterHigh: Ready -> Running;
trans WaterLow: Running -> Ready;
```

The *Ready* and *Running* states are declared in the place of their usage and the later usage is used as a reference to the existing state. This feature allows simplification of language sentences for common language users, as DSLs usually require syntax, which is more practical than technical.

Our method supports the automatic creation of language concepts in the place of their reference in case of non-existent variable declaration in the language sentence. It can be specified by the *create* parameter in the *@References* annotation (example in listing 15). This parameter is set to *false* by default, therefore a non-existent referenced concept is not created. Setting the *create* parameter to *true* allows the automatic creation of language concepts. The only requirement

is the existence of a constructor with a string parameter and the usage of global scope for referencing as well as identifier uniqueness.

Listing 15. References annotation with create parameter

```
@References(State.class,
            field = "source", create = true)
```

Automatic creation of undefined identifiers is the last phase of our method for resolving identifiers and references. It runs after the phase of resolving references and only if there is at least one *create* parameter with value *true*. After each creation of a new language concept it is required to resolve references as AST has been modified.

VI. Conclusion

Current language tools support creation of languages with one or more methods, but they fail in the area of an exact declaration of references and identifiers in the created languages. A language designer has to implement checking of identifier uniqueness and lookup methods for finding proper referenced identifiers manually, as it is common with other tools such as Beaver, JavaCC, JastAdd, ANTLR or at least programatically select language concepts from the scope using API in Xtext framework.

In this paper we have presented the method for declarative specification of computer languages with identifiers and references. We have discovered and described language patterns for identifiers and references. These language patterns have different possible parameters to adjust their impact. It is possible to define the scope of uniqueness of an identifier. We have described different levels of scopes for referencing other language concepts and we have explained different levels of scoping on illustrative examples. The scoping levels we discovered are global scope, simple local scope and complex scope. Complex scope is used mostly for advanced scoping of variable names in programming languages. During our research, we have analyzed various languages and discovered the pattern for identifiers without explicit declaration of identifier before the usage. In order to cope with such form of languages we provide an option to automatically create language concept on the first occurrence of identifier and to use it as a reference on every other occurrence.

Every mentioned pattern and option has been integrated in our method for working with references and identifiers in computer languages, with a special orientation to DSLs. Our method allows declarative definition of identifiers and references instead of more common imperative solutions of other methods. A summary of our method is a declarative transformation of AST to ASG. The method has been implemented in our YAJCo[1] tool, which can be found on the central Maven repository. All languages mentioned in this paper has been successfully tested in YAJCo and they serve as a proof of concept for the proposed method.

[1]https://code.google.com/p/yajco/

Acknowledgment

This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the artifacts written in domain-specific languages driven by language evolution.

References

[1] A. Kleppe, *Software Language Engineering: Creating Domain-Specific Languages Using Metamodels*, 1st ed. Addison-Wesley Professional, 2008.

[2] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 316–344, Dec. 2005. [Online]. Available: http://doi.acm.org/10.1145/1118890.1118892

[3] M. Tofte, *Compiler generators: what they can do, what they might do, and what they will probably never do*. New York, NY, USA: Springer-Verlag New York, Inc., 1990.

[4] S. C. Johnson, *Yacc: Yet another compiler-compiler*. Bell Laboratories Murray Hill, NJ, 1975, vol. 32.

[5] T. J. Parr and R. W. Quong, "Antlr: a predicated-ll(k) parser generator," *Softw. Pract. Exper.*, vol. 25, no. 7, pp. 789–810, Jul. 1995. [Online]. Available: http://dx.doi.org/10.1002/spe.4380250705

[6] A. Demenchuk, "Beaver-a lalr parser generator," 2006.

[7] M. Mernik, N. Korbar, and V. Žumer, "Lisa: a tool for automatic language implementation," *SIGPLAN Not.*, vol. 30, no. 4, pp. 71–79, Apr. 1995. [Online]. Available: http://doi.acm.org/10.1145/202176.202185

[8] M. Mernik, M. Lenič, E. Avdičaušević, and V. Žumer, "Lisa: An interactive environment for programming language development," in *Compiler Construction*. Springer, 2002, pp. 1–4.

[9] T. Ekman, G. Hedin, and E. Magnusson, "Jastadd," 2008.

[10] M. Eysholdt and H. Behrens, "Xtext: implement your language faster than the quick and dirty way," in *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, ser. SPLASH '10. New York, NY, USA: ACM, 2010, pp. 307–309. [Online]. Available: http://doi.acm.org/10.1145/1869542.1869625

[11] "Xtext @ONLINE," http://www.eclipse.org/Xtext/, Jun. 2013.

[12] J. Paakki, "Attribute grammar paradigmsa high-level methodology in language implementation," *ACM Comput. Surv.*, vol. 27, no. 2, pp. 196–255, Jun. 1995. [Online]. Available: http://doi.acm.org/10.1145/210376.197409

[13] G. Hedin, "Reference attributed grammars," 1999.

[14] T. Ekman and G. Hedin, "Rewritable reference attributed grammars," in *ECOOP 2004–Object-Oriented Programming*. Springer, 2004, pp. 147–171.

[15] G. Hedin, "An introductory tutorial on jastadd attribute grammars," in *Generative and Transformational Techniques in Software Engineering III*. Springer, 2011, pp. 166–200.

[16] J. Porubän, M. Forgáč, and M. Sabo, "Annotation based parser generator," in *Computer Science and Information Technology, 2009. IMCSIT '09. International Multiconference on*, Oct., pp. 707–714.

[17] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual*, 2nd ed. Addison-Wesley Professional, 2010.

[18] M. Fowler, "Inversion of control containers and the dependency injection pattern, jan. 2004," *URL: http://martinfowler. com/articles/injection. html*.

[19] I. R. Forman, N. Forman, D. J. V. Ibm, I. R. Forman, and N. Forman, "Java reflection in action," 2004.

[20] G. Nicol, *Flex: the lexical scanner generator*. Free Software Foundation, 1993.

[21] V. Kodaganallur, "Incorporating language processing into java applications: A javacc tutorial," *Software, IEEE*, vol. 21, no. 4, pp. 70–77, 2004.

[22] T. Parr, *The Definitive ANTLR Reference: Building Domain-Specific Languages*. Pragmatic Bookshelf, 2007.

[23] J. Clark, S. DeRose *et al.*, "Xml path language (xpath) version 1.0," 1999.

[24] C.-Y. Chan, P. Felber, M. Garofalakis, and R. Rastogi, "Efficient filtering of xml documents with xpath expressions," *The VLDB Journal*, vol. 11, no. 4, pp. 354–379, 2002.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1523–1528

SimpleConcepts: Support for Constraints on Generic Types in C++

Reed Milewicz
University of Alabama at Birmingham
Birmingham, AL 35294
rmmilewi@cis.uab.edu

Marjan Mernik
University of Maribor, Slovenia
marjan.mernik@uni-mb.si

Peter Pirkelbauer
University of Alabama at Birmingham
Birmingham, AL 35294
pirkelbauer@uab.edu

***Abstract*—Generic programming plays an essential role in C++ software through the use of templates. However, both the creation and use of template libraries is hindered by the fact that the language does not allow programmers to specify constraints on generic types. To date, no proposal to update the language to provide concepts has survived the committee process. Until that time comes, as a form of early support, this paper introduces SimpleConcepts, an extension to C++11 that provides support for concepts, sets of constraints on generic types. SimpleConcepts features are parsed according to an island grammar and source-to-source translation is used to lower concepts to pure C++11 code.**

***Keywords*—Generic Programming, C++ Templates, C++ Concepts**

I. Introduction

GENERIC programming is made possible in C++ through the use of templates [1]. Templates are language constructs that operate with generic types and that are instantiated as needed during compile-time [2]. Templates are ubiquitous in many C++ libraries and systems, most notably the Standard Template Library (STL), which provides generic implementations of commonly used containers and related algorithms. Essential to the STL are concepts, which are sets of constraints on types [3]. A type is called a model of a concept if that type satisfies all of its requirements, and templates can impose these concepts on their arguments; this is done to ensure type safety. For example, to be able to sort a list, its elements must support the == and < operators, and these requirements are expressed by the concepts EqualityComparable and LessThanComparable. If the concepts associated with a template class or function are not satisfied, then the instantiation of that template code will fail, and a compile-time error will result. It is important to note that concepts are not features of the C++ language, but are rather the products of what Gregor et al. refer to as a "grab-bag of template tricks" [4]. The first issue with concepts as they are currently known is that they do not lend themselves to informative error messages. Violations cannot be reported without exposing the programmer to the details of the implementation. This means that compile-time errors can lead to dense barrages of esoteric error messages that give the programmer little insight into what went wrong. The second issue is that the complex nature of template metaprogramming makes it difficult to map concepts as described in the documentation to the specifics of their implementation. For those who create and maintain generic libraries, this means that it can be extremely difficult to detect bugs and other issues in their code. The root problem is that the C++ language lacks a construct to perform a vital function, and this forces the developers of generic libraries to resort to bricolage, cobbling together a functional equivalent from whatever materials they have at hand. This satisfies the immediate needs of the developers. However, in that concepts do not formally exist, they cannot be formally reasoned about or analyzed. For users, this is what leads to incomprehensible error messages when they misuse templates. For developers, this means that writing and maintaining template code becomes unnecessarily burdensome. The main contribution of this work is the introduction of SimpleConcepts, a lightweight extension to the C++ language to facilitate that development, maintenance, and usage of generic libraries. SimpleConcepts introduces several useful abstractions that perform the same role as concepts via metaprogramming while being easier to write, read, and formally analyze. This paper is organized as follows. In §II, we characterize the previous work that has been done to help modernize concepts in C++. In §III, we describe the problem domain by investigating how concepts work in generic libraries such as the STL, and in §IV we introduce SimpleConcepts, give justifications for our approach, and show how the functionalities provided by concepts via metaprogramming map to the new model. In §V, we provide formalisms for the syntax and semantics of SimpleConcepts. In §VI, we compare our approach to previous and contemporary alternatives. Finally, in §VII and §VIII we provide discussion, conclusions, and plans for future work.

II. Background

The first comprehensive attempt to provide high level language support for concepts was the development of Tecton, a domain-specific language (DSL) for generic programming that was conceived by Stepanov, Kapur, and Musser in the late 1970s [5]. The work that was done on that language fed into the development of the STL by A. Stepanov [6]. After the Hewlett-Packard implementation of the STL was made publically available in 1994, the language evolved into one that specialized in concept specification, as seen in Musser's technical report in 1998 [7]; Tecton sought to provide a language-independent means of describing constraints on generic types. While it did not see widespread adoption, it did provide a formal framework through which concepts could be understood, laying the foundation for further developments. 1998 also marked the birth of the Boost libraries for C++ [8], and two years later the collection was extended to include the Boost Concept Check Library (BCCL), an effort spearheaded by

978-83-60810-53-8/$25.00

Siek [9]. The BCCL is meant to provide clean and accessible mechanisms for programmers to use concepts in generic code, an improvement upon concepts as known in the STL. In the last decade, the goal of researchers shifted away from providing support for concepts in C++ by means of DSLs and library support and towards the incorporation of concepts into the C++ language as an extension. A detailed review of this period has been provided by Voufo and Lumsdaine [10]. The most successful of these movements was ConceptC++, a project which culminated in a proposal to the C++ standardization committee that was not accepted in 2009 for want of simplification and more testing [11] [12]. It is from ConceptC++ and the work done in the intervening period that SimpleConcepts draws much of its inspiration.

III. Domain Analysis

Here we shall provide a description of concepts as they are known in the C++ Standard Template Library. As previously described, the term concept refers to a set of constraints or requirements on types. For instance, the concept EqualityComparable defines what it means for a type T to be comparable for equality. In C++, this means that the expressions `a == b` and `a != b` be valid for any two values `a` and `b` of type `T`. In the STL, concepts are implemented as assertions with the help of macros and made use of through statements such as

```
__STL_REQUIRES(X,_EqualityComparable);
```

which checks to see whether the type of a template parameter X models the concept EqualityComparable. Aside from requiring that a type support certain operations, concepts can also require that certain functions be supported or class members exist. For example, the concept Container requires that its models implement a size() method. Additionally, concepts may require certain associated types be defined (e.g. a size_type for the value returned by a call to a Container's size() method). Lastly, the documentation for a concept may also state that certain invariants must be satisfied (e.g. identity or transitivity), but these are assumed to be the case, and no actual verification occurs. Concepts can be defined as a refinement of any number of previously existing concepts. A refined concept adopts the requirements defined by the concepts that it is refining. For instance, the concept ForwardContainer is a refinement of the concepts Container, EqualityComparable, and LessThanComparable (if its elements model LessThanComparable). That is, a ForwardContainer is a Container that also requires that its elements be comparable to one another.

In order to produce an extension to the C++ language that captures what these concepts do, we have to provide a formalism that allows us to understand concepts independently of their realizations. It would be a mistake to begin with a deconstruction of the syntax and semantics of STL concepts, because we are less interested in what they are; rather, we seek to describe what they are meant to be. We ought to begin with seeing the task of identifying and expressing concepts as a problem domain that happens to intersect with the task of generic programming. This domain is not complete or self-contained as we cannot speak of constraints on types without dealing with the particulars of some type system. However, concepts, as entities in their own right, can be reasoned about. Following the work of van Deursen and Klint [13], we shall

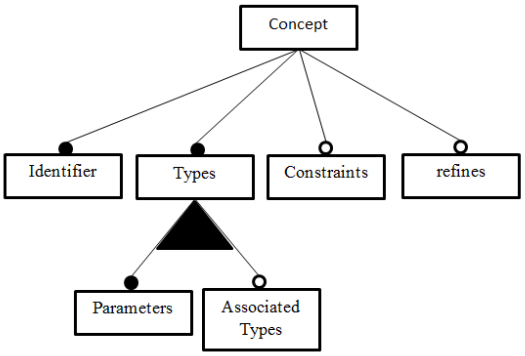


Fig. 1. FDL Diagram for Concepts

give a formal description of STL concepts using the Feature Description Language (FDL). From attempting to derive a formal description of concepts from our informal description, we can discern several truths. First, concepts cannot be anonymous. They must exist prior to and independently of the circumstances in which they are used, and therefore must be named. Second, concepts must have one or more generic type parameters. Concepts connect constraints to types, and a concept that does not do this is an invalid concept. Third, concepts are allowed to have no constraints, that is, they can be empty. A concept with no constraints is trivially satisfied; all types are models of an empty concept. This may seem puzzling at first, but consider that a concept that introduces no new constraints can still be useful if that concept is a refinement of two or more concepts, because it implicitly expresses the union of those concepts. Fourth, a concept may refine one or more other concepts. To refine a concept is to implicitly add its constraints to the concept. Lastly, comparing the informal description of STL concepts and the diagram shown in Fig. 1, one may note that invariants are not listed as a feature of concepts, and this is deliberate. Invariants are the consequences of satisfying both the syntactic and semantic requirements of a concept. To state that invariants are a feature of concepts means that there exists some has-a relationship between the two, when in fact this is not the case. For example, the STL documentation states that the reflexivity of the == operator is an invariant of the concept EqualityComparable, that is, `x == x` for all `x` of type `T` that is a model EqualityComparable. From a mathematical perspective, the invariant naturally follows if we assume the conventional definition of equality. However, in that C++ is a language that allows for operator overloading, knowing that a type supports the `==` and `!=` operators does not tell us whether those operators behave in some prescribed way. From this, we can conclude that there must be a limit to the enforceability of concepts with regards to their semantics. At the very least, finding the means to do so goes beyond the scope of this paper.

IV. Concepts

In SimpleConcepts, concepts are first-class representations of constraints on type parameters of templates. The concepts of SimpleConcepts obviate the need to use template metaprogramming to specify what a template requires of its type

```
concept EqualityComparable<typename T> {
        bool T::operator==(T rhs);
        bool T::operator!=(T rhs);
}
```

Fig. 2. SimpleConcepts concept definition

```
template<typename R> requires EqualityComparable<R>
bool foo(R x, R y) { /* ... */ };
```

Fig. 3. Requires clause

parameters. We designed SimpleConcepts with the following goals in mind:

1) To provide the same functionality as STL concepts while allowing programmers to write concept code that preserves readability and allows the compiler to produce meaningful error messages.
2) To make our extension to the C++ language as lightweight and undemanding as possible, providing more expressive power yet preserving the efficiency of template programming.

A concept definition consists of a declaration and a body containing concept member specifications. Fig. IV depicts the definition of the concept EqualityComparable.

A concept definition establishes what it means for a type T to be comparable for equality. Once a concept has been defined, it can be used in template code by means of a "requires" clause as is shown in Fig. IV. The requires clause places a restriction on the type R to being one which supports the == and != operators. Attempting to use the function foo with any x and y of a type that does not support these operators will lead to a compile-time error, informing the programmer that the EqualityComparable concept was not satisfied.

V. Implementation

As was stated previously, a major consideration in the design of SimpleConcepts was to produce an undemanding extension, one that did not require significant overhaul or that could break existing code. Our approach can be summarized as follows. First, the source code is parsed according to an island grammar that allows us to identify SimpleConcepts features [14]. These constructs are then translated to existing C++11 features, and the resulting source code is passed to an ordinary compiler. There are several advantages to this approach. First, this approach allows us to handle both the syntactic and semantic analysis of concepts without need to modify an existing compiler. Second, using an island grammar allows us to analyze the syntax of concepts in a context-free fashion. Lastly, by means of translational semantics, we are able to express the meaning of concepts in terms of language features whose semantics are already well-known. The remainder of this section addresses the details of this compilation model.

A. Abstract Syntax of SimpleConcepts

SimpleConcepts extends the C++ language to provide two new language constructs: concepts and constrained templates. A concept specifies a set of member function declarations and

```
concept Flammable<typename T>{
        double T::burn();
}
template<typename V> requires Flammable<V>
void makeCampfire(V v){
        double heat = v.burn();
        /*...*/
}
```

Fig. 4. Example of SimpleConcepts in action

```
template<typename T>
struct Flammable {
        Flammable_Requirements<T> rq0;
}

template<typename V>
void makeCampfire(V v){
        while(false) { Flammable<V>(); }
        /*...*/
        double heat = v.burn();
        /*...*/
}
```

Fig. 5. The first layer of the translation

associated types, and a list of other concepts that are refined by it. A constrained template is one that has a concept requirement clause, dictating what restrictions are placed on the template's parameters. Table I gives the abstract syntax and the syntactic domains for SimpleConcepts.

B. Concrete Syntax of SimpleConcepts

An island grammar is a context-free grammar which describes some subset of the features of a language and uses catch-all productions to ignore all else; the name of this technique is derived from the view that the features we are interested in capturing with the grammar are "islands" amidst a vast sea of other language features. In this case, we are only interested in parsing concept definitions and their uses so that we can perform the necessary translations, and therefore a grammar that allows us to identify salient features while skipping the rest is ideal. Table II gives an EBNF grammar for SimpleConcepts.

C. Summary of Translation Model

Our approach is based on the work done by Valentin and Magne [15], which describes a means of converting ConceptC++ code to pure C++03 code by translating the concepts of ConceptC++ into sets of class templates. Here we shall use a toy problem as a vehicle to explore the translation scheme. Consider the code fragment in Fig. V-C.

A type T models the concept Flammable if it has a member function called burn that takes no arguments and returns a double. In other words, Flammables burn, and burning produces an amount of heat expressed as a double. Below the concept definition we see an example of a function template that uses the concept Flammable. We shall describe, step by step, the translation process. The translation from concept code results in what can be seen as three distinct layers of template code. Fig. V-C shows the first layer of the translation.

The struct Flammable has, as a member, another struct representing the requirements associated with the concept

TABLE I. ABSTRACT SYNTAX AND SYNTACTIC DOMAINS OF SIMPLECONCEPTS

Type Variables	$\alpha \in TyVar$
Concept Names	$s \in CName$
Member Names	$f, a_t \in MemName$
Concept	$C \in \mathbb{C} := conceptC_{id}\ \{B\}; \| \ conceptC_{id}\ requires R\ \{B\}$
Concept Identifier	$C_{id} := s < P >$
Concept Parameters	P := $\alpha P \| \alpha$
Refinement Clause	R := $C_{id} R \| C_{id}$
Concept Body	B := $MB \| \epsilon$
Concept Member	M := $f\ func; \| typename\ a_t$
Constrained Template	$T \in \mathbb{T}_c := template\ < params >\ requires\ R\ template - body$

Here $\mathbb{C}$ refers to the set of all concepts, and $\mathbb{T}_c$ refers to the set of all constrained templates.

TABLE II. CONCRETE SYNTAX OF SIMPLECONCEPTS

<concept-id> := <concept-name> "<" <concept-parameter list> ">"
<concept-name> := **<identifier>**
<concept-definition> := "concept" <concept-id> <requires-clause>? <concept-body> ";"?
<concept-body> := "{" <concept-member-specification>? "}"
<concept-member-specification> := <concept-member-specifier> <concept-member-specification>?
<concept-member-specifier>:= <associated-function> | <associated-type>
<associated-function> := **<function-definition>**
<associated-type> := **<typename-specifier>**
<requires-clause> := "requires" <requirement-list> | "requires" "(" | <requirement-list> ")"
<requirement-list> := <requirement> "&&" <requirement-list> | <requirement>
<requirement> := <concept-id>
<declaration> := <concept-definition>
<template-declaration> := "template" "<" **<template-parameter-list>** ">" <requires-clause>?
<concept-parameter list> := **<template-parameter-list>**

Non-terminals that refer to pre-existing C++11 features are in bold for clarity.

```
CREATE_MEMBER_FUNC_SIG_CHECK(burn,double (T::*)(void));
template<typename T>
struct Flammable_Requirements {
        static_assert(has_member_func_burn<T>::value,
            "The member function 'burn' is not available"
            " or does not match signature.")
};
```

Fig. 6. The second layer of the translation

Flammable. The requirements are separated from the concept itself in order to support refinement; a concept that refines another concept "inherits" the requirements from its ancestor, and the ancestor's requirements struct is listed there as well. Attempting to instantiate the Flammable template will require the instantiation of the template Flammable_Requirements. If that instantiation fails, the code will fail to compile. To implement constraints on template functions, we cause trigger the instantiation of Flammable in the template function as seen in the template function code. The call to the constructor of Flammable is placed in an unreachable block of code to guarantee that no run-time overhead will result. Now we examine the second layer of the translation: expressing the requirements enumerated by a concept. A Requirements struct contains a set of static assertions that express the requirements that must be fulfilled by a type or set of types in order to model a concept. These assertions are checked when the compiler attempts to instantiate the Flammable_Requirements template. The code that we generate for the definition of Flammable_Requirements is provided in Fig. V-C.

The macro CREATE_MEMBER_FUNC_SIG_CHECK provides template code necessary to check the existence of a member function that matches both the name and the signature specified by the concept. The instantiation of has_member_func_burn will always succeed, but its member 'value' will be true if and only if T has a burn method that matches the signature specified in the concept definition. This, in turn, determines whether the corresponding static assertion succeeds or fails. The code generated by the pre-processor is shown in Fig. V-C.

This then is the third and innermost layer of the translation. Our mechanism verifies the existence of the member function burn in a way that gives a true or false value which is used in the static assertion. This allows us to report meaningful error messages.

D. Semantics of SimpleConcepts

We shall describe the translational semantics of SimpleConcepts. In that concepts and constrained templates are ultimately converted to template code, the semantics of concepts are a subset of the semantics of templates. With that in mind, we adapt the work done by Siek and Taha [16] to provide formalisms to describe the semantics of C++ templates. By providing a mapping from the abstract syntax of SimpleConcepts to that of C++ template code, we can then make the jump to the semantics. We represent our translation function as a set of functions that map SimpleConcepts code to C++11 code. The first subset of these translation functions, defined in Fig. V-D, describes the translation of concepts and their members (note that $\mathbb{T}$ refers to the set of all C++11 templates).

```
template<typename T, T>
struct match_signature : std::true_type {};
template<typename T, typename = std::true_type>
struct has_member_func_burn : std::false_type {};
template<typename T>
struct has_member_func_burn<T, std::integral_constant < bool , match_signature<fSig, &T::fName>::value >> : std::true_type {}
```

Fig. 7. The innermost layer of the translation.

$$
\begin{aligned}
F_C(C \in \mathbb{C}) = \{ & C_{id} \{ \{ requirements_{C_{id}} \langle P \rangle \} \\
& \cup \{ requirements_r \langle P_r \rangle \forall r \in R \} \in \mathbb{T}, requirements_{C_{id}} \langle P \rangle \{ F_{m1}(B) \} \in \mathbb{T}), F_{m2}(B) \} \\
F_{m1}(B) = \{ & static\_assert(has\_member\_func\_fname \langle P \rangle :: value, errormsg) \forall f \in R \} \cup \{ typedef a_t \forall a_t \in R \} \\
F_{m2}(B) = \{ & CREATE\_MEMBER\_FUNC\_SIG\_CHECK(fname, fsignature) \forall f \in B \}
\end{aligned}
$$

Fig. 8. A formal description of the translation function

```
// function member checking template macros here,
// one for each associated function f ∈ B
template<P>
struct ConceptName_Requirements {
  // A list of static assertions, one for each associated function f ∈ B,
  // and a list of typedefs, one for each associated type definition a_t ∈ B
};
template<P>
struct ConceptName {
        ConceptName_Requirements<P> rq_c;
        id_0_Requirements<P_0> rq_0;
        // . . . for every concept refined by this concept . . .
        id_n_Requirements<P_n> rq_n;
};
```

Fig. 9. Summary of the output of the translation function

In terms of the concrete syntax, this translation amounts to the code fragment depicted in Fig. V-D.

The second subset of the translation functions transform constrained function and class templates into C++11 legal templates. These can be summarized as follows: for a constrained function template, our translation moves the concept requirements into the body of the function as calls to constructors; for a constrained class template, the concept structs are made members of the class. The result of this translation scheme is that instantiations of constrained class templates amount to a chain of instantiations that perform the necessary checks to confirm that the types involved are models of the concepts required. Concept instantiations lead to requirement template instantiations, and those in turn lead to member function checking template instantiations. With that, the only way to make use of a constrained template class or function is to supply it with valid types, or compile-time errors will result. As for the concepts themselves, whose capabilities are limited to checking the existence of function members, we know that our translation does exactly that and nothing more.

VI. Related work

As has been stated in previously, the absence of concepts in C++ is not a new problem; each attempt at a solution has built upon the groundwork laid by predecessors. In this section, we shall attempt to compare and contrast past and current approaches with SimpleConcepts.

A. Concepts Lite

At this time, there exists another proposal to provide support for C++ concepts by Sutton and Stroustrup known as Concepts Lite [17] [18]. In this section, we shall attempt to compare and contrast that approach with SimpleConcepts.

Sutton and Stroustrup, the creators of Concepts Lite, have summarized their vision as "concepts = constraints + axioms" [19]. Concepts are abstract predicates that represent sets of requirements on generic types. These requirements can either be constraints or axioms. Constraints are syntactic requirements on the properties of generic types, which are checked at compile-time, and axioms are semantic requirements, analogous to the invariants of the STL documentation. As of the latest proposal, Concepts Lite supports constraints, but does not yet support axioms or concepts. With the understanding that the specifics of this proposal may be changed in the near future, we note the key differences between the two approaches.

First, while both SimpleConcepts and Concepts Lite share a similar notion of constraints, they differ greatly in terms of their implementation. In Concepts Lite, a constraint predicate is defined as a function template that contains a constant expression, referred to as a "use pattern". A type or set of types satisfies a constraint if the template can be legally instantiated, which is to say that the constant expression is valid. For example, a type `T` is Addable provided that for expression `a + b` is valid for any `a` and `b` of type `T`. In the syntax of Concepts Lite, this constraint might be expressed as follows:

```
   template <typename T>
constexpr bool Addable() { return
__is_valid_expr{bool={declval<T>() +
declval<T>()} }
```

This approach differs from that of SimpleConcepts in three ways. First, Concepts Lite decouples concepts and constraints, which allows us to define Addable as a stand-alone constraint; in SimpleConcepts, which keeps constraints and concepts coupled, Addable would be expressed as a concept with a single constraint member. Second, Concepts Lite requires extensions to the compiler to support new intrinsics such as `__is_valid_expr`; SimpleConcepts uses a preprocessor to

lower concepts to existing C++11. Third and most importantly, whereas Concepts Lite uses constant expressions to define constraints, SimpleConcepts relies on function signatures.

Next, in contrast to Concepts Lite, this paper rejects the inclusion of axioms in its definition of concepts as going outside of the role that concepts are intended to fulfill, which is to provide compile-time support for templates. According to its authors, axioms are not statically evaluable, which implies that the question of whether a generic type truly models a concept, both syntactically and semantically, cannot be decided at compile-time. While we recognize the fundamental relationship between the two kinds of requirements, and we appreciate the simplicity and beauty of an approach that unifies them, we do not see a pressing need to incorporate axioms.

B. ConceptsC++

SimpleConcepts draws inspiration from the ConceptsC++ proposal, and in a certain sense can be seen as an evolution of it. In ConceptsC++, as in SimpleConcepts, concepts are sets of requirements, which can be expressed as signatures and associated types, and these concepts can be refined from other concepts, and they can be imposed upon generic types through the use of requirement clauses. Key to ConceptsC++ is its emphasis on retroactive modeling, that is, the ability to extend types to model concepts without modifying those types. This is accomplished through the use of concept maps, which detail how a type satisfies the requirements of a concept; concept maps can contain implementations of the functions required by the concept, either providing new functionalities or overriding exisiting ones. In contrast, SimpleConcepts does not support retroactive modelling, and does not support concept maps. There are also several other differences between the two approaches, which we shall list here:

- ConceptsC++ allows for non-member associated functions. The implementations of these functions can either be defined via a concept map or a default implementation can be provided in the concept itself. SimpleConcepts, meanwhile, requires that all associated functions be member functions.
- ConceptsC++ distinguishes between refinement clauses, and associated requirements, both of which allow a concept to "inherit" requirements from other concepts. In SimpleConcepts, no such distinction is made; a concept can have a requires clause that lists all of the other concepts that it draws requirements from.
- ConceptsC++ supports axioms. As was explained in the previous subsection, SimpleConcepts does not support this language feature.

VII. Discussion

Our approach is not without limitations. Most notably, when translating refined concepts or constrained templates, it is assumed that the concepts that are being refined or used already exist, but this is not guaranteed to be the case. Referencing a non-existent concept will lead to a compile-time error, but that error is reported in terms of the translated code rather than the original source code, which could be problematic.

VIII. Conclusion and Future Work

It has been shown that C++ lacks language support to specify constraints on generic types, and that this lack is the underlying cause of many difficulties for both the users and developers of generic libraries. To that end, we introduced SimpleConcepts, an extension to the C++ language to provide such support. Our approach uses source-to-source transformations to provide an extension that is intended to be compatible with pre-existing C++11 compilers, while providing simple but powerful abstractions to aid in the design and use of C++ template libraries.

Moving forward, if the completed Concepts Lite is incorporated into the next iteration of the C++ language, then we shall turn our attention towards providing a formal analysis of such C++ concepts. In the short term, we hope to release a compiler front-end to provide experimental support for SimpleConcepts.

References

[1] B. Stroustrup, *The C++ programming language; 4th ed.* Addison-Wesley, 2013.

[2] A. Stepanov and P. McJones, *Elements of Programming*. Addison-Wesley, 2009.

[3] M. Austern, *Generic programming and the STL: using and extending the C++ Standard Template Library*. Addison-Wesley, 1998.

[4] D. Gregor, B. Stroustrup, J. Järvi, and G. D. Reis, "Concepts: Linguistic support for generic programming in C++," in *SIGPLAN Notices*. ACM Press, 2006, pp. 291–310.

[5] D. Kapur, D. R. Musser, and A. Stepanov, "Tecton: A framework for specifying and verifying generic system components," 1983.

[6] A. Stevens, "Al Stevens interviews alex stepanov," 1995.

[7] D. Musser, "Syntax of the tecton language," 1998.

[8] B. Dawes, "Proposal for a C++ library repository web site," 1998. [Online]. Available: http://www.boost.org/users/proposal.pdf

[9] J. Siek and A. Lumsdaine, "C++ concept checking: A better practice for C++ programming," 2001.

[10] L. Voufo and A. Lumsdaine, "A uniform terminology for C++ concepts," Indiana University Technical Report, Tech. Rep. TR 703, January 2013.

[11] D. Gregor and B. Stroustrup, "Proposed wording for concepts (revision 3)," no. N2421=07-0281, 10/2007 2007. [Online]. Available: http://www.open-std.org/JTC1/sc22/wg21/docs/papers/2007/n2421.pdf

[12] D. Kaley, "Bjarne stroustrup expounds on concepts and the future of C++," 2009.

[13] A. van Deursen and P. Klint, "Domain-specific language design requires feature descriptions," *Journal of Computing and Information Technology*, vol. 10, p. 2002, 2001.

[14] L. Moonen, "Generating robust parsers using island grammars," in *The 8th Working Conference on Reverse Engineering*. IEEE Computer Society Press, 2001, pp. 13–22.

[15] D. Valentin and H. Magne, "Concepts as syntactic sugar," in *SCAM*, 2009, pp. 147–156.

[16] J. Siek and W. Taha, "A semantic analysis of C++ templates," 2006.

[17] B. Stroustrup, A. Sutton, L. Voufo, and M. Zalewski, "A concept design for the STL," ISO/IEC JTC1/SC22/WG21—The C++ Standards Committee, Tech. Rep. N3351=12-0041, January 2012.

[18] A. Sutton and B. Stroustrup, "Concepts lite: Constraining templates with predicates," 2013.

[19] ——, "Design of concept libraries for C++," in *Proceedings of the 4th international conference on Software Language Engineering*, ser. SLE'11. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 97–118.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1529–1532

Concern-oriented Source Code Projections

Matej Nosáľ, Jaroslav Porubän and Milan Nosáľ
Department of Computers and Informatics
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
Email: matej.nosal@gmail.com, {jaroslav.poruban,milan.nosal}@tuke.sk

***Abstract*—The quality of the source code structure is a matter of the point of view, one programmer might consider one structure the best, the other not. A concrete structure can help in certain situations with the program understanding. Therefore we propose using dynamic structuring that allows assigning multiple structures to one source code to aid program comprehension. Concern-oriented source code projections facilitate this dynamic structuring expressed by custom metadata and provide multiple views of the source code that reflect logical structures provided by the dynamic structuring. This way in a specific situation a programmer can get a structure (by a view) that meets his/her current needs the best.**

I. Introduction

PROGRAM comprehension is a process of retrieving information and knowledge about a software system by studying its source code. Software system maintenance and evolution consumes up to 80 percent of system's lifetime [8]. Program comprehension tries to reduce this time. However, a more radical solutions, like for example literate programming [7] or elucidative programming [9], were not adapted in the industry, probably because they were too distant from the industrial practice.

We recognized that good design and source code structure are properties of *the point of view*. We believe that one static structure that is prevalent nowadays is not sufficient. Concern-oriented source code projections (or shortly code projections) contribute to the field of program comprehension by providing means to simultaneously express and use multiple structures of the source code.

II. Motivation

OOP uses classes and techniques such as dynamic binding to increase modularity, but on the other hand it also tears the source code structure to smaller parts (to submit it to single responsibility principle). AOP goes even further; it tears the structure even more according to concerns. It may help with modularity, but since the code as a whole (and thus the sequence of instructions) is scattered in more files, it is harder to follow the program logic. Considering which approach is better depends on the qualities taken into account – it is a matter of the point of view.

The same way it is difficult to find the best design, the best structure in a given paradigm. The quality of a design is a matter of point of view too. Let us consider a simple explanatory example from the AOP. An OOP method from listing 1 does some work and afterwards logs the process to the standard output.

Listing 1. Simple `doSomething()` method that is logged to standard output

```
public void doSomething() {
    ...
    // logging to standard output
    System.out.println
       ("Something was done!");
}
```

The AOP divides the source code into concerns that are implemented by aspects. Here a programmer Jack will identify a concern of logging in the system and will create a new aspect that will implement the logging concern. However, a programmer Jill will identify a concern of printing to standard output (for example to ensure that she will be able to easily switch from standard output to some other interface). So she would want to put the same line of code to the standard output aspect. So in this example Jack is interested in the logging concerns, while Jill in printing to standard output concern. However, these two concerns do not have to cover the same source code. Printing to standard output does not necessarily be logging. In this scenario Jack would say that creating logging aspect is better structuring that creating printing to standard output aspect, while Jill would say exactly the opposite.

Currently the source code has to have exactly one design, one structure. The concerns cannot overlap. The classes cannot either. Multiple design decisions can have good reasoning, but none of them tackles all the problems – e.g., sometimes the OO structuring is more useful, other times AOP structuring is advantageous. Usually it depends on whether the target of interest is a feature (OOP) or a concern (AOP).

The problem is based on the observation that the "best" structuring of the source code depends on the point of view. There are three main factors that influence the current best structuring:

- *Person* – each person has his/her own experience and opinion. Therefore each person has his/her own point of view.
- *Time* – even one person changes his/her opinion in time and in consequence his/her point of view. Usually the early structuring of the program evolves over the time to the required one even when there is only one programmer that authors it.

978-83-60810-53-8/$25.00

- *Character of the problem/solution* – of course the structuring also closely depends on the character of the problem or the solution. Not only the people involved in the software development do evolve in time, but also the character of the problem.

III. Concern-oriented Source Code Projections

In our work we recognize that the problem of multiple points of view is a consequence of *static structuring* of the source code. The problem of current approaches such as the AOP or OOP is that they allow the structure to meet some concrete needs, but does not support easy adaptation to new needs. Metaphorically speaking, using AOP instead of OOP is analogical to tearing down a house and building it again rotated in 90 degrees just to provide a view from side. Wouldn't it be more effective if the house would stay untouched and instead a programmer would walk around the corner? The same way as in the analogy, we don't want to change the building process, we just want to provide a programmer with a view that he/she wants.

A. Static vs. Dynamic Structuring

The problem is current technologies support merely one structure of the source code. This structure has to fully describe the system and does not allow any duplication of code. If current structure of the source code is not viable for current needs, the programmer has just bad luck. He/she has to work with the current structure, or refactor it and hope that the original structure won't be needed later. Or after refactoring the code[1] he/she has to hope that there will not be a need of the original aspect.

To deal with these problems we propose to use *dynamic structuring*. By dynamic source code structuring in this context we mean a case when a source code of one system has multiple different structures at the same time. In a particular situation the programmer would be able to choose the structure that currently the most relevant. This concept of dynamic structuring changes the role of refactoring. Instead of refactoring in sense of dropping the old structure and building a new one, with dynamic structuring a programmer is able to arbitrarily add a new structure to the source code or remove an existing one.

The usual representation of the structure is the source code itself. This will not suffice in case of dynamic structuring. It would be too difficult and cumbersome to write multiple versions of system source code and to keep them consistent through the time. Instead of this "physical" structures' representation dynamic structuring should use logical representation. Source code itself would be written (and stored) only once using some *base structure*. Adding new structures should be done by adding metainformation about their new relationships and properties.

[1]That is not an easy task even with current tool support for code refactoring. Tools support automation of merely trivial tasks such as changing a name of a variable, etc. However here we talk about structure on higher abstraction level, like the choice of using inheritance instead of composition.

The idea of dynamic structuring that allows having multiple different source code structures at the same time is the core of the concern-oriented source code projections method. We consider this idea the main contribution of our work.

B. Views

Code projections are based on dynamic structuring of program's source code. These structures have to be properly presented to programmer; otherwise they would be useless for program comprehension. Code projections map a set of base source code structures to a set of *views*. A view is an abstract structure of source code that is presentable to programmer. A view does not have to fully describe the system and multiple views can overlap (one code fragment can be a part of multiple views). The *Identity* projection defines a view that is identical to base source code structure; therefore it has to fully describe the system.

A single view consists of source code fragments that are somehow related. We will call these fragments view members. Relations between view members may be explicitly expressed in the view – a view can be graphical.

A code projection is specified by a sentence in a program query language[2]. Practically any PQL can be used; however, it has to support querying custom metadata too.

A programmer creates a *projection query* that specifies which concerns are relevant to his/her current situation. Projection queries can be shared and stored for later reuse, or modified if their current user is not satisfied with their current state. The concept of the code projections is outlined in figure 1.

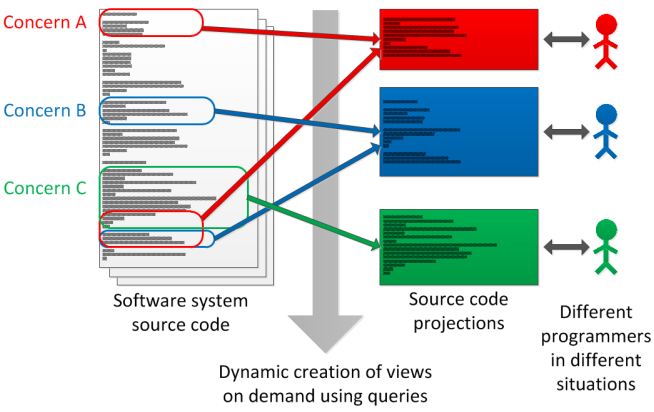


Fig. 1. The concept of the concern-oriented source projections

Our hypothesis in this work is that *current tools don't support flexible creating of views for viable price by using custom metadata*.

C. The Role of the Metadata

We will use term *software system metadata* (from now on only metadata) for the total sum (the set) of what one

[2]There are already programming query languages (PQL) in the world, this is not a new idea. However, PQLs are usually used to do source code checking (e.g., [4]). Our method uses PQL to provide a code projection.

can say about any program element, in a machine or human understandable representation.

Current IDEs provide programmers with code projections that operate on *intrinsic* metadata[3]. Navigator view uses the class intrinsic metadata to present the class members. Navigating to implementation through Ctrl and left mouse click uses the program element identifier. IDEs can use inheritance hierarchy to show implementations of interfaces or classes. Find Usages view uses also program element identifier to bind the implementation to its usages and provides a very useful projection of the source code.

In an example with Jack and Jill (listing 1) Jill would get desired result merely using intrinsic metadata. She can just query for any `System.out.println` call.

However, a situation would change if the logging was encapsulated in a `Logger` class (listing 2) that would provide a `PrintStream` that should be used for logging.

Listing 2. Logger implementation

```
public class Logger {
   private static PrintStream stream
      = System.out;

   public static getStream() {
      return stream;
   }
}
```

Listing 3 shows a modified `doSomething()` method with obscured printing to standard output. In this case it would be much more difficult to create a query that could find all the lines printing to standard output.

Listing 3. Obscured usage of standard output for logging

```
public void doSomething() {
   ...
   // logging to standard output
   Logger.getStream().println
      ("Something was done!");
}
```

Therefore we propose the utilization of the *custom* metadata to provide more information about program elements and to use that as a basis for different source code projections. Custom metadata are a tool that can be used to enrich the source code with metainformation about the semantic or design intents of the program elements at a higher abstraction level than the GPL is itself. In this way projections can use this metadata to present code to a programmer at a higher abstraction level too. A projection maps *concern-enriched* source code to a view. For listing 3 one could use for example simple marker annotations `@Log` and `@WritesToStandardOutput`.

D. The Role of the IDE

To provide code projections there has to be a tool that would be able to create a view while managing the source code in its base structure. In case of code projections we see as a best option utilization of the *Integrated Development Environment* (IDE) thanks to its approach to handle language in its infrastructure (considering IDE is an integrated set of language tools).

IDE infrastructure usually works with three language representations shown in figure 2, Notation, Model and View level. We want to utilize the editor to dynamically modify a view of the language.

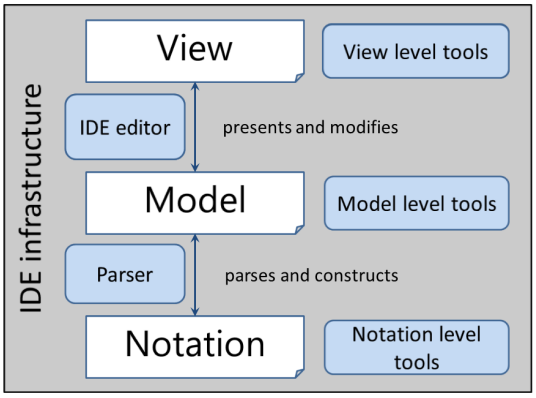


Fig. 2. Language representations in IDE infrastructure

If we will return to the analogy from the introduction to section III, concern-oriented source code projections provide a cheaper alternative to tearing down a house and building it in another angle. Instead of this invasive and inflexible solution code projections propose to change the view of an architect (analogy of programmer). Instead of moving the building merely the architect is moved to see what he/she needs to see.

IV. Related Work

We have applied a similar approach to *reduce the syntactic noise in internal domain-specific languages* in our previous work [6]. We were able to remove some undesired syntactic constructs (such as import section, class declaration, etc.) from the internal DSL based on Java language.

In modern IDEs there are many *standard projections* like the Navigator, TODOs, and others that we mentioned in section III-C. All these use projections to provide different views on the source code to provide a better orientation in the code.

Similar approach is used by Desmond et al. [1] in so called *Fluid source code views*. They allow viewing method bodies in place of their calls, thus reducing the need of browsing the source files. It is kind of similar to Go To Declaration projection of current IDEs, however using fluid source code views the body is shown directly in place of call using a tooltip.

Intentional source code views [5] are sets of related program elements that share some intention. In this sense they are

[3]Intrinsic metadata are standard metadata that define a program element. For example, for a class it is its name, its superclass, its interfaces, its methods and its attributes.

very similar to concern-oriented code projections. In Intentional views the intentions of the source code are specified using logic metaprogramming. Although they are close to our approach by providing means of defining architectural and conceptual information about source code, they differ in few rather important aspects. Intentional views require knowledge of logic metaprogramming. It is hard to expect every programmer to be a logic programmer. In our code projections we want to utilize common programmer's natural environment – code projections are to be made integral part of a modern IDE. Intentional views use code conventions that tend to be fragile (see [3]). And our projections can be used to edit the code, while Intentional views are read-only.

Source code annotations are used in [2], where Eisenberg et al. propose *simple edit-time metaobject protocol* that uses annotations as extensibility point of language. The editor is composed of multiple figures that have knowledge of what and how can be edited in them. These figures are configured by annotations.

V. Conclusion and Future Work

In this paper we argue that only one static structuring is not sufficient for software system source code. The quality of the source code structure is a matter of view, one programmer might consider one structure the best, the other not. A concrete structure can however play a significant role in program comprehension, since the point of view of looking at a code depends on the goal the programmer needs to achieve.

We presented the idea of so called dynamic structuring that allows assigning multiple structures to one source code. Our idea builds upon using logical structures that are expressed by metadata. When a new structuring is needed, it can be simply added to source code instead of overriding the existing one. Code projections utilize the view level of the language representation in the IDE infrastructure to reduce the price of their implementation and to still provide a modern professional IDE.

The *code projections* and *dynamic structuring* are the main contributions of the paper. Their purpose is to aid program comprehension. They can play significant role in documenting the source code, system maintenance and evolution.

Although we believe that dynamic structuring along with code projections may be a significant contribution to software engineering, we are aware of some problems with it. These problems we see as a space for our future work.

Dynamic structuring allows multiple, possibly uncountable, structures of the one source code. As we already argued, a good structure is a matter of the point of view. Therefore they may be just as many "good" structures as there are people working on the system. The *Babel effect*[4] is a consequence of the freedom in specifying the structure. Everybody has his/her own opinion and then it is hard to communicate. If two people are looking at two different views, it may be hard if not impossible to unambiguously talk about a specific source code fragment.

The second problem is the *price* of creating a projection. The source code annotations, or in general concern metadata only hardly can be created automatically. The source code has to be annotated explicitly either by it author or another programmer that recognize a need for a new structure. In this case a programmer has to consider the price annotating the source code to provide the space for code projections and the chance that the annotations will be reused later. If there is no real chance that the projection will be used, there is no good reason to create it. This is mainly a problem of development phase of software lifecycle, since the design decisions are made and the intended semantic properties are clearest in this phase.

[4]Named after famous story about building the Tower of Babel from the Bible.

Acknowledgment

This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the Artifacts Written in Domain-specific Languages Driven by Language Evolution.

References

[1] Michael Desmond, Margaret-Anne Storey, and Chris Exton. Fluid source code views. In *Proceedings of the 14th IEEE International Conference on Program Comprehension*, ICPC '06, pages 260–263, Washington, DC, USA, 2006. IEEE Computer Society.

[2] Andrew D. Eisenberg and Gregor Kiczales. A simple edit-time metaobject protocol: controlling the display of metadata in programs. In *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*, OOPSLA '06, pages 696–697, New York, NY, USA, 2006. ACM.

[3] Gregor Kiczales and Mira Mezini. Separation of concerns with procedures, annotations, advice and pointcuts. In *Proceedings of the 19th European conference on Object-Oriented Programming*, ECOOP'05, pages 195–213, Berlin, Heidelberg, 2005. Springer-Verlag.

[4] Michael Martin, Benjamin Livshits, and Monica S. Lam. Finding application errors and security flaws using pql: a program query language. *SIGPLAN Not.*, 40(10):365–383, October 2005.

[5] Kim Mens, Bernard Poll, and Sebastián González. Using intentional source-code views to aid software maintenance. In *Proceedings of the International Conference on Software Maintenance*, ICSM '03, pages 169–, Washington, DC, USA, 2003. IEEE Computer Society.

[6] Milan Nosál', Jaroslav Porubän, and Matej Nosál'. Reducing syntactic noise in internal domain-specific languages. In *Proceedings of CSE 2012: International Scientific Conference on Computer Science and Engineering*, CSE 2012, pages 111–118, 2012.

[7] James Dean Palmer and Eddie Hillenbrand. Reimagining literate programming. In *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*, OOPSLA '09, pages 1007–1014, New York, NY, USA, 2009. ACM.

[8] Michal Vagač and Ján Kollár. Improving program comprehension by automatic metamodel abstraction. *Computer Science and Information Systems*, 9(1):235–247, 2012.

[9] Thomas Vestdam. Elucidative programming in open integrated development environments for java. In *Proceedings of the 2nd international conference on Principles and practice of programming in Java*, PPPJ '03, pages 49–54, New York, NY, USA, 2003. Computer Science Press, Inc.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1533–1536

Teaching Programming through Problem Solving: The Role of the Programming Language

Nikolaos S. Papaspyrou
Email: nickie@softlab.ntua.gr

Stathis Zachos
Email: zachos@cs.ntua.gr

School of Electrical and Computer Engineering
National Technical University of Athens
Polytechnioupoli, 15780 Zografou, Athens, Greece

***Abstract*—In this short paper, we advocate the importance of problem solving for teaching "Introduction to Programming", instead of merely teaching the syntax and semantics of a programming language. We focus on the role of the programming language used for an introductory course. For this purpose we propose CAL, a C-like algorithmic language, which is essentially a well-defined and behaved subset of C with a small number of modest, "educational" extensions. We present the design rationale for CAL, its main features, syntax and illustrative examples.**

I. Introduction

IN THIS short paper, we present our experiences with teaching programming through *problem solving* in the School of Electrical and Computer Engineering of the National Technical University of Athens. We focus on the role of the programming language for this purpose and describe the approach that we have taken. Let us begin with two observations:

1) In some students' minds, algorithmic programming is strangely enough an intellectual process that is not connected to everyday problem solving.
2) Students often have no sense of what is good and what is bad in programming, even after taking a number of courses on the design of algorithms and complexity.

We believe that these are both due to the way we teach students how to program. Starting from secondary school, students often begin by learning a Pascal-like programming language. They learn to use variables, assignments, control-flow statements, arrays. They do not learn, however, because we do not teach them early enough, *what these should be used for!* Young children realize early that they need to solve problems; they are hungry and they want their parents to feed them, they want to play with that shiny car in the toy shop's window, etc. Later on, they learn to speak and use the *language* to communicate their needs. Children learn to speak after they know what they want to say! Why do we teach programming languages to students before they know what to use them for?

The main goal of our proposal is a quick introduction to programming for absolute beginners. Such an introduction would be useful for teenagers in high-school, who may not continue to be programming specialists but who want to support their literacy in mathematics by hands-on attractive algorithmically solvable problems. It would also be useful to first-year university students who have (somehow) escaped a proper exposure to programming in high school (and this is the majority of our students). Thus, this goal translates to:

- a quick educational introduction to self-evident programming concepts and tools, without cryptic, hardware-dependent and special purpose structures; but also
- fluency in a programming language which can be easily extended and/or modified to a language that is currently useful in practice, without a total rethinking of the basic algorithmic techniques.

Our historic prototype for a self-evident educational programming language is of course Pascal [10], whereas programming languages that are currently useful in practice are of course C [7], [5], C++ [9] and Java [2].

II. The Role of the Language

Since the 1950s, scores of different programming languages have been designed and implemented and many more are yet to come. Those with a relative experience in the field will agree that there is no such thing as "the best programming language" and this is what we need to explain to our students early on. Some languages are better than others *for some specific purpose* and indeed: (a) some specific languages are almost exclusively used for some specific purposes, and (b) for some specific purposes people use almost exclusively some specific languages.

For example, C [7], [5] is a very good language for systems programming. It is a low-level language, offering programmers the opportunity to directly interact with the hardware, but not the best language for numerical and scientific computing today. We believe that C is an inappropriate language for teaching "Introduction to Programming". Some of its characteristics are so low-level that tend to focus on the hardware, instead of on the algorithms. When you start learning how to program, you don't need twelve different types for integer numbers (including characters and Boolean values) and three more for real (floating-point) numbers. You don't need a `for` statement so powerful that you can use it to implement a binary search algorithm in just one line. You don't need to struggle to understand the meaning of `x = x++;` (most people, including some who teach C, think that it does something although opinions vary when it comes to what exactly this is; the truth is that this statement is illegal, or causes "undefined behaviour"

978-83-60810-53-8/$25.00

as the ANSI C standard puts it, because the value of variable `x` changes twice between two successive sequence points).

Pascal [10], [6] is arguably one of the best programming languages for teaching purposes. It is a concise, general purpose language which supports a systematic, structured and algorithmic approach to problem solving. Programs in Pascal are usually easy to read and understand with a clear structure that favours stepwise refinement. The language helps programmers to avoid programming mistakes and to be able to verify the correctness of their programs. On the other hand, Pascal is very little used today by software practitioners.

Java and other object-oriented languages are also poor candidates for teaching "Introduction to Programming". If the focus is on problem solving and algorithmic thinking, such languages add an unbearable level of noise. Using Java, the only logical approach is to teach programming in a purely object-oriented fashion and this necessarily takes the focus away from problem solving, although OOP might be a good choice for a second (e.g., data structures) course.

There is a trend towards Python, in the last few years. Python is a relatively good candidate; its syntax is concise and enforces proper indentation, it supports imperative and object-oriented programming equally well, and it is widely used in the software industry. The drawbacks for Python are: (a) it is dynamically typed and requires very few declarations; this is bad if you want the students to detect programming errors early and to learn to program in a disciplined way; (b) it is so high-level that students do not develop an intuition about how data are represented and how operations are implemented; this is bad if you want the students to understand the connection between the programming language and the underlying hardware; and (c) its data structures are so high-level that algorithmic complexity issues are obscured by the way data structures are implemented; e.g., in Python there are no arrays, but there are lists and dictionaries; however, in a data structure course, all three would need to be covered and with three different implementations, requiring $O(1)$, $O(n)$ and $O(n \log n)$ time for accessing an element, respectively.

The second drawback (b) is also true for functional languages, like Scheme, ML or Haskell, which are also very good from an educational point of view and are indeed used for teaching "Introduction to Programming" in several Computer Science departments [8], [3], [1], [4].

All this said, we decided to design a new educational programming language for an introductory course in which emphasis is on problem solving. However, this educational language will naturally evolve before the students' eyes to a full-scale programming language, useful later on. In the next sections we describe CAL, a *C-like algorithmic language*, starting from the design choices that we had to make, proceeding with the syntax, the main characteristics of the language and concluding with a few examples.[1]

[1]An implementation of CAL, based on GCC and using macros, is available from https://github.com/softlab-ntua/pazcal.

III. The Design of CAL

Disregarding some drawbacks, C is an adequate choice for teaching "Introduction to Programming", with emphasis on problem solving and algorithmic thinking. Its core is a quite simple algorithmic language, easy to explain and use. Moreover, it is a useful language to know, heavily used in practice, either directly or indirectly, through a line of descendants that share a large part of its syntax and semantics (C++, Java, C#, etc.). A list of drawbacks:

- Its syntax and semantics is often cryptic and obfuscated; e.g., allowing side-effects anywhere inside expressions.
- The use of "declarators" (as in `int*(f[3])(int);`) is counter-intuitive and hard to explain.
- Non trivial library functions (e.g., `printf` and `scanf`) are required for beginners to write programs that input and output data. The corresponding header files must be `#include`d. Pointers are required for `scanf`.
- Before the simplest program is written, students must see `int main()` and `return 0;` unless of course we want to teach them to be sloppy from the first lecture...
- The type system allows programmers to deliberately misuse data and to neglect declaring function prototypes. Both are bad from an educational point of view.

We therefore base CAL on an appropriate "educational subset" of C, which we extend with a number of macros, library functions and one extra feature (call by reference) to suit the needs of our introductory course. The result is a language reminiscent of Pascal but with C notation. All extensions are written with *uppercase letters* (e.g., **WRITE**), so that students immediately know if something that they have learnt exists in C or is one of our educational extensions. The main characteristics of CAL, whose complete syntax is defined in figure 1, are the following:

- A program is organized as a set of modules, each consisting of constant and type definitions, variable definitions, routine declarations, routine definitions and (optionally) the body of the main program, which must only be present in one module. The visibility of module definitions is controlled with **PRIVATE** and **extern**.
- There are *functions* and *procedures*, defined with **FUNC** and **PROC** respectively; the misleading type **void** is not used. The main program begins with the special keyword **PROGRAM**.
- The type system is simplified. There are types for Boolean values (**bool**, as in C99, with constants **true** and **false**), integers (**int**), characters (**char**) and real numbers (**REAL**). There are also enumerations, structures and unions, but these must be defined and given a name before they can be used. Arrays and pointers complete the picture of types. However, the syntax for declarators is very simplified in comparison with C; type synonyms (**typedef**) can be used for defining, e.g., double pointers, arrays or pointers, pointers to arrays, etc.
- Operators **NOT**, **AND**, **OR** and **MOD** are synonyms of C's (not so intuitive) standard operators `!`, `&&`, `||` and `%`.

⟨module⟩ ::= (⟨const_def⟩ | ⟨type_def⟩)* (⟨declaration⟩)* (⟨definition⟩)* [⟨program⟩]
⟨declaration⟩ ::= ["**PRIVATE**" | "**extern**"] (⟨var_def⟩ | ⟨routine_decl⟩)
⟨definition⟩ ::= ["**PRIVATE**" | "**extern**"] ⟨routine_def⟩
⟨const_def⟩ ::= "**const**" ⟨type⟩ ⟨declarator⟩ "**=**" ⟨initializer⟩ ("**,**" ⟨declarator⟩ "**=**" ⟨initializer⟩)* "**;**"
⟨type_def⟩ ::= "**typedef**" ⟨type⟩ ⟨declarator⟩ ("**,**" ⟨declarator⟩)* "**;**" | ⟨enum_def⟩ | ⟨struct_def⟩ | ⟨union_def⟩
⟨enum_def⟩ ::= "**enum**" ⟨id⟩ "**{**" ⟨id⟩ ("**,**" ⟨id⟩)* "**}**" "**;**"
⟨struct_def⟩ ::= "**struct**" ⟨id⟩ "**{**" (⟨type⟩ ⟨declarator⟩ ("**,**" ⟨declarator⟩)* "**;**")* "**}**" "**;**"
⟨union_def⟩ ::= "**union**" ⟨id⟩ "**{**" (⟨type⟩ ⟨declarator⟩ ("**,**" ⟨declarator⟩)* "**;**")* "**}**" "**;**"
⟨var_def⟩ ::= ⟨type⟩ ⟨declarator⟩ ["**=**" ⟨initializer⟩] ("**,**" ⟨declarator⟩ ["**=**" ⟨initializer⟩])* "**;**"
⟨routine_decl⟩ ::= ⟨routine_header⟩ "**;**"
⟨routine_def⟩ ::= ⟨routine_header⟩ ⟨block⟩
⟨routine_header⟩ ::= ("**PROC**" | "**FUNC**" ⟨type⟩) ⟨id⟩ "**(**" [⟨type⟩ ⟨formal⟩ ("**,**" ⟨type⟩ ⟨formal⟩)*] "**)**"
⟨formal⟩ ::= ⟨id⟩ ["**[**" "**]**"] ("**[**" ⟨expr⟩ "**]**")* | "*****" ⟨id⟩ | "**&**" ⟨id⟩
⟨type⟩ ::= "**int**" | "**bool**" | "**char**" | "**REAL**" | "**enum**" ⟨id⟩ | "**struct**" ⟨id⟩ | "**union**" ⟨id⟩ | ⟨id⟩
⟨declarator⟩ ::= ⟨id⟩ ("**[**" ⟨expr⟩ "**]**")* | "*****" ⟨id⟩
⟨initializer⟩ ::= ⟨expr⟩ | "**{**" ⟨initializer⟩ ("**,**" ⟨initializer⟩)* "**}**"
⟨program⟩ ::= "**PROGRAM**" ⟨id⟩ "**(**" "**)**" ⟨block⟩
⟨block⟩ ::= "**{**" (⟨local_def⟩ | ⟨stmt⟩)* "**}**"
⟨local_def⟩ ::= ⟨const_def⟩ | ⟨var_def⟩
⟨stmt⟩ ::= "**;**" | ⟨l_value⟩ ⟨assign⟩ ⟨expr⟩ "**;**" | ⟨l_value⟩ ("**++**" | "**--**") "**;**" | ⟨write⟩ "**(**" [⟨format⟩ ("**,**" ⟨format⟩)*] "**)**" "**;**"
| "**FOR**" "**(**" ⟨id⟩ "**,**" ⟨range⟩ "**)**" ⟨stmt⟩ | "**while**" "**(**" ⟨expr⟩ "**)**" ⟨stmt⟩ | "**do**" ⟨stmt⟩ "**while**" "**(**" ⟨expr⟩ "**)**" "**;**"
| "**if**" "**(**" ⟨expr⟩ "**)**" ⟨stmt⟩ ["**else**" ⟨stmt⟩] | "**break**" "**;**" | "**continue**" "**;**" | "**return**" [⟨expr⟩] "**;**"
| ⟨block⟩ | ⟨call⟩ "**;**" | "**switch**" "**(**" ⟨expr⟩ "**)**" "**{**" (("**case**" ⟨expr⟩ "**:**")+ ⟨clause⟩)* ["**default**" "**:**" ⟨clause⟩] "**}**"
⟨assign⟩ ::= "**=**" | "**+=**" | "**-=**" | "***=**" | "**/=**" | "**%=**"
⟨range⟩ ::= ⟨expr⟩ ("**TO**" | "**DOWNTO**") ⟨expr⟩ ["**STEP**" ⟨expr⟩]
⟨clause⟩ ::= (⟨stmt⟩)* ("**break**" "**;**" | "**NEXT**" "**;**")
⟨write⟩ ::= "**WRITE**" | "**WRITELN**" | "**WRITESP**" | "**WRITESPLN**"
⟨format⟩ ::= ⟨expr⟩ | "**FORM**" "**(**" ⟨expr⟩ "**,**" ⟨expr⟩ ["**,**" ⟨expr⟩] "**)**"
⟨expr⟩ ::= ⟨int-const⟩ | ⟨float-const⟩ | ⟨char-const⟩ | ⟨string-literal⟩ | "**true**" | "**false**" | "**(**" ⟨expr⟩ "**)**" | ⟨l_value⟩ | ⟨call⟩
| ⟨unop⟩ ⟨expr⟩ | ⟨expr⟩ ⟨binop⟩ ⟨expr⟩ | "**(**" ⟨type⟩ "**)**" ⟨expr⟩ | "**NULL**" | "**NEW**" "**(**" ⟨type⟩ ["**,**" ⟨expr⟩] "**)**"
⟨l_value⟩ ::= ⟨id⟩ | ⟨expr⟩ "**[**" ⟨expr⟩ "**]**" | "*****" ⟨expr⟩ | ⟨expr⟩ "**.**" ⟨id⟩ | ⟨expr⟩ "**->**" ⟨id⟩
⟨unop⟩ ::= "**+**" | "**-**" | "**NOT**" | "**!**"
⟨binop⟩ ::= "**+**" | "**-**" | "*****" | "**/**" | "**%**" | "**MOD**" | "**==**" | "**!=**" | "**<**" | "**>**" | "**<=**" | "**>=**" | "**&&**" | "**AND**" | "**||**" | "**OR**"
⟨call⟩ ::= ⟨id⟩ "**(**" [⟨expr⟩ ("**,**" ⟨expr⟩)*] "**)**"

Fig. 1. A context-free grammar defining the syntax of CAL in EBNF. Operator precedence and associativity are the same as in C.

- Assignment (simple or composite) is a statement. (Alas, for compatibility purposes we have to give up the assignment operator `:=` and accept the commonly used `=`, which we would prefer not to confuse with the equality operator known from mathematics.) There is just one type of increment and decrement operators (postfix, e.g., `x++`), used again as statements. Therefore, expression evaluation cannot contain direct side-effects. Also, we omit bitwise operators and conditional expressions (`?:`).

- We omit the **`for`** statement, which is too general for our purposes, and replace it with a **`FOR`** statement following the style of Pascal, mentioning the control variable and a (precomputed) range of values that the control variable will take. Using **`FOR`**, the maximum number of iterations is always finite and known before the loop starts executing; **`break`** and **`continue`** can be used to exit the loop and proceed with the next iteration.

- The **`switch`** statement is sanitized; **`case`** labels cannot appear everywhere. Furthermore, clauses are required to end either with a **`break`**, or with the new keyword **`NEXT`** in order to explicitly proceed to the next clause.

- Four kinds of **`WRITE`** statements are used for the output of data, allowing any number of arguments of any type, with a number of formatting options that are useful for an introductory course. Library functions `READ_INT`, `READ_REAL` and `getchar` are used to input data.

- The use of pointers has also been sanitized. The connection between pointers and arrays is still present, but it only allows the use of a pointer as an array; no pointer arithmetic is allowed and there is no "address of" operator (`&`). Pointers are used for dynamic memory allocation; **`NEW`** and **`DELETE`** help students for this purpose (in the spirit of C++, instead of `malloc` and `free` in C).

- Call by reference is allowed, using the same notation that

C++ uses for references.

IV. Introduction to Programming using CAL

Our course is based on the simplicity philosophy of the great teachers of the 1960s: Dijkstra, Hoare, and Wirth. We only give some highlights for lack of space.

- As early as in the second week, students are able to write simple programs that input, process and output data.

```
PROGRAM area_of_circle ()
{ WRITE("Give the radius: ");
  REAL r = READ_REAL();
  REAL a = 3.1415926 * r * r;
  WRITELN("The area is: ", a);
}
```

- Control flow and combinatorial calculations.

```
PROGRAM primes ()
{ int p;
  WRITELN(2);
  FOR (p, 3 TO 1000 STEP 2)
  { int t = 3;
    while (p MOD t != 0) t += 2;
    if (p == t) WRITELN(p);
  }
}
```

- Structured programming: modules, functions and procedures, parameter passing, stepwise refinement.
- Recursion. Euclid's algorithm for the greatest common divisor is one of the examples that we use:

```
FUNC int gcd (int i, int j)
{ if (i==0 OR j==0) return i+j;
  else if (i > j)   return gcd(i MOD j, j);
       else         return gcd(i, j MOD i);
}
```

- Call by reference: e.g., in `swap`, useful for sorting.

```
PROC swap (int &x, int &y)
{ int t = x; x = y; y = t; }
```

- Arrays: merge sort and quick sort, which are also examples of recursion.

```
PROC merge (int a[], int fst, int mid, int lst);

PROC mergesort (int a[], int first, int last)
{ if (first >= last) return;
  int mid = (first + last) / 2;
  mergesort(a, first, mid);
  mergesort(a, mid+1, last);
  merge(a, first, mid, last);
}
```

- Dynamic data structures, such as linked lists and trees, e.g., in-situ reversal of a simply linked list.

```
struct node { int data; struct node *next; };
typedef struct node *list;

PROC reverse (list &l)
{ list q = NULL;
  while (l != NULL)
  { list p = l;
    l = p->next; p->next = q; q = p;
  }
  l = q;
}
```

When grading, we reward the design of efficient algorithms for solving problems such as the following:

MAXSUM: Read a sequence of n integer numbers (positive, zero, or negative) and output the largest value of the sum of (arbitrarily many) consecutive numbers in the sequence.

We expect students who correctly solve this problem to come up with one of three general types of solutions, or variations thereof. The first, is the obvious $O(n^3)$ algorithm: for all possible starts (`i`) and ends (`j`) of subsequences, calculate the sum and find the largest. The second is the slightly less obvious $O(n^2)$ algorithm that, for every given start (`i`) avoids recomputing the sum of a subsequence from scratch but reuses the sums of smaller subsequences. Finally, the third is a linear algorithm — $O(n)$ time and requiring $O(1)$ memory — which works in a greedy fashion.

```
PROGRAM maxsum_On ()
{ int n = READ_INT();
  int i, sofar = 0, best = 0;
  FOR (i, 0 TO n-1)
  { int x = READ_INT();
    sofar += x;
    if (sofar < 0)          sofar = 0;
    else if (sofar > best) best = sofar;
  }
  WRITELN(best);
}
```

V. Concluding Remarks

We have presented CAL, a C-like algorithmic language that we advocate for teaching "Introduction to Programming" focusing on problem solving. We discussed the design of CAL, which is essentially a controlled subset of C with some appropriate extensions. We would like to see CAL, or appropriate subsets of it, as a vehicle to teach computer programming to high-school students, making a bridge between mathematics and computer science in secondary education.

References

[1] M. Felleisen, R. B. Findler, M. Flatt, and S. Krishnamurthi, *How to Design Programs: An Introduction to Computing and Programming*. MIT Press, 2001.

[2] J. Gosling, B. Joy, G. L. S. Jr., G. Bracha, and A. Buckley, *The Java Language Specification*, java se 7 ed. Addison-Wesley, 2013.

[3] P. Hudak, *The Haskell School of Expression: Learning Functional Programming through Multimedia*. Cambridge University Press, 2000.

[4] G. Hutton, *Programming in Haskell*. Cambridge University Press, 2007.

[5] *ISO/IEC 9899:2011 Standard, Information technology – Programming languages – C*, International Organization for Standardization, 2011.

[6] K. Jensen and N. Wirth, *Pascal user manual and report — ISO Pascal standard, 4th Edition*. Springer, 1991.

[7] B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1988.

[8] L. C. Paulson, *ML for the Working Programmer*, 2nd ed. Cambridge University Press, 1996.

[9] B. Stroustrup, *The C++ Programming Language*, 3rd ed. Addison-Wesley, 1997.

[10] N. Wirth, "The programming language Pascal," *Acta Informatica*, vol. 1, pp. 35–63, 1971.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1537–1544

Compilation to Quantum Circuits for a Language with Quantum Data and Control

Yannis Rouselakis*† Nikolaos S. Papaspyrou* Yiannis Tsiouris* Eneia N. Todoran‡

*School of Electrical and Computer Engineering
National Technical University of Athens
Polytechnioupoli, 15780 Zografou, Athens, Greece
Email: {nickie, gtsiour}.softlab.ntua.gr

†Department of Computer Science
University of Texas at Austin
2317 Speedway, Stop D9500, Austin, TX 78712, USA
Email: jrous@cs.utexas.edu

‡Computer Science Department
Technical University of Cluj-Napoca
Baritiu Street 28, 400027, Cluj-Napoca, Romania
Email: Eneia.Todoran@cs.utcluj.ro

***Abstract*—In this paper we further investigate nQML, a functional quantum programming language that follows the "quantum data and control" paradigm. We define a semantics for nQML, which translates programs to quantum circuits in the category FQC of finite quantum computations, following the approach of Altenkirch and Grattage's QML. This semantics, which coincides with the denotational semantics for nQML over density matrices and unitary transformations, serves as a compiler from nQML programs to quantum circuits. We also provide an implementation of this compiler, written in Haskell, as well as an interpreter for quantum circuits.**

I. Introduction

QUANTUM computing processes data that is stored in the form of quantum bits (qubits) and, for doing so, it employs quantum mechanical phenomena such as the superposition and entanglement of quantum states. Roughly speaking, a qubit may contain the digit "0", the digit "1", or any superposition of these two. Although research towards the manufacturing of quantum computers has not yet led to mature results, *quantum circuits* seem to be today a commonly accepted model for quantum hardware. Such circuits consist of appropriate formations of quantum gates, acting upon qubits in the same way that classical logic gates act upon bits in ordinary computers.

Most quantum programming languages that have been proposed so far are based on the principle "quantum data, classical control", that is, on the idea that the execution of a quantum program follows a specific control-flow, exactly as the execution of a program in a classical computer. Such languages allow programmers to use quantum data, in addition to classical, and through their manipulation to implement quantum algorithms. On a different track, we see languages following the "quantum data and control" paradigm. Such languages use quantum control flow; in other words, they allow the execution flow of a program to be in a superposition of various different states in exactly the same way as the quantum data that the program manipulates.

nQML [15], [14] is a high-level functional language based on the concept of "quantum data and control." It was defined by Lampis *et al.*, inspired by Altenkirch and Grattage's QML [1], [8], [9], and its main design goal was to give programmers sufficient expressive power to implement quantum algorithms easily, while preventing them from breaking the rules of quantum computation. nQML includes constructs which allow any unitary transformation to be expressed as a program in nQML quite naturally, more or less using the same notation that is used by the designers of quantum algorithms. It also permits quantum measurements to be carried out at any point during the execution of a program.

As explained in the paper defining nQML [14], the relative ease of use of the language comes at the cost of putting aside a number of important practical issues, such as the existence of imperfect quantum hardware, the need for quantum error correction and the fact that every quantum program will eventually have to be implemented as a quantum circuit using only a finite set of quantum gates and, therefore, some of the unitary transformations that nQML allows will have to be approximated. Similar problems were a source of concern for the founders of the classical programming model many decades ago. Fortunately they have been resolved and their solutions have been abstracted in such a way that people who use modern high-level programming languages need not know anything about them. The same can and must be done for quantum programming languages and, therefore, such issues should be tackled not by the designer and users of a quantum programming language, but by the architect of a quantum computer, the designer of its operating system and, to a lesser extent, the designer of the compiler.

In order to demonstrate the feasibility of using nQML as a quantum programming language and to draw attention to the assumptions that are necessary and to the problems that remain to be resolved, in this paper we define a compiler for nQML, targeting quantum circuits in the category FQC of finite quantum computations [1], [8]. We also provide an

978-83-60810-53-8/$25.00

implementation of the compiler in Haskell, as well as an interpreter for quantum circuits in FQC. The combination of these two can be used for the execution of quantum programs and for a direct comparison with the original definition of the language, using denotational semantics over density matrices and unitary transformations [14].

The rest of the paper is structured as follows. In section II we give the syntax of the language nQML and explain its constructs. Section III contains a description of the quantum circuits that we will use as the target language for our nQML compiler, which is in turn defined in section IV. Section V contains a number of examples in nQML, corresponding to well known quantum algorithms, and the quantum circuits in which they are compiled. We conclude with an exposition of related work, followed by some remarks and directions for future work.

Due to space limitations, in this paper we do not have the luxury to explain how quantum programming works. It is assumed that the reader is familiar with the basics of quantum computation and quantum circuits. There are several introductory books [3], [12], [24], [16], as well as publicly available manuscripts and course material on this field.

II. The language nQML

The syntax of nQML is given by the following grammar. It is assumed that x is a variable identifier and λ is a complex constant. The grammar defines two syntactic classes. Quantum expressions are denoted by e; they represent quantum programs and their syntax is similar to that of QML. Classical expressions are denoted by c; they are only needed in the quantum transformation construct $|e\rangle \to x, x'.\, c$ and they can represent two types of information: a structure of classical bits or a complex number.

$$
\begin{array}{rcl}
e & ::= & x \mid \{\,(\lambda)\,\mathbf{qfalse} + (\lambda')\,\mathbf{qtrue}\,\} \\
& \mid & \mathbf{let}\ x = e_1\ \mathbf{in}\ e_2 \\
& \mid & (e_1, e_2) \mid \mathbf{let}\ (x_1, x_2) = e_1\ \mathbf{in}\ e_2 \\
& \mid & \mathbf{if}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 \mid \mathbf{ifm}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 \\
& \mid & |e\rangle \to x, x'.\, c \\
c & ::= & x \mid \mathbf{false} \mid \mathbf{true} \mid \lambda \mid \mathbf{let}\ x = c_1\ \mathbf{in}\ c_2 \\
& \mid & (c_1, c_2) \mid \mathbf{let}\ (x_1, x_2) = c_1\ \mathbf{in}\ c_2 \\
& \mid & \mathbf{if}\ c\ \mathbf{then}\ c_1\ \mathbf{else}\ c_2 \mid c_1 = c_2 \mid c_1 < c_2 \\
& \mid & \mathbf{int}\, c \mid c_1 + c_2 \mid c_1 - c_2 \mid c_1 * c_2 \mid c_1/c_2 \mid c_1^{c_2}
\end{array}
$$

Variables in nQML are viewed as references to quantum information that is stored in a global quantum state. There are two types of quantum information: qubits and products. A new qubit is allocated in the quantum state when the superposition operator $\{\,(\lambda)\,\mathbf{qfalse} + (\lambda')\,\mathbf{qtrue}\,\}$ is used, in the same way that new objects are allocated on the heap when a data constructor is used in a functional programming language. Products are introduced and eliminated with the constructs (e_1, e_2) and $\mathbf{let}\ (x_1, x_2) = e_1\ \mathbf{in}\ e_2$. nQML also features three control constructs:

- $\mathbf{ifm}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2$: It conducts a measurement on e, which must be of type qubit. Depending on the result, it executes one of its branches. It is similar to a classical random branching, based on a toss of a biased coin with probabilities depending on the state of the qubit being measured.
- $\mathbf{if}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2$: It allows the programmer to perform quantum branching. If e, which must be of type qubit, is in a classical state, then the effect is what we would expect from **ifm**. But if e is in a quantum superposition, the program proceeds in a quantum superposition of both branches, most likely creating entanglement among the qubits of the quantum state.
- $|e\rangle \to x, x'.\, c$: A generic means of expressing any unitary transformation, which has to be relied upon when a transformation can not be easily broken down to a series of controlled operations, expressible with **if**. Its advantage is that, rather than forcing programmers to precompute and provide the whole unitary matrix of the transformation, whose size is exponential in the number of qubits that it affects, it allows them to express that matrix as a complex function of the input and output state of the transformed qubits. This leads to a succinct and clear expression of many useful quantum algorithms, such as the Deutsch-Jozsa or Grover's algorithm that are described in Section V.
  In quantum pseudocode notation, all unitary transformations can be expressed in the form:
  $$|i\rangle \to \sum_{j=0}^{2^n-1} f(i,j)\,|j\rangle$$
  where $f(i,j)$ is a function of the input state i of the quantum register and its output state j. The construct $|e\rangle \to x, x'.\, c$ allows the programmers to use precisely this natural notation: the classical variables x and x' denote the register's input and output state and the classical expression c denotes the function's body.
  From this notation, if the function f is known, the unitary matrix can be easily constructed by taking $S_{j,i} = f(i,j)$. Of course, not all functions f result in unitary matrices and the type system of nQML cannot efficiently decide whether the resulting transformation is indeed unitary. The type system of Altenkirch and Grattage's QML is able to do that, at the expense of making the size of the program exponential and complicating the typing with orthogonality constraints.

nQML admits a simple type system and denotational semantics [14]. By simple, we mean that both use structures and techniques that are typical in the study of classical programming languages of similar size and complexity.

The main novelty of nQML's type system is that the type of a quantum expression conveys information which reveals the exact qubits of the quantum state in which the expression's value resides. Qubit aliasing is allowed in such a way that the "no cloning" and "no dropping" principles are not violated. Programmers have the look-and-feel of a classical programming language, without linearity restrictions. The type system can be extended to support polymorphic higher-order

functions, where polymorphism is over the exact qubits of the quantum state that are used for representing data [14].

Quantum types, in the type system of nQML, are defined by the grammar

$$\tau \ ::= \ \mathbf{qbit}[n] \ | \ \tau_1 \otimes \tau_2$$

where n is the exact qubit of the state that is used, e.g., an expression has type $\mathbf{qbit}[5]$ if its value is stored in the 5th qubit of the state. This information is used to make sure that the same qubit cannot be used twice in a transformation.

There are two typing relations: $\Gamma; n \ \vdash^{\circ} \ e \ : \ \tau; m$ is for type checking pure quantum expressions (i.e. without measurements); on the other hand, $\Gamma; n \ \vdash \ e \ : \ \tau; m$ is for type checking arbitrary quantum expressions, We refer to both by $\Gamma; n \ \vdash^{\alpha} \ e \ : \ \tau; m$, allowing the superscript $^{\alpha}$ to be either $^{\circ}$ or empty. As the types of nQML convey information regarding the position of qubits in the quantum state, the typing relation is forced to process and propagate such information. In $\Gamma; n \vdash^{\alpha} e : \tau; m$, the natural number n appearing on the left side of the relation stands for the number of qubits of the original quantum state, before e starts evaluating. The natural number m appearing on the right side of the relation stands for the number of new qubits that are allocated during the evaluation of e. The typing rules are defined in [14].

The denotational semantics of nQML is based on the use of density matrices to describe quantum states. The meaning of a well-typed nQML program is a function from density matrices to density matrices and describes the program's effect on an arbitrary quantum input state. Pure well-typed programs, i.e., programs which conduct no measurements, are also assigned a meaning in the form of a unitary matrix which describes the transformation they perform on the quantum state. The execution of an nQML program can be seen as a sequence of steps which affect the quantum state by allocating new qubits, by applying unitary transformations to existing qubits or by measuring existing qubits.

III. Quantum circuits

Quantum circuits are one possible model of quantum computation. We extend the Haskell data type proposed by Altenkirch and Grattage [6], [8] by adding one more constructor for arbitrary unitary matrices (`Unit`), which will be the target of nQML's $|e\rangle \to x, x'.\, c$ construct.

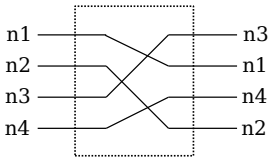
```
data Circ = Rot (C,C) (C,C)
          | Wire [Int]
          | Par Circ Circ
          | Seq Circ Circ
          | Cond Circ Circ
          | Unit (Matrix C)
```

The set of quantum circuits operating on a state of n qubits are defined inductively using these constructors:

- Rotation `Rot` $(\lambda_0, \lambda_1)\,(\kappa_0, \kappa_1)$: introduces a new unitary transformation on one qubit ($n = 1$), defined by the following matrix, where $\lambda_0^* \kappa_0 + \lambda_1 \kappa_1^* = 0$.

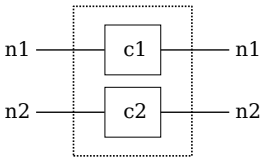
$$\begin{pmatrix} \lambda_0 & \lambda_1 \\ \kappa_0 & \kappa_1 \end{pmatrix}$$

- Wire reordering `Wire` p: reorders the qubits in the state. The parameter p must be a permutation of the sequence $[0\,..\,n-1]$. If the i-th element of this permutation is j, this means that the wire at the i-th position in the input state becomes the j-th wire of the output state. The identity permutation corresponds to the identity unitary matrix which leaves the state unchanged. When drawing quantum circuits, we will not use quantum gates to implement wire reordering; we will just draw crossing wires, e.g., as follows:

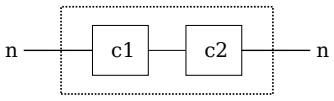


  In all circuits, the numbers next to the wires denote multiplicity, i.e., the number of qubits in the state. If $n_1 = n_2 = n_3 = n_4 = 1$, the reordering shown above would be encoded in Haskell by `Wire [1,3,0,2]`.

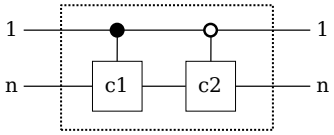
- Parallel composition `Par` $c_1\, c_2$: combines c_1 and c_2 in parallel, adding the number of qubits in their states.



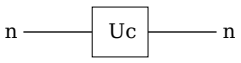
- Sequential composition `Seq` $c_1\, c_2$: combines c_1 and c_2 in sequence, The two circuits must have a state of n qubits, where n is also the number of qubits in their composition.



- Conditional `Cond` $c_1\, c_2$: creates a conditional circuit that is controlled by an extra qubit. The two circuits must have a state of n qubits, whereas the number of qubits in the conditional circuit is $n + 1$.



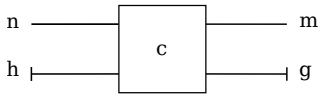
- Arbitrary unitary matrix `Unit` C: creates a circuit with a state of n qubits corresponding to the unitary matrix C. Such a circuit must in general be approximated by an appropriate composition of elementary quantum gates.



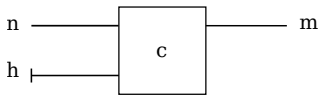
Reversible finite quantum circuits form the category $\mathsf{FQC}^{\approx}$, whose objects are states of n qubits and morphisms are unitary

transformations. Following Altenkirch and Grattage [6], [8], we define two larger categories, FQC and FQC$^\circ$.

Circuits in FQC are not necessarily reversible: they can also contain measurements and/or qubit initializations (which also amount to measurements). To model circuits in FQC, we separate a number of qubits of the input state, which we call *heap*, and a number of qubits of the output state, which we call *garbage*. Qubits in the heap are considered to be initialized to $|0\rangle$. Qubits in the garbage are measured and discarded. When drawing circuits, we denote the heap and garbage by terminating lines. It must be $n + h = m + g$.



Also, the category FQC$^\circ$ is a subset of FQC where circuits are allowed to have a heap, but not garbage. Such circuits are pure, in the sense that they do not contain measurements, and can be modelled by unitary transformations between pure quantum states. It must be $n + h = m$.

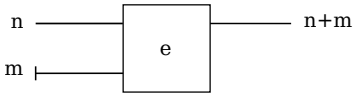


Obviously, FQC$^\approx$ $\subset$ FQC$^\circ$ $\subset$ FQC.

IV. A COMPILER FOR NQML

Following the compilation approach used by Altenkirch and Grattage [6], [8] we use the typing relation for compiling (pure and impure) quantum expressions. However, in contrast to the approach used for QML, the process is not guided by the linear type system, deciding how to split the wires of the input state. Instead, purity information and the numbers n and m from nQML's typing information are used.

If e is a pure quantum expression such that $\Gamma; n \vdash^\circ e : \tau; m$, then e is compiled to a circuit in FQC$^\circ$ which has an input state of n wires plus m wires of heap and an output state of $n + m$ wires (without garbage). We draw this as follows:



On the other hand, if e is an impure quantum expression such that $\Gamma; n \vdash e : \tau; m$, then e is compiled to a circuit in FQC which has an input state of n wires plus $h \geq m$ wires of heap and an output state of $n + m$ wires plus g wires of garbage. It must be $h = m + g$. We draw this as follows:

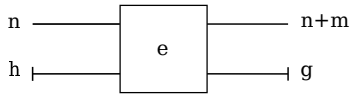


Fig. 1 shows how nQML constructs are compiled to circuits. The compilation process is based on the typing of expressions.

- **Superposition.** A new qubit is added to the state corresponding to $\{ (\lambda)\,\mathbf{qfalse} + (\lambda')\,\mathbf{qtrue} \}$. The remaining n qubits of the state are unaltered, whereas the new qubit is initialized with the transformation matrix:

$$\begin{pmatrix} \lambda & \lambda' \\ \lambda' & -\lambda \end{pmatrix}$$

- **Let construct and products.** Although the typing rules for these three constructs (simple let, product formation and product elimination) are different when it comes to the types of the participating expressions, they are all the same w.r.t. the number of qubits in the state and they produce the same quantum circuit, which is essentially the sequential composition of two expressions.

- **Quantum conditional.** In the typing of $\mathbf{if}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2$, the condition is the k-th qubit of the state and the pure expressions e_1 and e_2 are not allowed to refer to the k-th qubit (as the environment $\Gamma|_k$ suggests in the figure). The circuit corresponding to condition e is generated first and the k-th qubit of the output state is isolated. This qubit controls the conditional circuit of e_1 and e_2. Notice that it is not strictly true that this conditional circuit is composed of e_1 and e_2. First of all, we have to translate away the (unused) k-th qubit, by inductively transforming the circuits corresponding to e_1 and e_2. Then, we have to extend the input state of the smallest of the two circuits, so that both expect an input state of $n + m - 1 + \max(m_1, m_2)$ qubits.

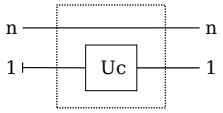
- **Measurement.** The difference between the quantum conditional and the measurement is that (a) impure expressions are allowed in branches, (b) the branches can use the qubit of the condition, and (c) the qubit of the condition is measured at the end of the circuit. In order to be used by the two branches and (at the same time) be measured at the end, the qubit of the condition must be duplicated (creating a quantum entanglement). This is achieved by using one extra qubit and the controlled CNOT gate. The two expressions may, of course, use the measured value of this qubit. Notice that this is the only circuit which explicitly creates garbage, by measuring the qubit of the condition. Also, this is one of the two circuits that explicitly use qubits from the heap (the other one is generated by superposition).

- **Unitary transformation.** In $|e\rangle \to x, x'.\, c$, it is assumed that $c(x, x')$ defines an arbitrary unitary transformation on states of $n + m$ qubits, and this transformation is applied to the result of expression e.

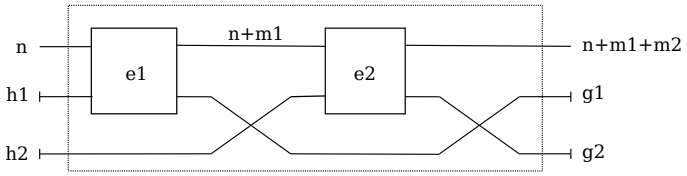
The implementation of our compiler applies several simple optimizations to the generated quantum circuits.[1] In general,

[1]The implementation of nQML can be found at http://www.softlab.ntua.gr/~nickie/Research/nqml/. It consists of approximately 3,200 lines of Haskell code. Parts of it have been written by Michael Lampis.

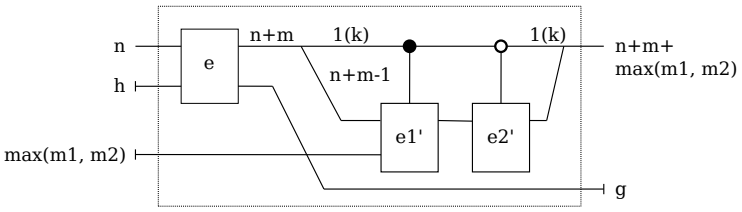
Superposition:
$\Gamma; n \vdash^{\circ} \{ (\lambda)\, \mathbf{qfalse} + (\lambda')\, \mathbf{qtrue} \} : \mathbf{qbit}[n]; 1$



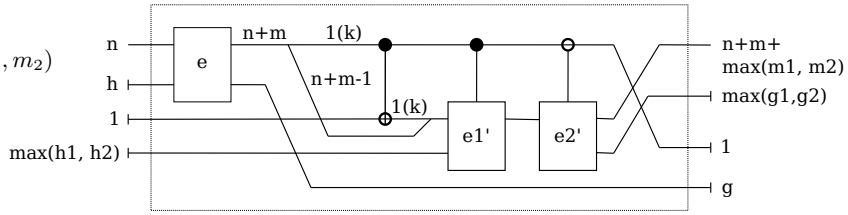
Let and products:
$\Gamma; n \vdash^{\alpha} \mathbf{let}\ x = e_1\ \mathbf{in}\ e_2 : \tau; m_1 + m_2$
$\Gamma; n \vdash^{\alpha} (e_1, e_2) : \tau; m_1 + m_2$
$\Gamma; n \vdash^{\alpha} \mathbf{let}\ (x_1, x_2) = e_1\ \mathbf{in}\ e_2 : \tau; m_1 + m_2$
where:
$\Gamma_1; n \vdash^{\alpha} e_1 : \tau_1; m_1$
$\Gamma_2; n + m_1 \vdash^{\alpha} e_2 : \tau_2; m_2$



Quantum conditional:
$\Gamma; n \vdash^{\alpha} \mathbf{if}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 : \tau; m + \max(m_1, m_2)$
where:
$\Gamma; n \vdash^{\alpha} e : \mathbf{qbit}[k]; m$
$\Gamma|_k; n + m \vdash^{\circ} e_1 : \tau; m_1$
$\Gamma|_k; n + m \vdash^{\circ} e_2 : \tau; m_2$



Measurement:
$\Gamma; n \vdash \mathbf{ifm}\ e\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 : \tau; m + \max(m_1, m_2)$
where:
$\Gamma; n \vdash e : \mathbf{qbit}[k]; m$
$\Gamma; n + m \vdash e_1 : \tau; m_1$
$\Gamma; n + m \vdash e_2 : \tau; m_2$



Unitary transformation:
$\Gamma; n \vdash^{\alpha} |e\rangle \to x, x'.\, c : \tau; m$
where:
$\Gamma; n \vdash^{\alpha} e : \tau; m$
$c(x, x')$ defines a unitary transformation on $n+m$ qubits

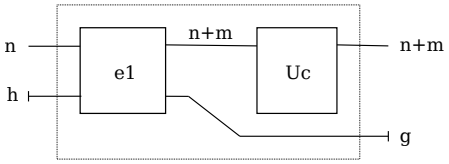


Fig. 1. Compiling nQML expressions to quantum circuits: A typing-directed approach.

the implementation is written in Haskell; it targets the polymorphic language described in [14] and consists of:

- a parser,
- a type checker,
- a first interpreter, based on the denotational semantics of nQML, using density matrices and unitary transformations,
- the compiler from nQML to quantum circuits, defined in this paper, and
- a second interpreter, based on the simulation of quantum circuits generated by our compiler.

The implementation checks that the outputs of the two interpreters coincide, thus testing the correctness of our compiler.

V. Examples

In this section, we outline the use of nQML and its compiler with two relatively simple but historically important examples: Deutsch's algorithm for testing whether a function on one bit is balanced or constant [4], and Grover's algorithm for searching an unsorted database [11].

We begin by providing a couple of auxiliary functions, **not** and **had**, that will be useful in both examples. They correspond to the NOT gate and the Hadamard gate. Their definitions can be given by simple unitary transformations.

```
def not q = |q> -> x, x'.
  if x' = x then 0 else 1;

def had q = |q> -> x, x'.
  (if x then (if x' then -1 else 1) else 1)
    / sqrt(2);
```

The syntax of function definitions in nQML follows the proposed extension with polymorphic functions [14]. Such functions could be treated as macros by the compiler.

We will also abbreviate tuples of more than two elements by writing (x, y, z) instead of $(x, (y, z))$. Furthermore, we will

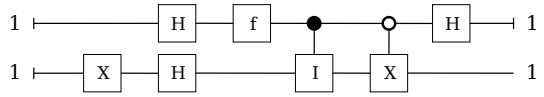


Fig. 2. The circuit produced for Deutsch's algorithm, where f is the parameter: the function that we want to determine if it's constant or balanced.

use **qtrue** as syntactic sugar for $\{\,(0)\,\mathbf{qfalse} + (1)\,\mathbf{qtrue}\,\}$ and **qfalse** as syntactic sugar for $\{\,(1)\,\mathbf{qfalse} + (0)\,\mathbf{qtrue}\,\}$.

A. Deutsch's Algorithm

Deutsch's algorithm (later generalized by Deutsch and Jozsa) was one of the first quantum algorithms to be studied. Supposing that we have a function $f(x) : \{0,1\} \to \{0,1\}$, we want to determine whether this function is *constant*, i.e., $f(0) = f(1)$, or *balanced*, i.e., $f(0) \neq f(1)$, by just computing it once.

There is obviously no classical solution to this problem. The quantum solution employs the trick of computing the function once, with a superposition of the two inputs, then appropriately measuring the result. (The interested reader is refered to the introductory literature in quantum computations for analyses of the algorithm and proofs of correctness.) In nQML, it can be written as follows. The measurement of **had** i gives 1 if function f is balanced and 0 if it is constant.

```
def Deutsch f =
  let (i, j) = (had qfalse, had qtrue) in
  let r = if f i then j else not j in
  ifm had i then qtrue else qfalse;
```

The circuit that our compiler produces for this program (just measuring the result and excluding the new qubits for the branches of **ifm**), is shown in Fig. 2.

B. Grover's Algorithm

As a second example, let us see an implementation of Grover's fast database search. Consider an unsorted database with $N = 2^n$ entries and the problem of finding the index of a particular database entry that satisfies some criterion. To simplify things, let us assume that c denotes the index that we are searching for. We first need to implement the query operator, which is a transformation corresponding to a matrix which has 0 everywhere, 1 along the primary diagonal and -1 at the element with coordinates (c, c).

```
def query q = |q> -> x, x'.
  if x = x' then
    if int x = c then -1 else 1
  else
    0;
```

We now define the diffusion operator, a transformation corresponding to the matrix $2P - I$, where P a matrix with 2^{-n} everywhere.

```
def diffusion q = |q> -> x, x'.
  if x = x' then 2 / 2^n - 1 else 2 / 2^n;
```

The algorithm proceeds by repeated iterations of queries and diffusions. Let us now consider the most simple application of Grover's algorithm: searching in a space of size $N = 4$ (with $n = 2$ qubits). In this special case, one iteration is enough to produce the correct result with certainty:

```
def grover2 =
  let qs = (had qfalse, had qfalse) in
  diffusion (query qs);
```

In the general case, $O(\sqrt{N})$ iterations of the two operators are required to obtain the result with a high probability. Consider $N = 16$ (with $n = 4$ qubits). Three iterations suffice:

```
def grover4 =
  let qs = (had qfalse, had qfalse,
            had qfalse, had qfalse) in
  let step1 = diffusion (query qs) in
  let step2 = diffusion (query qs) in
  let step3 = diffusion (query qs) in
  qs
```

The circuit that our compiler produces for grover4 is shown in Fig. 3. The result is implicitly measured.

VI. Related work

The design of quantum algorithms, such as Shor's algorithm for the factorization of integer numbers in polynomial time [21] and Grover's algorithm for searching an unordered list of n elements in $O(\sqrt{n})$ time [11], has shown that the quantum model of computation is strictly more powerful than the classical model; although both can compute the same set of functions, some functions can be computed in the quantum model strictly faster than in the classical one. Quantum algorithms are usually studied at a low level, either expressed directly in the form of quantum circuits or using appropriate mathematical models. The fact that reasoning about quantum circuits is no easier than reasoning about their classical counterparts has given rise to quantum programming languages, that is, languages that allow programmers to implement quantum algorithms and make use of the added power of the quantum computational model, while respecting its special restrictions.

Knill's conventions for quantum pseudocode [13] was the first proposed formal language for the description of quantum algorithms, tightly connected with the Quantum Random Access Machine. Since then, several quantum programming languages have been proposed; the reader is referred to an excellent (although slightly outdated) survey of the emerging field [5]. Ömer's QCL is an imperative language with quantum primitives and automatic quantum scratch space management [17]. Moreover, van Tonder has proposed a λ-calculus for higher-order quantum programs without measurements [22]. Both languages, however, do not compile to quantum circuits and, in the case of van Tonder's λ-calculus, it is not clear how this can be done. Sanders and Zuliani have defined qGCL, an extension of Dijkstra's guarded command language [18], and they have shown how to compile qGCL to a form of assembly language for a quantum computer [25].

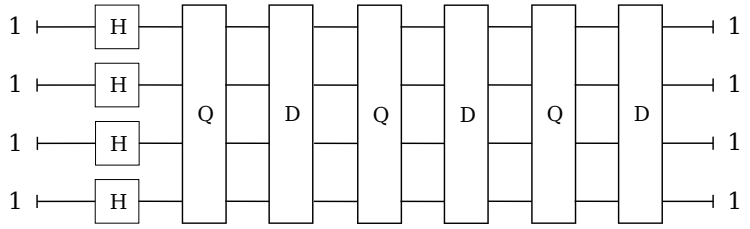


Fig. 3. The circuit produced for Grover's algorithm, where $N = 16$ ($n = 4$). The transformations Q and D correspond to the query and diffusion operators, which are applied iteratively.

Selinger's QPL is a language following the paradigm "quantum data, classical control" [20]. It is functional in nature, although from a programmer's point of view it looks more imperative than functional. QPL allows the programmer to access both classical and quantum memory and includes high-level features such as loops and recursion. Program control is strictly classical and quantum branching can only be implemented indirectly with appropriate unitary transformations. The denotational semantics of QPL is given in the form of superoperators on density matrices. A higher-order extension of QPL in the form of a quantum lambda calculus has also been proposed by Selinger and Valiron [19]. In the same paradigm, Green *et al.* have recently defined Quipper [10], a functional, higher-order quantum programming language designed to be used for implementing large-scale quantum algorithms. They have shown how programs can be compiled to quantum circuits consisting of a large number of gates.

On the other hand, Altenkirch and Grattage's QML is a functional language that follows the paradigm "quantum data and control" [1], [7], [8]. QML comes with a linear type system prohibiting implicit weakening, which would lead to implicit measurements and quantum collapse. The authors describe a way to compile QML programs to quantum circuits in the category FQC of finite quantum computations [6], [9]. Variables in QML correspond to wires in the produced quantum circuit and thus have to be shared implicitly when they are used in several places in a program so as not to break the "no cloning" rule. The sharing of wires is monitored by a linear type system. Altenkirch and Green have recently presented a monadic purely functional interface to quantum programming (the QIO monad) and they provide an implementation in the form of a quantum DSL in Haskell [2]. Again, there is an almost direct translation from QIO to quantum circuits. A similar embedding in Haskell, in the form of arrows, is proposed by Vizzoto *et al.* [23].

VII. Concluding remarks

We have defined a compiler for quantum programs written in the language nQML that follows the paradigm "quantum data and control". The compiler targets quantum circuits in the category FQC of finite quantum computations, defined by Altenkirch and Grattage. We have implemented our compiler as part of nQML's implementation, which is publicly available.

The real challenge in quantum programming, and a definite direction for future work, is the integration of features that are at a higher-level than quantum gates and unitary transformations, for example, reversible binary arithmetic, quantum data structures, etc. The proper integration of such features in quantum programming languages is a hard problem in terms of language design and usability, especially if one wants to keep compatibility with the way in which quantum algorithms are expressed (mostly by non-programmers) today.

Acknowledgment

This research is partially funded by the research project "SemNatComp: Semantic models and technologies for natural computations" (ΓΓΕΤ 11 ROM 11_1_ET30), funded by the Greek General Secretariat for Research and Technology and the European Regional Development Fund, through the operational program "Competitiveness Entrepreneurship & Regions in Transition", action "Billateral Co-operation Greece-Romania 2011-2012".

References

[1] T. Altenkirch and J. Grattage, "A functional quantum programming language," in *Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society, 2005, pp. 249–258.

[2] T. Altenkirch and A. S. Green, "The quantum IO monad," in *Semantic Techniques in Quantum Computation*, S. Gay and I. Mackie, Eds. Cambridge University Press, 2009, p. 173205.

[3] J. Brown, *Quest for the Quantum Computer*. Simon and Schuster, 2001.

[4] D. Deutsch and R. Jozsa, "Rapid solutions of problems by quantum computation," *Proceedings of the Royal Society of London*, vol. A 439, pp. 553–558, Dec. 1992.

[5] S. J. Gay, "Quantum programming languages: Survey and bibliography," *Mathematical Structures in Computer Science*, vol. 16, no. 4, pp. 581–600, Aug. 2006.

[6] J. Grattage and T. Altenkirch, "A compiler for a functional quantum programming language," Jan. 2005, manuscript, available from the authors' web page.

[7] ——, "QML: Quantum data and control," Feb. 2005, manuscript, available from the authors' web page.

[8] J. Grattage, "QML: A functional quantum programming language," Ph.D. dissertation, School of Computer Science and School of Mathematical Sciences, The University of Nottingham, Sep. 2006. [Online]. Available: http://etheses.nottingham.ac.uk/archive/00000250/

[9] ——, "An overview of QML with a concrete implementation in Haskell," *Electronic Notes in Theoretical Computer Science*, vol. 270, no. 1, pp. 165–174, 2011, proceedings of the *4th Workshop on Developments in Computational Models* (DCM '08), doi:10.1016/J.ENTCS.2011.01.015, arXiv:0806.2735. [Online]. Available: http://fop.cs.nott.ac.uk/qml

[10] A. S. Green, P. L. Lumsdaine, N. J. Ross, P. Selinger, and B. Valiron, "Quipper: A scalable quantum programming language," in *Proceedings of the 34th annual ACM SIGPLAN conference on Programming Language Design and Implementation*, Jun. 2013, to appear.

[11] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing*, Philadelphia, PA, May 22-24 1996, pp. 212–219.

[12] M. Hirvensalo, *Quantum Computing*, 2nd ed. Springer, 2004.

[13] E. Knill, "Conventions for quantum pseudocode," Los Alamos National Laboratory, Tech. Rep. LAUR-96-2724, 1996.
[14] M. Lampis, K. G. Ginis, M. A. Papakyriakou, and N. S. Papaspyrou, "Quantum data and control made easier," *Electronic Notes in Theoretical Computer Science*, vol. 210, pp. 85–105, Jul. 2008.
[15] M. Lampis, K. G. Ginis, and N. S. Papaspyrou, "Quantum data and control made easier," in *Preliminary Proceedings of the 4th International Workshop on Quantum Programming Languages*, P. Selinger, Ed., Oxford, UK, Jul. 2006, pp. 73–86. [Online]. Available: http://www.mscs.dal.ca/~selinger/qpl2006/
[16] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. Cambridge University Press, 2010.
[17] B. Ömer, "Structured quantum programming," Ph.D. dissertation, Institute of Information Systems, Technical University of Vienna, May 2003.
[18] J. W. Sanders and P. Zuliani, "Quantum programming," in *Proceedings of the 5th International Conference on Mathematics of Program Construction*, ser. Lecture Notes in Computer Science, vol. 1837. London, UK: Springer-Verlag, 2000, pp. 80–99.
[19] P. Selinger and B. Valiron, "A lambda calculus for quantum computation with classical control," *Mathematical Structures in Computer Science*, vol. 16, no. 3, pp. 527–552, 2006.
[20] P. Selinger, "Towards a quantum programming language," *Mathematical Structures in Computer Science*, vol. 14, no. 4, pp. 527–586, 2004.
[21] P. W. Shor, "Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, 1997.
[22] A. van Tonder, "A lambda calculus for quantum computation," *SIAM Journal on Computing*, vol. 33, no. 5, pp. 1109–1135, 2004.
[23] J. K. Vizzotto, A. R. D. Bois, and A. Sabry, "The arrow calculus as a quantum programming language," in *Logic, Language, Information and Computation*, ser. Lecture Notes in Computer Science. Springer, 2009, vol. 5514, pp. 379–393.
[24] N. S. Yanofsky and M. A. Mannucci, *Quantum Computing for Computer Scientists*. Cambridge University Press, 2008.
[25] P. Zuliani, "Compiling quantum programs," *Acta Informatica*, vol. 41, no. 7, pp. 435–474, Jun. 2005.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1545–1552

Grammar-Driven Development of JSON Processing Applications

Antonio Sarasa-Cabezuelo, José-Luis Sierra
Fac. Informática. Universidad Complutense de Madrid. 28040 Madrid (Spain)
{asarasa,jlsierra}@fdi.ucm.es

***Abstract*—This paper describes how to use conventional parser generation tools for the development of JSON processing applications. According to the resulting grammar-driven development approach, JSON processing applications are architected as syntax-directed translators. Thus, the core part of these components can be described in terms of translation schemata and can be automatically generated by using suitable parser generators. It makes it possible to specify critical parts of the application (those interfacing with JSON documents) by using high-level, grammar-oriented descriptions, as well as to promote the separation of JSON processing concerns from other application-specific aspects. In consequence, the production and maintenance of JSON processing applications is facilitated (especially for applications involving JSON documents with intricate nested structures, as well as for applications in which JSON formats are exposed to frequent changes and evolutions in their surface structures). This paper illustrates the approach with JSON-P as the generic JSON processing framework, with ANTLR as the parser generation tool, and with a case study concerning the development of a player for simple man-machine dialogs shaped in terms of JSON documents.**

***Keywords*—JSON, Grammar-Driven Development, Translation Schemata, Parser Generator, ANTLR, JSON-P**

I. Introduction

JSON (JavaScript Object Notation) [15][34] is a data exchange format based on a subset of the JavaScript programming language that in recent years has achieved enormous relevance in industry. Indeed, many times JSON results in a more natural mechanism for representing data structures than other alternative formats (e.g., XML [25], which is more suitable for representing hierarchical data). In fact, JSON makes it possible to use collections of name-value pairs and ordered sequences of values, which mirrors the typical data included in mainstream programming languages structures (records, objects or hash tables for collections of name-value pairs, and arrays or lists for ordered sequences of values). This JSON feature makes it natural to map JSON documents to data structures in a target programming language. Still, since JSON is based on text encoding, it is independent from particular programming languages and binary formats; indeed, it can be inspected and, with some effort, interpreted by humans, which facilitates development, debugging and system interconnection tasks. Finally, JSON has also found an important application area as a storage format in non-relational database systems [12][24].

As any other data interchange enabling technology, the success of any development based on JSON relies on finding suitable ways of processing JSON documents in the resulting applications. For this purpose, multiple technologies for processing JSON documents have been proposed, which can be classified into two broad categories:

- *Specific processing technologies*. With these artifacts, it is possible to carry out specific-purpose processing tasks (e.g, querying and document transformation). Examples of these proposals are [5][10][18]. While these technologies are easy to use, due to their specific and task-oriented nature, the main drawback of this task-specific approach is the need to find suitable specific technologies for each particular processing task.
- *Generic processing technologies*. These technologies make it possible to achieve any processing task. They are provided by libraries and frameworks for JSON manipulation embedded as part of a general-purpose programming language. Examples of these technologies include those that perform marshalling and unmarshalling between JSON documents and data structures [23], and frameworks for parsing JSON documents [4][11][13][14][16][19][22][29]. In addition, although these technologies can be used to address any processing task, they are substantially more difficult than specific technologies, resulting in higher development and maintenance efforts.

Regardless of their scope of applicability, the aforementioned processing approaches are *data-oriented* in nature, in the sense of conceiving JSON documents as mere data containers, and JSON processing as the mapping of this data into data structures in the host languages. However, since JSON is a formal language, an alternative, language-oriented, approach is possible. This approach will be focused on computer language processing aspects instead of a data marshaling / un-marshaling perspective. In particular, it will be possible to characterize types of JSON documents as *formal grammars*, and then to orchestrate the processing of these documents according to a syntax-directed processing model. Indeed, the characterization of JSON documents as formal grammars is consistent with schema languages like JSON Schema [17]. Thus, the proposed language-oriented (or, more specifically, *grammar-oriented*) approach goes a step further, by conceiving processing tasks of JSON documents being carried out by syntax-directed language processors operating on these

978-83-60810-53-8/$25.00

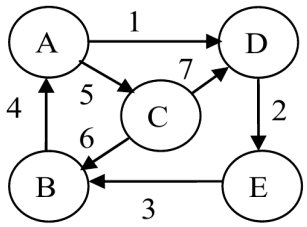
JSON documents. In turn, these language processors can be developed by using dedicated compiler construction tools (parser generators like JavaCC [21], ANTLR [27] or CUP [3], in particular). This approach exhibits the advantages of the variety and stability of these tools, the high level of abstraction to specify the processing (indeed, the approach brings the advantages of task-specific strategies to general-purpose processing settings), greater simplicity in application maintenance, and naturalness for addressing efficient stream-based processing.

This paper describes this grammar-oriented approach to the development of JSON processing applications. The rest of the paper is structured as follows: Section II provides a short introduction to JSON. Section III outlines the grammar-oriented approach. Section IV shows how the approach can be actually implemented by combining a concrete JSON processing framework (JSON-P) with a concrete parser generation tool (ANTLR). Section V illustrates how the approach can be applied to concrete scenarios with the development of a JSON-based application for playing human-computer dialogs. Finally, Section VI provides some conclusions and lines of future work.

II. JSON

As indicated earlier, JSON is a lightweight text-based notation for encoding data structures. Thus, this notation rules how to encode data structures as text entities, known as JSON *documents*. For this purpose, JSON distinguishes among the following kind of data:

(a)



(b)

```
{"nodes":["A","B","C","D","E"],
 "arcs":[
   {"source":"A","target":"D","label":"1"},
   {"source":"A","target":"C","label":"5"},
   {"source":"B","target":"A","label":"4"},
   {"source":"C","target":"B","label":"6"},
   {"source":"C","target":"D","label":"7"},
   {"source":"D","target":"E","label":"2"},
   {"source":"E","target":"B","label":"3"}]}
```

Figure 1. (a) A labelled directed graph, (b) a JSON encoding of the graph in (a).

- *Basic data:* (double precision floating-point) numbers, strings (double-quoted sequences of Unicode characters, with standard scape conventions), Booleans (true and false), and the `null` value.
- *Compound data: arrays* (ordered sequences of comma-separated values, delimited by `[` and `]`), and *objects* (unordered, comma-separated, collections of key-value pairs delimited by `{` and `}`; each key-value pair is in the form *key: value,* where *key* is, in turn, a string).

Using these somewhat simple conventions, JSON makes it possible to represent data structures of arbitrary complexity (it is very similar to what happens with s-expressions in LISP [1], or with XML markup). This flexibility, together with its seamless integration with JavaScript, explains the successful adoption of JSON as an enabling technology for web development, where, for instance, it has become a de facto standard for data exchange in RESTFul service-oriented architectures [30].

Figure 1 illustrates the use of JSON to represent a labeled directed graph with a JSON document. The encoding conventions followed should be apparent from the JSON document itself. It reveals another important feature of JSON: since it is a text-based format, with a little effort it can become understandable to developers. Thus, it facilitates making a good amount of system internals accessible both for humans and machines in terms of JSON documents.

III. The Grammar-Driven approach to the development of JSON applications

In addition to the textual encoding of data structures, JSON documents can be conceived as sentences in a formal language. Indeed, when JSON is used to encode a particular kind of data structure (e.g., labeled directed graphs, as in Figure 1), it is possible to distinguish a subset of JSON documents that meaningfully represents instances of such a data structure. This subset of documents, in turn, can be thought of as defining another formal language: the language of the JSON documents allowable in the particular application domain. Thus, it is possible to apply to JSON similar principles to those used in other analogous fields, like XML (i.e., distinction between *well-formed* and *valid* documents, and characterization of document types with formal grammars). In particular, the use of formal grammars to describe JSON documents in a given application domain acquires full meaning. This paradigmatic bias (i.e., going from a data-oriented perspective to a linguistic, grammar-oriented one) leads to the grammar-driven approach presented in this paper.

The grammar-driven approach can be derived by first considering the structure of a standard syntax-directed translator (Figure 2a). This structure comprises two basic components:

- The *scanner* that is in charge of tokenizing input sentences.
- The *translator,* a *parser* augmented with semantic actions that, when acting on the token sequence produced by the scanner, is able: (i) to recognize this sequence of tokens as belonging to the input language, or otherwise to reject it as invalid, (ii) to arrange it according to its underlying syntactic structure, and, (iii) to process it by firing the semantic actions.

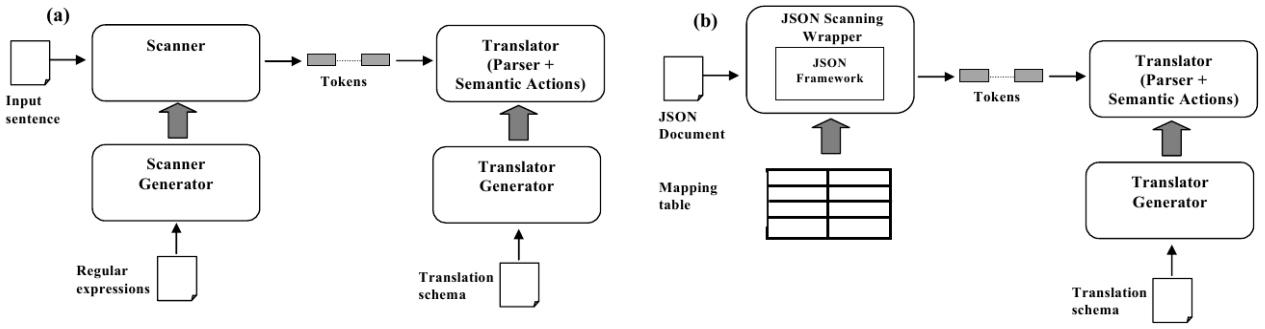


Figure 2. (a) Structure of a Syntax-Directed Translator and the automationaccomplish JSONof its development; (b) modification of the structure despicted in (a) to processing

As is widely acknowledged by the programming language community, this organization results in a processing model especially suited for stream processing, which, under reasonable assumptions, is able to behave in an extremely efficient way [2]. In addition, both the scanner and the translator can be automatically generated from high-level specifications (regular expression-based ones concerning the scanner; *translation schemata* –i.e., context-free grammars augmented with semantic actions, concerning the translator) [2]. Indeed, generation tools like JavaCC, ANTLR or CUP greatly facilitate this development task.

The next step is to adapt classic syntax-directed organization to JSON processing. For this purpose, the scanner in Figure 2a can be replaced by a new component: the *JSON scanning wrapper* (Figure 2b). When operating on JSON documents, this component will map the logical structure of these documents into sequences of tokens, as expected by a syntax-directed translator. It is important to notice that the provision of this component does not rely on the programming of a new generic JSON processor. On the contrary, this component can be meaningfully piggybacked on an existing JSON processing framework (like JSON-simple [19] or JSON-P [16]).

Once this replacement is accomplished, the rest of the organization remains unchanged, as evidenced by Figure 2b. In particular, it is still possible to specify processing (this time of JSON documents) by using high-level translation schemata, and to automatically turn these specifications into efficient implementations by using parser generation tools. Therefore, tools like JavaCC, ANTLR and CUP take a new and unpredicted role, as tools for developing efficient, stream-oriented, JSON processing applications.

Thus, notice that this grammar-driven development approach makes it possible to make up grammar-driven production environments for JSON processing applications by combining a suitable parser generation tool with a general purpose JSON processing framework. The adaptation between the two components will be performed by means of a JSON *scanning wrapper*, which will be dependent on the particular parser generation and processing framework. Beyond this specific component, the approach is nicely independent of the particular parser generator and processing framework chosen.

Concerning the use of this kind of grammar-oriented environments in the actual development of JSON applications, it involves:

- Customizing the JSON *scanning wrapper* to tokenize the logical structure of the JSON documents involved in the application. It can be readily done by providing a *mapping table* associating a distinct token with: (i) each key in each object, (ii) the object opening and closing marks (i.e., { and }), (iii) each possible basic value in the document (i.e., *number, string, true* and *false*). The other structure in the document (e.g., ordered sequences in lists) can be characterized in purely grammatical terms. For instance, Figure 3 despicts the mapping table for the example of labelled graphs in section II.

JSON Element	Token Kind
"nodes"	NODES
"arcs"	ARCS
"source"	SOURCE
"target"	TARGET
"label"	LABEL
{	OO
}	CO
string	STRING

Figure 3. Mapping table for documents like those of Figure 1.

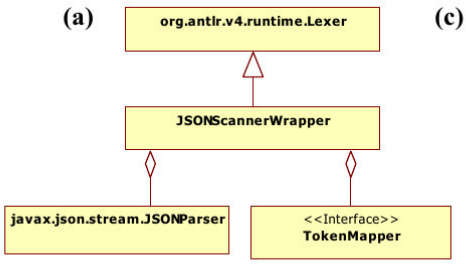
- Characterizing the grammatical structure of the source JSON documents. It can be done by using standard BNF or EBNF notation, augmented with some facilities for describing the structure of objects.

(a)
graph → {nodes: **nodes**, arcs: **arcs**}
nodes → (STRING)*
arcs → (**arc**)*
arc → {source: STRING, target: STRING, label: STRING}

(b)
arc → source: STRING, target: STRING, label: STRING |
source: STRING, label: STRING, target: STRING |
target: STRING, source: STRING, label: STRING |
target: STRING, label: STRING, source: STRING |
label: STRING, source: STRING, target: STRING |
label STRING, target: STRING, source: STRING

Figure 4. (a) Structure of graph-description JSON documents, (b) description of the structure of an arc object using standard EBNF notation.

(a)



(b)

```
public interface TokenMapper {
  int typeForKeyName(String keyName);
  int typeForObjectOpening();
  int typeForObjectClosing();
  int typeForStringValue(String data);
  int typeForNumericValue(String value);
  int typeForTrueValue();
  int typeForFalseValue();
  int typeForNullValue();
}
```

(c)

```
public Token nextToken() {
 try{
  if (parser.hasNext()) {
   JsonParser.Event eventCode=parser.next();
   while (parser.hasNext()&&
          eventCode!=JsonParser.Event.KEY_NAME &&
          eventCode!=JsonParser.Event.START_OBJECT &&
          eventCode!=JsonParser.Event.END_OBJECT &&
          eventCode!=JsonParser.Event.VALUE_NULL &&
          eventCode!=JsonParser.Event.VALUE_STRING &&
          eventCode!=JsonParser.Event.VALUE_NUMBER &&
          eventCode!=JsonParser.Event.VALUE_FALSE &&
          eventCode!=JsonParser.Event.VALUE_TRUE)
          eventCode=parser.next();
   if (eventCode == JsonParser.Event.KEY_NAME) {
        String name=parser.getString();
        JsonLocation location= parser.getLocation();
        return newToken(
                 tokenmapper.typeForKeyName(name),
                 name, location);}
   else if (eventCode == JsonParser.Event.VALUE_STRING) {
         String data = parser.getString();
         JsonLocation location = parser.getLocation();
         return newToken(
                  tokenmapper.typeForStringValue(data),
                  data, location);
    }
    else ...
```

Figure 5. (a) `JSONScannerWrapper` and its relationships with JSON-P and ANTLR, (b) the `TokenMapper` interface, (c) excerpt of `nextToken in JSONScannerWrapper`

In particular, we propose to describe objects by expressions in the form $\{k_1 m_1: V_1, \ldots, k_n m_n: V_n\}$, where each k_i is a distinct key name, each V_i is a EBNF expression characterizing the structure of the allowable values for k_i, and each m_i is a modifier controlling the ocurrence of key-value pairs of the kind k_i:v in actual documents. In this expression, the order of appearance of the key-value pairs does not matter. In addition, key-value pairs in the form k_i: v must occur (i) exactly one time if m_i is omitted, (ii) zero or one time if m_i is set to `?` , (iii) zero or more times if m_i is set to `*`, and (iv) one or more times if m_i is set to `+`. For instance, Figure 4a characterizes the grammatical structure of the documents involved in the graph example of the previous section using standard EBNF augmented with this convention.

- Encoding the grammatical structure in the parser generation tool. While in principle it could be possible to translate such a structure to pure EBNF (and thus, to pure BNF) notation(s), the lack of order of key-value pairs in objects can make this direct approach cumbersome, since it could involve enumerating all the possible permutations of key-value sequences (see Figure 4b). Thus, it is possible to use additional semantic facilities in the generator to facilitate such an encoding (e.g., validating semantic actions, semantic predicates …)
- Augmenting the grammar with semantic contexts and semantic actions in order to characterize the processing task as a syntax-directed translation process. The result is a translation scheme, which will be dependent on the parser generator adopted.
- Providing the additional machinery necessary to complete the processing application. Depending on the kind of application, it could include data visualization facilities, database support, a domain model to be instantiated as result of processing the JSON document, etc. In any case, it is interesting to provide a suitable façade in terms of which of the semantic actions in the translation scheme can be written. This façade will be called a *semantic module.*
- Generating the JSON processing component from its specification as a translation scheme. For this purpose, the parser generator is used.
- Gluing it all together in a suitable main program able to launch the application itself.

It is worthwhile to notice that, as a consequence of this grammar-oriented approach, applications are split into two well-differentiated layers:

- A *linguistic layer,* which is declaratively described as a translation scheme expressed in the specification language of the parser generator.
- An *application logic layer,* which is given in terms of conventional software components interfaced by the semantic module.

It leads to an interesting division of labor among developers specialized in JSON processing using formal grammars, and more conventional developers specialized in the development of more conventional application / business logics. The linguistic layer takes care of the orchestration of conventional application logic components, each of which can largely be provided in isolation from the others. In turn, this orchestration is directed by the grammatical structure that underlies the

JSON documents, and it can be described and maintained at a high level, using declarative, grammar-based, specifications, instead of being expressed in a more conventional general-purpose programming language.

IV. Grammar-Driven development with JSON-P and ANTLR

In this section we show how to enable the grammar-driven approach by combining JSON-P and ANTLR:

- JSON-P (Java API for JSON Processing) is a general-purpose JSON processing framework for Java [16]. It defines an API for mapping JSON documents into tree-like representations (the equivalent to DOM in the XML world), and another API to process JSON documents in a streaming fashion.
- ANTLR [27] is a multi-language parser generation tool, which is able to generate recursive descent parsers that combine many of the more recent parsing tendencies: the use of prediction automata for unlimited lookahead (achieved by the LL(*) parsing method), the use of semantic predicates, and the use of backtracking and tabulation to mimic *packrat* parsing [8] (see [28] for a more in-depth description of ANTLR internals). This combination of parsing technologies, together with their support for multiple implementation languages (among them, Java) makes this tool one of the more widely used worldwide.

Concerning JSON-P, this combination uses its facilities for JSON streaming processing. In particular, JSON-P provides a *pull* API similar to StAX in the XML world, which is especially well suited for its combination with ANTLR-generated parsers, since it can naturally work as a scanner for such a parser. In this way, the JSON *Scanning Wrapper* in this combination (see Figure 5a):

- Encloses a `JSONParser` instance (i.e., an instance of the *pull* streaming processing artifact provided by JSON-P)
- Can be customized by an instance of a suitable `TokenMapper` implementation, which actually characterizes the mapping tables (Figure 5b)
- Extends the ANTLR `Lexer` class. In particular, it implements the `nextToken` method to return ANTLR `CommonToken` instances representing the tokens associated with the JSON logical elements. Figure 5c shows an excerpt of this method in our combination.

```
arc locals [boolean sourceOn=false,
            boolean targetOn=false,
            boolean labelOn=false]:
 OC ( ({$sourceOn}? SOURCE STRING {$sourceOn=true;}) |
      ({$targetOn}? TARGET STRING {$targetOn=true;}) |
    ({$labelOn}? LABEL STRING {$labelOn=true;}) )* CC
  {$sourceOn && $targetOn && $labelOn}? ;
```

Figure 6. ANTLR encoding of the rule for `arc` in Figure 4a.

In this way, in the resulting environment:

- The customization of the JSON *Scanning Wrapper* involves: (i) providing a suitable implementation of the `TokenMapper` interface (each method in this interface determines how to map relevant JSON elements into types for ANTLR tokens), and (ii) specializing `JSONScannerWrapper` to use such an implementation as a customization table.
- The encoding of the grammatical structure can take advantage of ANLTR validating semantic predicates to simplify the description of object expressions in the augmented EBNF notation. Indeed, $\{k_1\ m_1\colon V_1, \ldots, k_n\ m_n\colon V_n\}$ is represented by $OC\ ((vp_1\ K_1\ V_1\ a_1) \mid \ldots \mid (vp_n\ K_n\ V_n\ a_n))^*\ CC\ v_f$ where: (i) K_i is the type of token corresponding to k_i, (ii) a_i is a semantic predicate registering the number of times that k_i has occurred, (iii) vp_i is a semantic predicate that validates whether K_i can occur, (iv) v_f is a semantic predicate validating whether all the mandatory key-value pairs have appeared, and (v) *OC* and *CC* are respectively the object opening and closing tokens. Figure 6 pro-

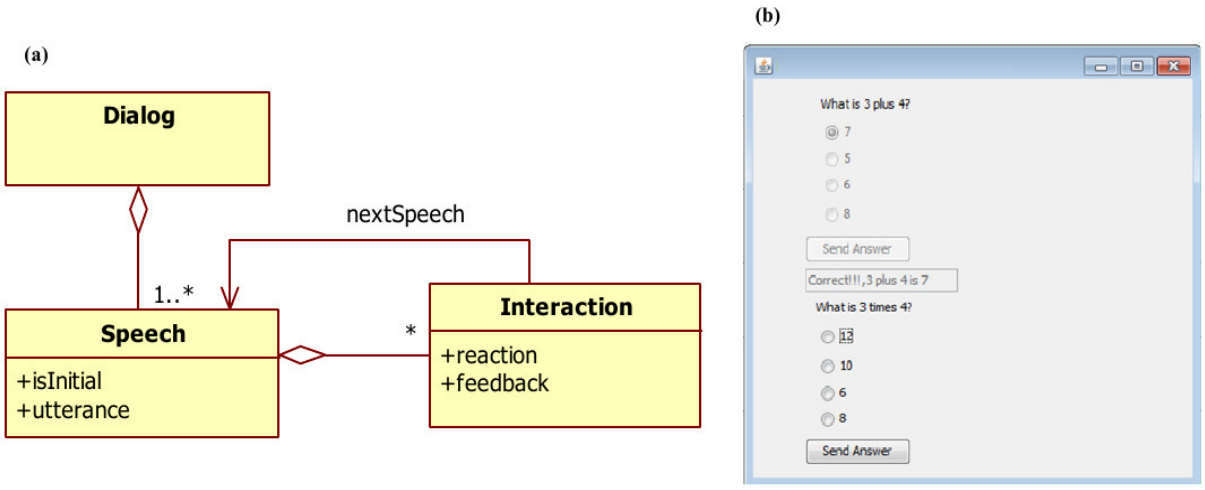


Figure 7. (a) Dialog semantic model;(b) Snapshot for the Dialog player.

vides an example concerning the EBNF structure provided in Figure 4a.
- Then, additional semantic context and semantic actions can be added to the resulting ANLTR grammar to configure the specification of the JSON processing component as a translation scheme. Once this translation scheme is processed with the ANTLR tool to yield the Java implementation, the corresponding *Lexer* class must be replaced by the customization of the JSON Processing Wrapper, in order to make the parser actually operate on the logical structure of the JSON documents.

V. Case Study: Grammar-Driven Development of a JSON-customizable Interactive Application

This section describes how we have applied the grammar-driven development model defined in the previous sections to the implementation of an interactive application oriented to play dialogs between the user and the computer[1]. These dialogs can be described using JSON documents. Subsection V.A describes how the interactive application behaves and how the dialog documents are structured. Then, subsection V.B details the grammar-driven development of this application.

A. The dialog player

The dialogs played by our application are based on the *Socratic Tutorials* developed by Prof. Alfred Bork's team during the eighties of the past century [6]. They obey the semantic model in Figure 7a. Thus, when playing a dialog (Figure 7b):
- The application proffers a *speech*. It is a chunk of text that can be read by the user.
- Then it displays a repertory of possible *interactions* and lets the user select one of them.
- When the user selects the interaction, the application gives him/her an associated *feedback.*
- Finally, it either continues with other speeches or ends the execution.

The application can be customized with dialogs described as JSON documents. Indeed, these documents are a possible concrete syntax for the semantic model depicted in Figure 7a. In this way:
- Dialogs are represented by objects with a "`Dialog`" key. The value of this key is a sequence of speeches.
- Each speech is, in turn, represented by an object that includes: (i) a mandatory "`idSpeech`" key, whose value identifies the speech in the JSON document (this value will be used for referencing the speech from other places in the document), (ii) an optional "`isInitial`" key, whose value, when `true`, indicates the speech starts the dialog, (iii) a mandatory "`utterance`" field that includes the text to be proffered by the machine, and (iv) an optional "`interactions`" field containing a sequence with all the possible interactions (if a speech without interactions is reached, the player ends the execution).
- Finally, interactions are represented by objects including: (i) a mandatory "`reaction`" key representing the actual text of the interaction, (ii) a mandatory "`feedback`" key representing the text of the feedback, and (iii) a mandatory "`nextSpeech`" representing the speech that continues the dialog; its value can be either another speech description, or the *id* of other speech in the dialog.

Figure 8 shows an excerpt of JSON document describing a dialog.

```
{"Dialog":[{
   "idSpeech":"1",
   "isInitial":true,
   "utterance":"what is 3 plus 4?",
   "interactions":[
          {
          "reaction":"7",
          "feedback":"Correct!!,3 plus 4 is 7.",
          "nextSpeech":{
                "idSpeech":"2",
                "isInitial":false,
                "utterance":"what is 3 times 4?",
                "interactions":[{
                        "reaction":"12",
                        "feedback":"Correct!!,3 times 4 is 12.",
                        "nextSpeech":3 },{
                        "reaction":"10",
                    "feedback":"Sorry!!, 3 times 4 is not 10. Try again",
                        "nextSpeech":2}, {
                        "reaction":"6",
                        "feedback":"Sorry!!, 3 times 4 is not 6.Try again"
                        "nextSpeech":2},
....
```

Figure 8. Excerpt of a Dialog Description JSON Document.

B. Grammar-Driven Development

The operation of the dialog player described in the previous subsection is as follows:
- It processes the input JSON document in order to instantiate the semantic model in Figure 7a.
- Then it plays the dialog by a direct interpretation of the semantic model instance.

While the interpretation stage is straightforward once the semantic model has been instantiated, the instantiation process is considerably more cumbersome, due in part to the changing and evolving nature of the concrete JSON encoding. Thus, the player can be meaningfully architected according to the grammar-driven approach as follows:
- The instantiation of the semantic model is developed in grammatical terms, using JSON-P and ANTLR.
- The semantic module is implemented as a façade class providing instantiation operations as methods, as well as a couple of tables required during the instantiation process: one table mapping *ids* into `Speech` instances, and another table mapping ids into `Interation` instances whose next speeches are those associated to such *ids.*
- The semantic model itself, along with the player shell, constitute the application-specific logic.

1It is actually a simplification of the real system in order to fit the space constraints of the paper. The actual system is closer to that described in [33]

(a)

```
public class DialogTokenMapper implements TokenMapper {
  public int typeForKeyName(String keyName) {
  if (keyName.equals("idSpeech"))
     return DialogParser.IDSPEECH;
  if (keyName.equals("isInitial"))
     return DialogParser.IS_INITIAL;
  ...
  }
  ...
}
```

(b)

```
public class DialogLexer extends JSONScannerWrapper {
  public DialogLexer(CharStream in) {
       super(in,new DialogTokenMapper());
  }
}
```

(c)

```
grammar Dialog;
@parser::header {import java.util.*;}
@parser::members {private DialogSemM sem =
                                  new DialogSemM();}
dialog returns [Dialog d]:
  {List<Speech> speeches = sem.newSList();}
  (s=speech {speeches.add($s.speechvalue);} )+
  {$d = sem.mkDialog(speeches);} ;

speech returns [Speech speechvalue]
  locals[boolean id=false, boolean isInitial = false,
         boolean ut=false, boolean inter=false]:
  OO (
    ({!$id}? IDSPEECH  sid=STRING {$id=true;} )      |
    ( {!$isInitial}? IS_INITIAL bool {$isInitial=true;} ) |
    ( {!$ut}? UTTERANCE sut=STRING {$ut=true;})            |
    ( {!$inter}? INTERACTIONS interactions {$inter=true;} )
  )* CO
  {$id && $ut}?
  {$speechvalue = sem.newSpeech($sid.text, $isInitial?$bool.b:false,
                    $sut.text, ($inter)? $interactions.inters: null); } ;

interactions returns [List<Interaction> inters] :
   {$inters = sem.newIList(); }
   (i=interaction {$inters.add($i.inter);})* ;

interaction returns [Interaction inter]
   locals[boolean reaction=false, boolean feedback=false,
          boolean ns = false]:
   {Speech sp=null;}
   OO (
        ({!$reaction}? REACTION r=STRING {$reaction=true;}) |
        ({!$feedback}? FEEDBACK f=STRING {$feedback=true;}) |
        ({!$ns}? NEXT_SPEECH (s = speech {sp=$s.speechvalue;}|
                              ref=STRING
                              {sp = sem.getSpeech($ref.text);})
         {$ns=true;})
   )* CO
   {$reaction && $feedback && $ns}?
   {$inter = sem.newInteraction($r.text,$f.text,sp);
    if (sp == null)
       sem.toBePatched($ref.text,$inter);} ;
bool returns [boolean b] : TRUE {$b=true;} | FALSE {$b=false;};
   // Lexical rules will be not used, but they are necessary
   // for completing the ANTLR grammar
IDSPEECH : 'IDSPEECH';
IS_INITIAL : 'ISINITIAL';
...
```

(d)

```
public class DialogSemM {
  private Map<String,Speech> speeches;
  private Map<String,List<Interaction>> patchMap;

  public DialogSemM() {
     speeches = new HashMap<>();
     patchMap = new HashMap<>();
 }

  public Dialog mkDialog(List<Speech> speeches) {
   return new Dialog(speeches);
  }
  public List<Speech> newSList() {
     return new LinkedList<Speech>();
  }
  public List<Interaction> newIList() {
     return new LinkedList<Interaction>();
  }
  public Speech newSpeech(String id, boolean isInitial,
                          String ut,List<Interaction> is)
     Speech sp = new Speech(isInitial,ut,is);
     speeches.put(id,sp);
     List<Interaction> pl = patchMap.get(id);
     if (pl != null) {
       for(Interaction i: pl)
           i.putNextSpeech(sp);
       patchMap.put(id,null);
     }
     return sp;
  }
  public Speech getSpeech(String sid) {
   return speeches.get(sid);
  }
  public Interaction newInteraction(String r, String f,
                                    Speech ns) {
     return new Interaction(r,f,ns);
  }
  public void toBePatched(String si,Interaction i) {
     List<Interaction> pl = patchMap.get(i);
     if (pl == null) {
      pl = new LinkedList<Interaction>();
      patchMap.put(si,pl);
     }
     pl.add(i);
  }
}
```

Figure 9. (a) Implementation of the mapping table for the Dialog case-study; (b) specialization of the JSONScannerWrapper ; (c) ANTLR grammar for the processing of JSON Dialog Documents; (d) semantic module.

- The main gluing program performs the instantiation, and then activates the player with the resulting semantic model instance.

Thus, the organization is very similar to that of applications built using DSL construction frameworks such as Eclipse XText [7] (in this case, input descriptions are encoded in JSON instead on a domain-specific syntax, however).

Concerning development details, Figure 9a shows an excerpt of the token mapping table. As indicated in Section IV, it involves implementing the `TokenMapper` interface, as made apparent in Figure 9a. Notice that, in this implementation, actual token codes are taken from the `DialogParser` class. This will be the parser class generated by ANTLR. Therefore, token names must be kept consistent throughout this mapping table and the subsequent ANTLR grammar.

Once the mapping table is available, it is possible to customize the JSON Scanner Wrapper. As indicated in Section IV it involves to subclass `JSONScannerWrapper` in order to install an instance of the mapping table provided (Figure 9b). The name given to this subclass must be consistent with the name of the lexer to be generated by ANTLR.

Next step, the most relevant one, is to characterize the syntactic structure of the JSON documents, then to encode this structure as an ANTLR grammar following the patterns given in Section IV, and finally to augment this grammar with suitable semantic actions. Figure 9c shows the resulting ANTLR translation scheme.

Then the semantic class that implements the semantic module can be provided (Figure 9d). In this class, in addition to creating a new speech, the `newSpeech` method back-patches all the interactions referring to such a speech, which is consistent with the usage of the operations in the ANTLR grammar.

Next step is to generate all the parsing code from the ANTLR grammar, and to replace the `DialogLexer` generated by that shown in Figure 9b. Finally, the application-specific logic and the main launching program must be provided, which constitutes a routine programming task.

VI. Conclusions And Future Work

In this paper, we have shown how to combine generic, stream-oriented, JSON processing frameworks with parser generators in order to facilitate the development of JSON processing applications. The resulting approach is aware of the grammatical nature of JSON documents and enables the specification of JSON processing tasks at a higher and more declarative level than that provided by general-purpose programming languages. Contrary to proposals like [9], formal grammars in our proposal operate on the logical structure of JSON documents instead of on the raw text of these documents. In this sense, our proposal is aligned with our previous works in XML processing [31][32], in which we proposed similar grammar-driven models for processing XML using grammars and parser generators.

Currently we are working on the implementation of an environment for providing more assistance to our grammar-driven development process model. We also are planning to use attribute grammars [20][26] as specification mechanisms of JSON processing tasks, paralleling our previous work in the XML world [31]. Finally, and although our firsts tests with developers are satisfactory, we plan to carry out a more systematic comparative study of our approach with more conventional approaches to JSON processing.

Acknowledgment

This work was partially supported by the project grant TIN2010-21288-C02-01.

References

[1] Abelson, H., Sussman, G.J. (1993). *Structure and Interpretation of Computer Programs*. MIT Press

[2] Aho, A.V., Lam, M.S., Sethi, R., Ullman, J.D. Compilers: principles, techniques and tools (2nd ed.). Addison-Wesley. 2007

[3] Appel, A.W. Modern Compiler Implementation in Java (2002). Cambridge University Press

[4] Berg, J. (2012). Utvärdering av bibliotek för generering och "parsning" av JSON. Degree Dissertation. KTH

[5] Beyer, K. S., Ercegovac, V., Gemulla, R., Balmin, A., Eltabakh, M., Kanne, C. C., ... & Shekita, E. J. (2011). Jaql: A scripting language for large scale semistructured data analysis. In Proceedings of 37th VLDB Conference.

[6] Bork, A. 1985. Personal Computers for Education. New York, NY, USA: Harper & Row Publishers, Inc.

[7] Eysholdt, M., Behrens, H (2010). Xtext: implement your language faster than the quick and dirty way. ACM international conference on Object Oriented Programming Systems Languages and Applications Companion (SPLASH '10), 307-309.

[8] Ford, B (2002). Packrat Parsing : Simple, Powerful, Lazy, Linear Time, Functional Pearl. 17th ACM SIGPLAN international conference on Functional programming, pp. 36-47.

[9] Gerasika, A (2011). How to convert JSON to XML using ANTLR. http://www.gerixsoft.com/blog/xslt/json2xml2 (last access: April 11, 2013)

[10] Goessner, S (2007). JSONPath – XPath for JSON. http://goessner.net/articles/JsonPath/ (last access : April 11, 2013)

[11] Gson. Google-gson - A Java library to convert JSON to Java objects and vice-versa. https://code.google.com/p/google-gson/ (last access: April 11, 2013)

[12] Han, J., Haihong, E., Le, G., & Du, J. (2011, October). Survey on NoSQL database. 6th international conference on Pervasive computing and applications (ICPCA'11), pp. 363-366.

[13] iJSON. https://pypi.python.org/pypi/ijson/ (last access: April 11, 2013)

[14] Jackson. http://jackson.codehaus.org/ (last access: April 11, 2013)

[15] JSON. http://www.json.org/ (last access: April 11, 2013)

[16] JSON-P. Java API for JSON Processing (JSON-P). http://json-processing-spec.java.net/ (last access: April 11, 2013)

[17] JSON-Schema. http://json-schema.org/ (last access: April 11, 2013)

[18] JSONSelect . http://jsonselect.org/ (last access : April 11, 2013)

[19] Json-simple. Json-simple – A simple Java Toolkit for JSON. https://code.google.com/p/json-simple/ (last access: April 11, 2013)

[20] Knuth, D. E. Semantics of Context-free Languages. Mathematical System Theory 2(2), 127–145. 1968.

[21] Kodaganallur, V (2004). Incorporating language processing into Java applications: a JavaCC tutorial. IEEE Software 21(4), 70-77.

[22] Litjson. http://lbv.github.io/litjson/ (last access : April 11, 2013)

[23] Maeda, K. (2012). Performance evaluation of object serialization libraries in XML, JSON and binary formats. 2nd Conference on Digital Information and Communication Technology and its Applications (DICTAP), pp. 177-182.

[24] Membrey, P., Plugge, E., & Hawkins, T. (2010). The definitive guide to MongoDB: the noSQL database for cloud and desktop computing. Apress.

[25] Nurseitov, N., Paulson, M., Reynolds, R., & Izurieta, C. (2009). Comparison of JSON and XML data interchange formats: A case study. Computer Applications in Industry and Engineering (CAINE), 157-162.

[26] Paakki, J. Attribute Grammar Paradigms – A High-Level Methodology in Language Implementation. ACM Computing Surveys, 27, 2, 196-255. 1995

[27] Parr, T (2007). The Definitive ANTLR Reference: Building Domain-Specific Languages. Pragmatic Bookshelf.

[28] Parr, T., Fisher, K (2011). LL(*): the Foundation of the ANTLR Parser Generator. 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI'11), pp. 425-436.

[29] Rapidjson. Rapidjson - A fast JSON parser/generator for C++ with both SAX/DOM style API. https://code.go ogle.com/p/rapidjson/ (last access: April 11, 2013)

[30] Richardson, L., & Ruby, S. (2007). RESTful Web Services. O'Reilly.

[31] Sarasa-Cabezuelo, A., Sierra, J.L. (2013). The grammatical approach: A syntax-directed declarative specification method for XML processing tasks. Comp. Stand. & Interfaces 35(1), 114-131

[32] Sarasa-Cabezuelo, A., Temprado-Battad, B., Rodríguez-Cerezo, D., Sierra, J. L. (2012). Building XML-driven application generators with compiler construction tools. Computer Science and Information Systems, 9(2), 485-504.

[33] Sierra, J.L., Fernández-Valmayor, A., Fernández-Manjón, B (2008). From Documents to Applications Using Markup Languages. IEEE Software, 25(2), 68-76

[34] Zakas, Z. N (2012). Professional JavaScript for Web Developers 3rd Edition. Wrox Press.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1553–1558

Alvis Language with Time Dependence

Marcin Szpyrka, Piotr Matyasik, Michał Wypych
AGH University of Science and Technology
Department of Applied Computer Science
Al. Mickiewicza 30, 30-059 Krakow, Poland
Email: {mszpyrka,ptm,mwypych}@agh.edu.pl

***Abstract*—The paper presents the semantics for the time version of the Alvis modelling language. Alvis combines possibilities of formal models verification with flexibility and simplicity of practical programming languages. The considered time Alvis language is suitable for formal verification of real-time systems. The paper contains description of: the Alvis time model, states and transitions between states and snapshot reachability graphs that represent models state spaces in the form of directed graphs.**

I. Introduction

COSTS of creating and maintaining embedded software draw attention of producers to formal methods. There are more and more attempts to provide methods and tools to improve the concurrent systems development [6], [9]. Alvis combines a formal approach with engineering-like look and style. It is hiding most of the formal side from users but not losing any part of it. Alvis is a modelling language being developed at AGH-UST in Krakow, Department of Applied Computer Science (*http://fm.kis.agh.edu.pl*).

Previous research on Alvis has been mainly concerned with the untimed version of the language with α^0 system layer (multiprocessor environments). The syntax of Alvis which is common for all language versions can be found in [21]. Formal semantics of the untimed version of Alvis has been presented in [22]. This version of Alvis has been successfully used for formal verification of concurrent systems e.g. for BPMN models [23] which may include rule-based systems [18] designed with the XTT2 method [11], [15] or as D-nets [24].

The aim of the paper is to present a draft of semantics for the time version of Alvis with α^0 system layer which allows users to assign to every model statement its duration. Then the set of reachable states of such a model is represented in the form of SR-graph and is used for its verification with model checking techniques [2]. SR-graphs provide the possibility of formal verification of real-time requirements. In contrast to other formalisms like time automata [1], Petri nets with time [10], [17] or multi-agent systems [5], Alvis syntax is very similar to procedural programming languages and the method of model states description is similar to information provided by software debuggers. The idea of SR-graphs has been shortly introduced in [19]. This paper contains formalised and more detailed description of it.

II. Alvis at a Glance

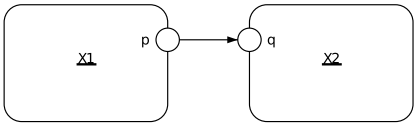
Alvis combines advantages of high level programming languages with a graphical language for modelling interconnections between subsystems (called agents) of a concurrent system. Agents are divided into *active* and *passive*. *Active agents* perform some activities and are similar to tasks in Ada programming language [4]. By contrast, *passive agents* do not perform any individual activities and are similar to protected objects (shared variables). Passive agents provide other agents with a set of procedures (services). An Alvis model is composed of three layers. A *communication diagram* (graphical layer) is used to describe a modelled system from the control and data flow point of view. Examples of such diagrams are given in Fig. 1 and 6. Active agents are drawn as rounded boxes while passive ones as rectangles. Ports used for communication are drawn as circles placed at the edges of the corresponding figures. Alvis agents communicate with each other using communication channels drawn as lines. The code layer is used to define behaviour of agents. It uses a set of Alvis statements and some elements of the Haskell functional programming language [16]. Despite of the fact that Alvis has its origin in the CCS process algebra [14] and the XCCS language [3], [20], it does not use algebraic equations to describe the behaviour of agents but a high level programming language. The *system layer* is predefined and defines the hardware environment for a model. In this paper we consider models with the α^0 *system layer* that denotes that each active agent has access to its own processor and if possible agents perform their steps concurrently. For more details see [21] or the project website.

An Alvis model semantics find expression in a *labelled transition system* (LTS graph). Execution of any language statement is expressed as a transition between formally defined states of such model. An LTS graph is an ordered graph with nodes representing states of the considered system and edges representing transitions among states. Examples of Alvis LTS graphs are given in Fig. 2, 3, 5 and 7. Alvis LTS graphs can be verified using the CADP toolbox [8]. We use CADP *evaluator* tool to check whether the model satisfies requirements given as regular alternation-free μ-calculus formulas [7], [13].

III. Alvis Time Model

The Alvis time model is based on the idea of a *global clock* used to measure the duration of model steps. The language provides carefully selected set of statements sufficient to describe the behaviour of individual agents. Each of them can have duration assigned which is provided by a user as

978-83-60810-53-8/$25.00




```
agent X1 { loop { out p; } }
agent X2 { loop { in q; } }
```

Figure 1. Communication between active agents

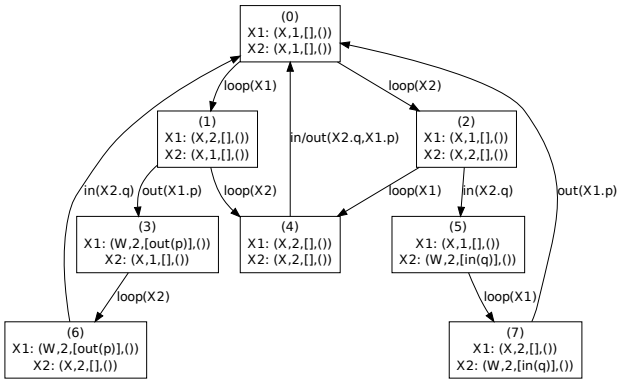


Figure 2. LTS graph for model in Fig. 1

a verification parameter. The time units used in a given model are strictly connected with the model interpretation.

Let us consider the simple model of two communicating active agents shown in Fig. 1. Each agent performs two steps: entering a loop and a communication. Agent $X1$ sequentially sends signals via port $X1.p$ ($X1.p$ denotes port p of agent $X1$), while agent $X2$ sequentially collects signals via port $X2.q$. If an untimed Alvis language is considered, the LTS graph represents all possible execution paths as shown in Fig. 2. The LTS graph labels point out steps performed by agents.

Definition 1: A *state of an agent* X is a tuple

$$S(X) = (am(X), pc(X), ci(X), pv(X)) \tag{1}$$

where $am(X)$, $pc(X)$, $ci(X)$ and $pv(X)$ denote *agent mode*, *program counter*, *context information list* and *parameters values* of the agent X respectively.

The following modes are possible. *Finished* (F) means that an agent has finished its work. *Init* (I) is the default mode for agents that are inactive in the initial state. *Running* (X) means that an agent is performing one of its statements. *Taken* (T) means that one of the passive agent procedures has been called and the agent is executing it. For passive agents *waiting* (W) means that the corresponding agent is inactive and is waiting for another agent to call one of its accessible procedures. For active agents the mode means that the corresponding agent is waiting either for a communication with another active agent or for a currently inaccessible procedure of a passive agent.

The *program counter* points out the current statement of an agent. The context information list contains additional

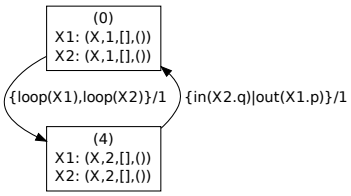


Figure 3. LTS graph for model in Fig. 1 – timed version

information about the current state e.g. if an agent is in the *waiting* mode, ci contains information about events the agent is waiting for. The set of admissible entries used in ci lists is given in Table II. The *parameters values list* contains the current values of the corresponding agent parameters, if such parameters (variables) have been defined in the agent code.

A state of a model is represented as a sequence of agents states [22], [12]. We will use letter S with possible index to denote states. If necessary am, pc, ci, pv will be indicated by indexes S, S' etc. to point out the state they refer to.

If durations of steps are taken under consideration, we cannot consider states of a system in the same way as previously. For example, state 3 in the untimed LTS graph shown in Fig. 2 represents the situation when agent $X1$ has already finished two of its steps, while agent $X2$ still remains in its initial state. Such situation is not possible in time models. Assume steps durations for all steps in the considered model are equal to 1. It means that both agents start execution of their first steps in the same time, so after 1 time-unit the system changes its state from 0 to 4. Finally, the LTS graph for the model is reduced to the one shown in Fig. 3. Labels of edges in the presented graph are of the form $steps/t$, where t stands for the duration of the steps performed simultaneously. The change of the state from 4 to 0 is the result of synchronous communication between agents which is denoted by symbol `|` used instead of a comma.

Alvis uses three statements that use time explicitly:

- **`delay`** `t` – postpones an agent for a given time;
- **`alt`**`(`**`delay`** `t) {...}` – defines a branch of the *select* statement that is open after the given time;
- **`loop`** `(`**`every`** `t) {...}` – repeats loop contents every specified number of time-units.

Let us focus on the *step* idea. It is necessary to distinguish between code statements and steps. Most of the Alvis statements e.g. *exec*, *exit*, etc. are *single-step* statements. By contrast, *if*, *loop* and *select* are *multi-step* statements. We use recursion to count the number of steps for multi-step statements. For each of them, the first step enters the statement interior. Then we count steps of statements put inside curly brackets. From theoretical point of view steps are described as transitions. The formal description of Alvis provides definitions of results of any transition execution. Such formal semantics for untimed models is presented in [22]. The time aspect of transitions is considered in Section IV.

Suppose the code layer for the communication diagram in Fig. 1 is implemented as shown in Fig 4. Agent $X1$ starts its

```
agent X1 {
  loop (every 10) { out p; }}        -- 1, 2

agent X2 {
  loop {                             -- 1
    select {                         -- 2
      alt (ready [in(q)]) {
        in q; delay 1; }             -- 3, 4
      alt (delay 2) { null; }}}}     -- 5
```

Figure 4. Communication between active agents version 2 – new code layer for communication diagram in Fig. 1

Table I
STEP DURATION FOR MODEL IN FIG. 4

Agent $X1$	Step duration	Agent $X2$	Step duration
loop every	1	*loop*	1
out	3	*select*	2
		in	2
		delay	1
		null	1

loop every 10 time-units and sends a signal via port p inside the loop. Behaviour of agent $X2$ is defined as an infinite loop with a **select** statement inside. The statement contains two branches. First one is open (can be performed) if port q can be immediately used to collect a signal (i.e. agent $X1$ has already sent a signal via port p which is connected with q). Inside the branch agent $X2$ collects a signal via port q and is postponed for 1 time-unit. Second branch is open 2 time-units after entering the **select** statement. Inside the branch agent $X2$ performs the empty statement.

Assume steps durations for all steps performed by agents $X1$ and $X2$ are defined as given in Table I. Let us focus on the initial state $S_0 = ((\mathsf{X}, 1, [\,], ()), (\mathsf{X}, 1, [\,], ()))$. When the α^0 system layer and timed Alvis language are considered it is assumed that agents execute their steps as soon as possible. Thus, both agents are running their first steps ($loopevery(X1)$ and $loop(X2)$) concurrently and after one time-unit the state $S_1 = ((\mathsf{X}, 2, [timer(1, 9)], ()), (\mathsf{X}, 2, [\,], ()))$ is received. The $timer(1, 9)$ entry used in $X1$ agent context information list points out that the next loop course can start after 9 time-units. There are two steps $out(X1.p)$ and $select(X2)$ enabled in the state 1. Because step $out(X1.p)$ takes 3 time-units, while $select(X2)$ takes 2 time-units, we cannot present the result of these transitions execution as a state similar to state 1. After 2 time-units the step $out(X1.p)$ is still under execution and after 3 time-units when step $out(X1.p)$ is finished, agent $X2$ could be executing another step – this is not the case in this model due to the lack of an open branch for the **select** statement. The solution for the problem is a *snapshot* [19] i.e. a state that presents the considered system with some steps under execution. We can take a snapshot every 1 time-unit but we are interested only in such snapshots when at least one step has finished its execution.

An LTS graph with snapshots will be called *snapshot reachability graph* or SR-graph for short. A part of the SR-graph for the model in Fig. 4 is shown in Fig. 5. State 2 represents the time point when agent $X2$ has finished step 2 and is waiting for an open branch of the **select** statement, while agent $X1$ is still performing step $out(X1.p)$. The $sft(n)$ (*step finish time*) entry used in $X1$ context information list points out the number of time-units necessary to finish the current step. State 4 represents the time point when agent $X1$ is waiting for a timer event to restart the loop. The event will be generated in 4 time-units. State 5 represents the time point when both agents are waiting for timers' events. Agent $X2$ is waiting for the end of the postpone time. The $time$ label of the edge from state 5 to 6 denotes the passage of time.

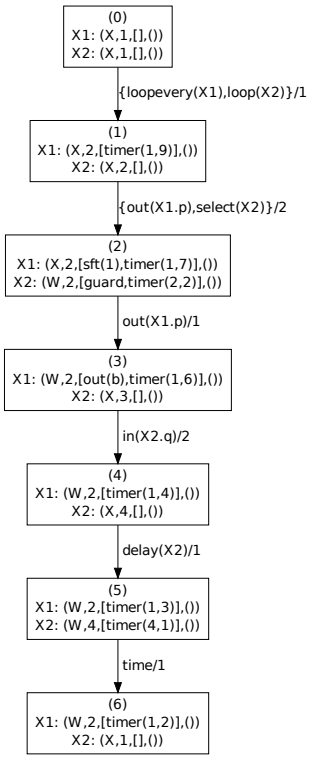


Figure 5. Part of SR-graph for model in Fig. 4

IV. TRANSITIONS

Let $\mathcal{P}$ denote the set of all model ports. For this paper we define Alvis models as follows [22].

Definition 2: A *communication diagram* is a triple $D = (\mathcal{A}, \mathcal{C}, \sigma)$, where: $\mathcal{A} = \{X_1, \ldots, X_n\}$ is the set of *agents* consisting of two disjoint sets, $\mathcal{A}_A$, $\mathcal{A}_P$ such that $\mathcal{A} = \mathcal{A}_A \cup \mathcal{A}_P$, containing *active* and *passive* agents respectively; $\mathcal{C} \subseteq \mathcal{P} \times \mathcal{P}$ is the *communication relation*, such that: (1) a connection cannot be defined between ports of the same agent; (2) procedure ports are either input or output ones i.e. ports defined as procedures are used to transfer signals (values) either to or from a passive agent; (3) a connection between an active and a passive agent must be a procedure call; (4) a connection between two passive agents must be a procedure call from a non-procedure port. Function $\sigma\colon \mathcal{A}_A \to \{False, True\}$ is the

start function that points out initially activated agents.

Definition 3: An Alvis *model* is a triple $\mathbf{A} = (D, B, \alpha^0)$, where $D = (\mathcal{A}, \mathcal{C}, \sigma)$ is a *communication diagram*, B is a syntactically correct *code layer*, and α^0 is the α^0 *system layer*. Moreover, each agent X belonging to the diagram D must be defined in the code layer and each agent defined in the code layer must belong to the diagram.

Definition 4: A *state* of a model $\mathbf{A} = (D, B, \alpha^0)$, where $D = (\mathcal{A}, \mathcal{C}, \sigma)$ and $\mathcal{A} = \{X_1, \ldots, X_n\}$ is a tuple $S = (S(X_1), \ldots, S(X_n))$. The *initial state* is defined as follows:

- $am(X) = \mathsf{X}$, for any active agent X such that $\sigma(X) = True$; $am(X) = \mathsf{I}$, for any active agent X such that $\sigma(X) = False$; $am(X) = \mathsf{W}$, for any passive agent X;
- $pc(X) = 1$ for any active agent X in the *running* mode and $pc(X) = 0$ for other agents.
- $ci(X) = [\,]$ for any active agent X; and $ci(X)$ contains names of accessible procedures for any passive agent X.
- For any agent X, $pv(X)$ contains X parameters with their initial values.

Table II contains all possible entries that can be included into a context information list and the relationships between the entries and an agent mode.

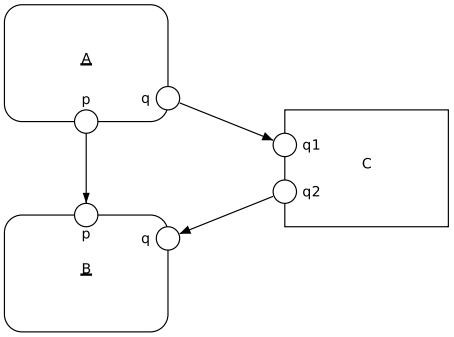
Let $B(X)$ denote an agent X code, $card(B(X))$ denote the number of *steps* in $B(X)$, $B_i(X) \in \{delay, exec, exit, if, in, jump, loop, loopevery, null, out, select, start\}$ denote the name of the agent X i-th step, and $\mathcal{N}(t)$ denote the name of the transition t (possible values are the same as for steps). The set of all transitions available for a particular model will be denoted by $\mathcal{T}$. Moreover, let $\Delta(X, k)$ denote the duration of the k-th step of agent X.

Let us consider the model given in Fig. 6. It contains two active agents A and B that can communicate directly or using passive agent C. Agent A inside the infinite loop performs a **select** statement and waits at most 3 time-units for a communication via port p. In case of a timeout, agent A sends a signal via port q. Agent B inside its periodic loop picks 0 or 1 at random and depending on the result collects a signal via port q or p. Agent C provides two procedures that are accessible depending on the value of parameter n. It works like a buffer for a single signal. The comments included into the code contain steps numbers and durations.

Definition 5: Assume $\mathbf{A} = (D, B, \alpha^0)$ is an Alvis model with the current state S and $X \in \mathcal{A}_A$. A transition $t \in \mathcal{T}$ is *enable* in the state S *with respect to* X if and only if X is in the *running* mode, the program counter points out step t, X has not called a procedure and the step t is not already in progress. The fact that a transition t is enabled in a state S with respect to an agent X and that a state S' is the result of executing t in S will be denoted by $S{-}t(X){\rightarrow}S'$.

The paper [22] contains formal description of all possible transitions for untimed Alvis models. In this section we will focus on description of the differences between untimed and time versions of the language.

Let $pv_S(X)|_{x=w}$ denote the list of parameters values $pv_S(X)$, but with the parameter x assigned to a new value w. If $X \in \mathcal{A}_A$, $S{-}t_{exec}(X){\rightarrow}S'$, and a parameter x is




```
agent A {
  loop {                                        -- 1/1
    select {                                    -- 2/1
      alt (ready [out(p)]) { out p; } -- 3/2
      alt (delay 3) {out q; } } } }    -- 4/2

agent B {
  i :: Int = 0;
  loop (every 6) {                              -- 1/1
    i = pick [0,1];                             -- 2/1
    if(i == 1) { in q; }          -- 3/1, 4/3
    else { in p; } } }                          -- 5/2

agent C {
  n :: Bool = False;
  proc q1 (n == False) {
    in q1; n = True; }            -- 1/2, 2/1
  proc q2 (n == True) {
    out q2; n = False; } }        -- 3/2, 4/1
```

Figure 6. A time model with a passive agent

assign a value w with the corresponding *exec* statement, then for an untimed model the state S' is defined as folows: $S'(X) = (\mathsf{X}, nextpc(S(X)), ci_S(X), pv_S(X)|_{x=w})$, if $nextpc(S(X)) \neq 0$, and $S'(X) = (\mathsf{F}, 0, [\,], pv_S(X)|_{x=w})$ otherwise, where $nextpc$ function determines the next program counter for an agent [22]. Moreover, $S'(Y) = S(Y)$ for any other agent Y.

The transition is defined in a similar way for the time Alvis language. The basic difference concerns ci list with entries referring to time. Let Δ denote the duration of the considered step. Then, in case of $nextpc(S(X)) \neq 0$, we have $S'(X) = (\mathsf{X}, nextpc(S(X)), update(ci_S(X), \Delta), pv_S(X)|_{x=w})$, where the function $update$ replaces entries $timer(s, n)$ with $timer(s, n - d)$ if $n > d$ and with $timeout(s)$ otherwise.

It should be stressed that the $update$ function must be applied to context information lists of all agents in the considered model but it is not enough to determine the new state for the model. If after an ci update the list contains a $timeout(s)$ entry and the agent is in the *waiting* mode in the current state, then the corresponding agent may change its mode (to *running*) and program counter. For example, after execution of the **delay** d statement, agent switches to the *waiting* mode. Then after d time-units (if the statement is not the last one in the main block or a procedural block) the agent switches back to the

Table II
RELATIONSHIPS BETWEEN THE MODE AND THE CONTEXT INFORMATION LIST OF AN AGENT

agent X	$am(X)$	$ci(X)$ entry	description
active passive	X T	$sft(n)$	the current step will be finished in n time-units
active	X, W	$proc(Y.b, a)$	X has called the $Y.b$ procedure via port a and this procedure is being executed in the X agent context
active passive	X, W T	$timer(n, t)$ $timeout(n)$	a time event for the step number n will be generated in t time-units a time event for the step number n has been generated but it has not yet been served
active passive	W T	$in(a)$, $in(a\|T)$ $out(a)$, $out(a\|T)$ $guard$	X waits for a communication via port a (a is the input port for the communication); T is the type of the expected value X waits for a communication via port a (a is the output port for this communication) X waits for an open branch of a *select* statement
passive	T	$proc(Y.b, a)$	X has called the $Y.b$ procedure via port a and this procedure is being executed in the same context as X
passive	W	$in(a)$ $out(a)$	input procedure a is accessible output procedure a is accessible

running mode, its program counter is set to the next value and the $timeout(s)$ entry associated with the considered statement is removed from the ci list. When the $timeout(s)$ entry is *consumed* immediately then it does not appear at all in the agent state in the SR-graph.

The process of a new state determination is complex due to necessity of consideration of all concurrent steps and optimisation of the number of states in the SR-graph. The optimisation refers to skipping snapshots that differ from their predecessors only in parameters of sft and $timer$ entries in the corresponding context information lists. We can distinguish the following stages of a new SR-graph node generation:

- determination of the set $\mathcal{T}_1$ of all transitions that are in progress;
- determination of the set $\mathcal{T}_2$ of all transitions that start performing a new step;
- determination of the new state S' on the assumption that all steps from $\mathcal{T}_1 \cup \mathcal{T}_2$ are performed concurrently.

A state S is called *dead*, iff sets $\mathcal{T}_1$ and $\mathcal{T}_2$ are empty in S and does not exist an agent with ci list containing a $timer$ entry.

Assume all steps have assigned non-zero durations. Firstly, we determine the new state S' as the state 1 time-unit later than S. If state S' differs from S only in parameters of sft and $timer$ entries (parameters are decreased by 1) then we skip that state and calculate the new state 2 time-units later than S, etc. Otherwise, the state S' is a new node in the SR-graph.

If we allow zero duration for at least one step then as additional state *separators* are used changes of agents program counters values. In other words, a label in an SR-graph cannot contain two steps performed by the same agent.

Let us focus on t_{delay} and $t_{loopevery}$ transitions. Suppose, $X \in \mathcal{A}_A$, $S{-}t_{delay}(X){\rightarrow}S'$, $d > 0$ is the argument of the **`delay`** statement and Δ is the duration of the considered step. Then: $S'(X) = (\mathsf{W}, pc_S(X), update(ci_S(X), \Delta) \oplus timer(pc_S(X), d), pv_S(X))$, where $\oplus$ adds the $timer$ entry at the end of the list.

Suppose, $X \in \mathcal{A}_A$, $S{-}t_{loopevery}(X){\rightarrow}S'$ and $d > 0$ is the loop period. Then: $S'(X) = (\mathsf{X}, nextpc(S(X)), update(ci_S(X) \oplus timer(pc_S(X), d), \Delta), pv_S(X))$.

Activity of passive agents is defined similarly as for active ones but a passive agent context (i.e. the active agent that called the procedure in progress) must be taken under consideration [22].

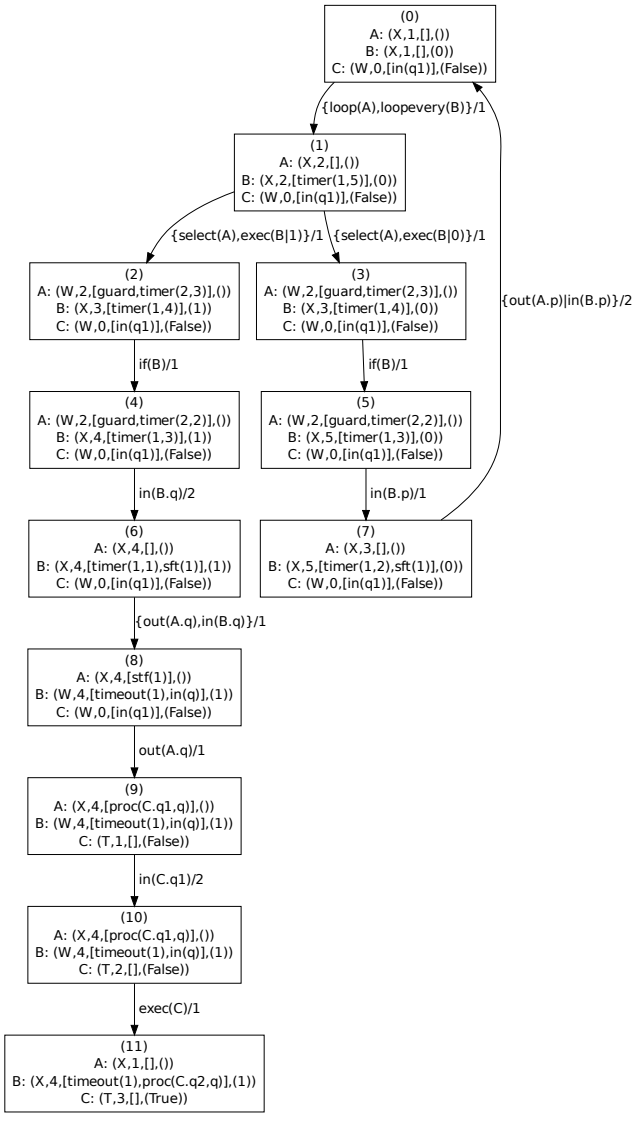


Figure 7. Time model – part of SR-graph

To illustrate presented definitions let us consider a part of the SR-graph for the considered model that is shown in Fig. 7. Let us consider sample states and transitions between them.

- Edge $0 \rightarrow 1$: Steps $loop(A)$ and $loopevery(B)$ are executed simultaneously.
- State 1: The ci list of agent B contains entry $timer(1,5)$ referring to the periodic loop.
- States 2 and 3: The states differ in the value of the parameter of agent B. After execution of **`select`** step, agent A switches to *waiting* mode, because all branches are closed; ci list contains $guard$ and $timer(2,3)$ entries, because A waits either for guard satisfaction or timeout.
- Edge $4 \rightarrow 6$: The duration of $in(B.q)$ step is 3 time-units, but state 6 is present in the SR-graph, because of agent A state change. After lapse of 2 time-units (referring to state 4) entry $timer(2,2)$ was updated to $timeout(2)$ and the agent switched mode to *running* and its program counter was set to 4. At the same time ci list of agent B contains $sft(1)$ entry.
- Edge $5 \rightarrow 7$: In the case of time models readiness of a port for a communication is stated just after commencement (rather than completion) of a communication via this port. After the lapse of 1 time-unit from starting executing $in(B.p)$ step, the condition of the first branch of **`select`** statement is satisfied, so agent A switches to *running* mode and performs steps from the branch.
- Edge $7 \rightarrow 0$: Steps $out(A.p)$ and $in(B.p)$ are performed at the same time as a synchronous communication. In time models communication is considered as synchronous, when intervals of execution of steps in and out overlap partially at least. Such communication is completed when both steps are finished.
- State 8: The $timer(1,1)$ entry in agent B context information list was updated to $timeout(1)$. The periodic loop cannot be restarted because agent B still waits for availability of the called procedure.
- Edge $10 \rightarrow 11$: The execution of agent C $exec$ step finishes procedure $q1$. Because procedure $q2$ has been already called agent C starts it immediately.

V. SUMMARY

The formal description of time Alvis models and the set of transition rules for such models have been considered in the paper. The transition rules provide in fact an algorithm for SR-graphs generation that represent state spaces for such models. It should be stressed that an SR-graph is strictly dependent on the steps duration. For example, if we change the integers presented in Table I we will receive another SR-graph with possibly another paths. An SR-graph enables to check whether a given path (a sequence of steps) is possible to be executed for a given steps durations. We can also determine the minimal and maximal times of passing between two given states, i.e. we can, for example, determine the maximal time of reaction of our system to an event. Moreover, SR-graphs enable us to verify all classic properties like live-locks, deadlocks, process starvation etc. What is more important, the verification of these properties takes time dependencies under consideration. The future work will focus on implementation of algorithms for verification time requirements automatically.

REFERENCES

[1] R. Alur and D. Dill, "A theory of timed automata," *Theoretical Computer Science*, vol. 126, no. 2, pp. 183–235, 1994.

[2] C. Baier and J.-P. Katoen, *Principles of Model Checking*. The MIT Press, 2008.

[3] K. Balicki and M. Szpyrka, "Formal definition of XCCS modelling language," *Fundamenta Informaticae*, vol. 93, no. 1-3, pp. 1–15, 2009.

[4] J. Barnes, *Programming in Ada 2005*. Addison Wesley, 2006.

[5] A. Byrski and M. Kisiel-Dorohinicki, "Agent-based model and computing environment facilitating the development of distributed computational intelligence systems," in *Computational Science – ICCS 2009*, ser. LNCS, Springer-Verlag, 2009, vol. 5545, pp. 865–874.

[6] A. M. K. Cheng, *Real-time Systems. Scheduling, Analysis, and Verification*. Wiley Interscience, 2002.

[7] E. Emerson, "Model checking and the mu-calculus," in *DIMACS Series in Discrete Mathematics*. Amer. Math. Soc., 1997, pp. 185–214.

[8] H. Garavel, F. Lang, R. Mateescu, and W. Serwe, "CADP 2006: A toolbox for the construction and analysis of distributed processes," in *Computer Aided Verification*, ser. LNCS, vol. 4590. Springer-Verlag, 2007, pp. 158–163.

[9] S. Gnesi and T. Margaria, Eds., *Formal Methods for Industrial Critical Systems. A Survey of Applications*. Hoboken, John Wiley & Sons, 2013.

[10] K. Jensen and L. Kristensen, *Coloured Petri nets. Modelling and Validation of Concurrent Systems*. Springer-Verlag, 2009.

[11] K. Kluza, T. Maślanka, G. Nalepa, and A. Ligęza, "Proposal of representing BPMN diagrams with XTT2-based business rules," in *Intelligent Distributed Computing V – IDC 2011*, ser. Studies in Computational Intelligence, Springer-Verlag, 2011, vol. 382, pp. 243–248.

[12] L. Kotulski, M. Szpyrka, and A. Sędziwy, "Labelled transition system generation from Alvis language," in *Knowledge-Based and Intelligent Information and Engineering Systems – KES 2011*, ser. LNCS, Springer-Verlag, 2011, vol. 6881, pp. 180–189.

[13] R. Mateescu and M. Sighireanu, "Efficient on-the-fly model-checking for regular alternation-free μ-calculus," INRIA, Tech. Rep. 3899, 2000.

[14] R. Milner, *Communication and Concurrency*. Prentice-Hall, 1989.

[15] G. Nalepa, A. Ligęza, and K. Kaczor, "Formalization and modeling of rules using the XTT2 method," *International Journal on Artificial Intelligence Tools*, vol. 20, no. 6, pp. 1107–1125, 2011.

[16] B. O'Sullivan, J. Goerzen, and D. Stewart, *Real World Haskell*. O'Reilly Media, 2008.

[17] M. Szpyrka, "Analysis of VME-Bus communication protocol – RTCP-net approach," *Real-Time Systems*, vol. 35, no. 1, pp. 91–108, 2007.

[18] ——, "Exclusion rule-based systems – case study," in *International Multiconference on Computer Science and Information Technology*, vol. 3, Wisła, Poland, 2008, pp. 237–242.

[19] M. Szpyrka and L. Kotulski, "Snapshot reachability graphs for Alvis models," in *Knowledge-Based and Intelligent Information and Engineering Systems – KES 2011*, ser. LNAI, Springer-Verlag, 2011, vol. 6881, pp. 190–199.

[20] M. Szpyrka and P. Matyasik, "Formal modelling and verification of concurrent systems with XCCS," in *Proc. of the 7th Int. Symposium on Parallel and Distributed Computing (ISPDC 2008)*, Krakow, Poland, 2008, pp. 454–458.

[21] M. Szpyrka, P. Matyasik, and R. Mrówka, "Alvis – modelling language for concurrent systems," in *Intelligent Decision Systems in Large-Scale Distributed Environments*, ser. Studies in Computational Intelligence. Springer-Verlag, 2011, vol. 362, ch. 15, pp. 315–341.

[22] M. Szpyrka, P. Matyasik, R. Mrówka, and L. Kotulski, "Formal description of Alvis language with α^0 system layer," *Fundamenta Informaticae*, 2013, (to appear).

[23] M. Szpyrka, J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V – IDC 2011*, ser. Studies in Computational Intelligence, Springer-Verlag, 2011, vol. 382, pp. 249–255.

[24] M. Szpyrka and T. Szmuc, "Decision tables in Petri net models," in *Rough Sets and Intelligent Systems Paradigms*, ser. LNAI, Springer-Verlag, 2007, vol. 4585, pp. 648–657.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1559–1566

Relaxing Queries to Detect Variants of Design Patterns

Patrycja Węgrzynowicz, Krzysztof Stencel
Institute of Informatics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
Email: {patrycja, stencel}@mimuw.edu.pl

***Abstract*—Design patterns codify general solutions to frequently encountered design problems. They also facilitate writing robust and readable code. Their usage happens to be particularly profitable if the documentation of the resulting system is lost, inaccurate or out of date. In reverse engineering, detection of instances of design patterns is extremely helpful as it aids grasping high level design ideas. However, the actual instances of design patterns can diverge from their canonical textbook templates. Useful pattern detection tools should thus be able to identify not only orthodox implementations but also their disparate variants. In this paper, we present a method to generate queries to detect canonical instances of design patters. We formulate these queries so that they are language-agnostic. They precisely reflect the intents of the canonical implementations of design patterns. However, they abstract from any peculiarities of programming languages. Next, we show a systematic technique to relax these queries so that they also cover variant implementations of patterns. We discuss our proof-of-concept implementation of this approach in our prototype tool D-CUBED. Finally, we report the results of an experimental comparison of D-CUBED and state-of-the-art detectors.**

I. Introduction

A DESIGN pattern [1] is a general reusable solution to a commonly occurring problem in software design. Design patterns facilitate forming quality designs. As well as being useful in the construction of software systems (forward engineering), they also aid analysing existing systems (reverse engineering).

Detection of design patterns is an important part of reverse engineering. There are a significant number of large software systems without proper documentation that nevertheless need to be maintained, extended, or modified. In such cases, reverse engineering is necessary. However, the process is usually time-consuming and error-prone, as most of the core analysis must be performed manually and some important aspects can be omitted. Detection of design patterns automates extraction of high-level design concepts, which helps gaining a better understanding of code and makes analysis more efficient in terms of time and cost. Moreover, detection of design patterns can aid documenting code (e.g., generating or verifying documentation) or assessing its quality (e.g., using metrics based on design patterns).

In recent years, we have observed a continual improvement in the field of automatic detection of design patterns in source code. Existing approaches [2]–[13] can detect a fairly broad range of design patterns, targeting structural as well as behavioural aspects of patterns. Until recently, the research in the field of pattern detection has focused on novel approaches. However, the papers [14], [15] highlight the importance of the accuracy (precision and recall) of detection methods.

To achieve high recall, we need to reduce the number of false negative results. For a design pattern detection method, this minimization of false negatives means that the method should be capable of detection of numerous implementation variants that preserve the meaning of a design pattern, even though the details of their implementations do not follow the canonical implementation. Therefore, following the advice of [14] that emphasises the importance of '*a common set of patterns, both structural as well as behavioural, with well-defined implementation variants*', we focus on a systematic approach to detect variants of design pattern.

The analysis of implementation variants revealed highly diverse ecosystem of possibilities. For a simple design pattern like the Singleton, we have identified 7 elemental variants as presented in [16]. For another design pattern like the Visitor, we have also identified several significantly different implementations as presented in [17]. Considering further combinations of those elemental variants, it seemed infeasible to enumerate all available variants of design patterns, thus we sought an automated way to generate them. As a starting point we assumed the canonical implementation of design patterns as described in [1]. In [18], we introduced *pattern-preserving transformations* that enable transforming an implementation variant of a design pattern into a new one while preserving the design pattern.

In this paper, we extend our method of detecting design patterns (based on first-order logic formulae as described in [13]) to include a systematic approach to relaxing queries capable of detecting implementation variants of design patterns.

Contributions of this paper are as follows:

1) We present a systematic approach to the construction of queries to detect implementation variants of design patterns. The approach is based on the application of specific and generic pattern-preserving transformations of a Prolog query representing the canonical variant of a design pattern.
2) As a proof-of-concept for Contribution 1, we present queries capable of detecting variants of the Singleton

978-83-60810-53-8/$25.00

design pattern.
3) We evaluate our prototype tool (D-CUBED) with relaxed variant queries and compare it with two state-of-the-art pattern detectors (PINOT, DPD Tool).

II. METHOD

Our method of detecting variants of design patterns consists of the following steps (see Figure 1):

1) A transformation of the UML class diagram of the canonical implementation of a design pattern to a logic query;
2) An application of specific pattern-preserving query transformations;
3) An application of generic pattern-preserving query transformations;

The starting point of the method is the UML class diagram of *the canonical implementation* of a design pattern. We define *a canonical implementation* as the original implementation provided by the authors of a design pattern (e.g., Gang of Four in their book [1]). Through a set of transformations applied to the UML diagram of the canonical implementation of a design pattern, we obtain a disjunction query used to detect the variants of this pattern.

In [18], we introduced the concept of *pattern-preserving code transformations*, i.e., code transformations that preserve the intent of a design pattern. They were used to generate the implementation variants of design patterns in order to create test cases for pattern detectors. Here, we introduce a corresponding concept of *pattern-preserving query transformations*, which relax queries to extend their capabilities of detecting more implementation variants of design patterns.

In Step 1: Direct Transformation of UML to Logic, we transform the class diagram of the canonical implementation of a design pattern represented in UML to a formula using our custom metamodel (see Section II-A) expressed in logic. The output formula is a conjunction of predicates from our metamodel, describing such features as classes along with their fields and methods. Occasionally, we may need to add additional clauses to a conjunction query in order to represent important implementation details of a design pattern which are described in UML comments, other UML diagrams, or in the description of the pattern.

In Step 2: Specific Pattern-Preserving Query Transformations, we apply *specific pattern-preserving query transformations* to the query constructed in Step 1. *Specific pattern-preserving query transformations* are based on the concept of *specific pattern-preserving transformations* introduced in [18]. Each of them is a code transformations characteristic to a particular design pattern that preserves its intent. If such a transformation is applied to a correct implementation variant of a design pattern, it will produce a new correct variant. Exemplary specific pattern-preserving query transformations can be found in Section III.

In Step 3: Generic Pattern-Preserving Query Transformations, we apply *generic pattern-preserving query transformations* to the queries obtained in Step 2. Similarly to a specific pattern-preserving transformation, a *generic pattern-preserving query transformation* is based on the concept of a *generic pattern-preserving transformation* introduced in [18]. It is a code transformation referring to generic programming (e.g., abstractness, invocation, access modifiers) that preserves the design intent. The detailed description of generic pattern-preserving query transformations can be found in Section II-D.

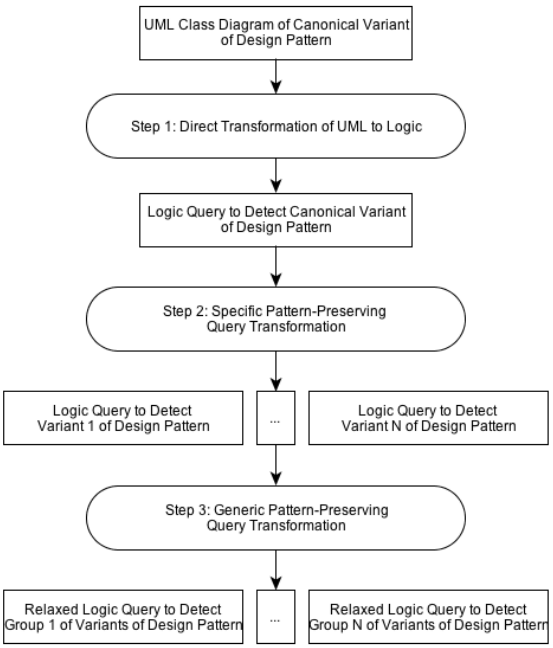


Fig. 1. The overview of the construction of the queries to detect the variants of a design pattern.

The final query to detect the variants of a given design pattern is the disjunction of the relaxed queries constructed in Step 3.

A. Program Metamodel

The program metamodel used in our detection method has been introduced in [13]. It consists of a set of core elements and a set of relationships among those elements, both structural and behavioural. The metamodel has been designed to be *"as simple as possible, but not simpler"*, yet it is powerful enough to model a large set of object-oriented languages.

The program metamodel consists of the following core elements (Figure 2): types, fields-or-variables, operations, and instances. Most of them have their obvious object-oriented meaning. A type denotes either a class, an interface or any other language-specific data type construct (like Java enum). A field-or-variable includes two cases: (1) an instance field or static field of a class or an interface, and (2) a variable that is not a field e.g. a global variable (irrelevant to Java but not to C++). Similar to the field-or-variable element, an operation also covers two cases: (1) a method declared in a type, and (2) a function e.g., a global function. An instance has been defined as an equivalence class of the relation "objects

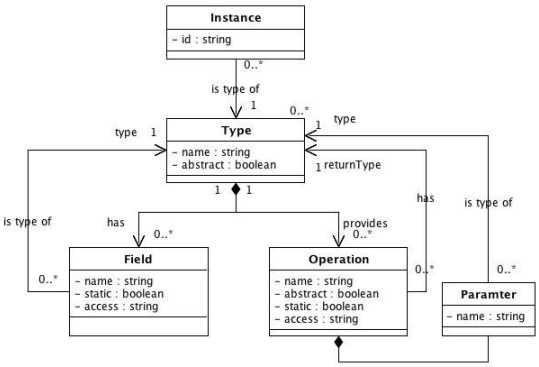


Fig. 2. The core elements of the program metamodel

constructed by the same `new`". All objects instantiated by the same `new` statement are treated as a single instance.

The extensional Prolog predicates describing core elements include: $isClass$, $isInterface$, $isEnum$, $isPrimitive$, $isType$, $isInstance$, $isField$, $isVariable$, $isFunction$, $isMethod$, $isParameter$, and others.

The elemental structural relations describe the relationships among core elements of a program, including memberships (i.e., fields and methods), modifiers (e.g., static, abstract, and access modifiers), and type system. The extensional Prolog predicates representing the elemental structural relations include: $isFieldOf$, $isMethodOf$, $hasParameter$, $hasReturnType$, $hasModifier$, $isTypeOf$, $isSubtypeOf$, and others. The intensional predicates $isTypeOf^*$ and $isSubtypeOf^*$ are the transitive closures of $isTypeOf$ and $isSubtypeOf$ respectively.

The elemental behavioural relations mostly refer to a data-flow and a call-flow, including instantiation as a specific call to a `new` operator. The extensional Prolog predicates representing the elemental behavioural relations include: $invokes$, $instantiates$, $hasInput$, $hasOutput$, and others. The intensional predicates $invokes^*$ and $instantiates^*$ are the transitive closures of $invokes$ and $instantiates$ respectively.

In order to illustrate the semantics of the predicates representing relations, here we present the definitions of $hasInput$ and $hasOutput$, two exemplary predicates related to the call-flow of a program:

$hasInput$

$hasInput(F, I)$, if and only if there is a potential execution path where the instance I is passed as one of the input parameters or as a part of an input parameter to the operation F.
$hasInput(F, T)$, if and only if $hasInput(F, I)$ and $isTypeOf^*(I, T)$.

$hasOutput$

$hasOutput(F, I)$, if and only if there is a potential execution path where the instance I is the output value or a part of the output value of the call to the operation F. The output means a return value as well as an output parameter.
$hasOutput(F, T)$, if and only if $hasOutput(F, I)$ and $isTypeOf^*(I, T)$.

B. UML to Logic Transformation

By a logic query we understand a set of Horn [19] clauses as used in logic programming. These logic queries operate on the program metamodel from Section II-A. We treat UML class diagrams of design patterns as inquires to a codebase. Therefore, we translate the codebase queries in the form of the UML class diagrams to the logic queries operating on the metamodel.

The algorithm to transform a class diagram to a logic query is as follows:

1) For each class and interface in a class diagram, we produce a conjunction of predicates, describing the type ($isClass$ or $isInterface$), its supertypes ($isSubtypeOf$), and its modifiers ($hasModifier$).
2) For each method in a class or interface, we produce a conjunction of predicates, describing its enclosing type ($isMethodOf$), its signature ($numberOfParameters$, $hasParameter$, $hasReturnType$), and its modifiers ($hasModifier$).
3) For each field in a class or interface, we produce a conjunction of predicates, describing its enclosing type ($isFieldOf$), its type ($isTypeOf$), and its modifiers ($hasModifier$).
4) We model additional information (e.g., comments) contained in a class diagram on per-case basis.
5) The output query is the conjunction of previously generated queries (i.e., class queries, method queries, field queries, additional information queries).

Optionally, we may need to transform to a logic query other UML diagrams (e.g., a sequence diagram) of a design pattern or encode features described purely textually.

C. Specific Query Transformations

A *specific pattern-preserving query transformation* is characteristic to a particular design pattern. It deals with, so-called, query logical fragments (i.e. parts of a query representing logical fragments of a pattern) transforming them into queries corresponding to different implementations valid in the context of a pattern.

A *pattern logical fragment* is a robust fragment of a design pattern which can be implemented using various programming techniques, e.g. the instantiation of a singleton instance (lazy or eager implementation). It is possible for a design pattern to have only one logical fragment (itself).

The identification of logical fragments of a design pattern as well as the transformations of these fragments are a purely manual process because they require understanding and abstraction of the semantics of a design pattern.

The application of the specific query transformations is an automated process, once we have query logical fragments and their transformations. Let us assume that we identified N logical fragments $F_i : 1 \leq i \leq N$ and for each fragment F_i we found a set of query transformations $T_{F_i} = T_j^i : 1 \leq j \leq K_i$. Then our algorithm to apply specific pattern-preserving query transformations is as follows:

1) We produce a Cartesian product of transformations over logical fragments, eliminating impossible combinations. As the result we obtain a set of possible transformation tuples $T = (t_1, ..., t_N) : t_i \in T_{F_i} \cup NONE$.
2) We apply the transformations tuples to the query obtained in Step 1.

D. Generic Query Transformations

Generic pattern-preserving query transformations are based on generic pattern-preserving code transformations. Figure 3 presents the generic pattern-preserving transformations as identified in [18]: (1) abstractness transformations, (2) invocation transformations, (3) inheritance transformations, (4) aggregation transformations, (5) method signature transformations, and (6) access transformations.

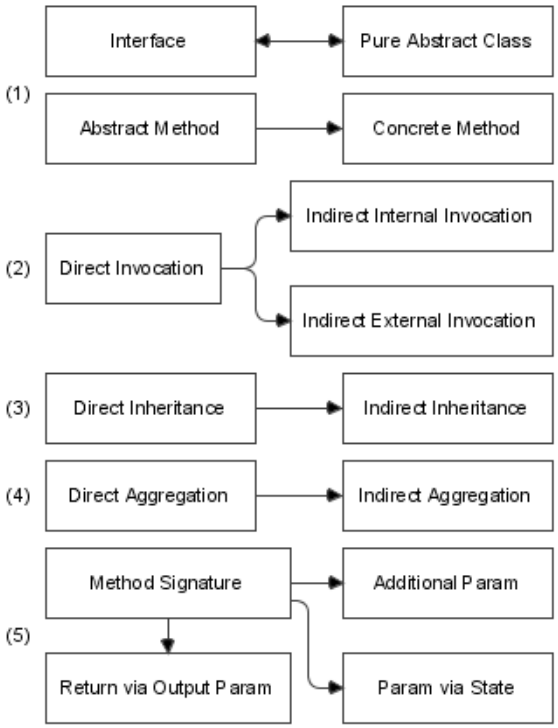


Fig. 3. The generic pattern-preserving code transformations.

a) Abstractness Transformations: These transformations refer to the property of being abstract for a class or a method. There is a two-sided transformation between interface and pure abstract class. This transformation is obvious since there exists a direct correspondence between these two constructs. They are often used interchangeably. The next transformation converts an abstract method to a concrete method. In our opinion this also is a natural transformation. It often happens that in the real world development programmers provide a default implementation instead of leaving a method abstract. By combining these transformations, we can transform an interface into a concrete class. Summing up, as long as we can take advantage of polymorphic calls, abstractness and concreteness do not impact the intent of a pattern code.

In terms of logic queries, the abstractness transformations lead to the following rewrite rules:

1) $isClass(Type) \rightarrow isClassOrInterface(Type)$
2) $isInterface(Type) \rightarrow isClassOrInterface(Type)$
3) $hasModifier(Class, 'abstract') \rightarrow true$
4) $hasModifier(Method, 'abstract') \rightarrow true$

b) Inheritance Transformations: The inheritance transformation introduces an intermediary level of inheritance, i.e. a direct subclassing is transformed into indirect. In real world software code such a construct can be the effect of a particular functional requirement or the complexity of a design problem. The length of an inheritance chain does not impact the intent of a pattern.

In terms of logic queries, the inheritance transformations lead to the following rewrite rule:

1) $isSubtypeOf(Type1, Type2) \rightarrow isSubtypeOf^*(Type1, Type2)$

c) Invocation Transformations: This group refers to transformations of invocation and instantiation statements. It is a popular refactoring. When a piece of code becomes complex, it is extracted into a separate method or class. This means that an invocation (or an instantiation), which remained direct until refactoring, is transformed into indirect. Depending on whether a new method or a new class is introduced, we call this indirect invocation internal or external respectively. Similarly to the length of an inheritance chain, the length of an invocation chain does not influence the logic of a pattern variant.

In terms of logic queries, the invocation transformations lead to the following rewrite rules:

1) $invokes(Method1, Method2) \wedge isMethodOf(Method1, Class) \wedge isMethodOf(Method2, Class) \rightarrow invokes^*(Type1, Type2) \wedge isMethodOf(Method1, Class)$
2) $invokes(Method1, Method2) \rightarrow invokes^*(Type1, Type2)$

d) Aggregation Transformations: These transformations follow the reasoning presented for invocations. Replacing a direct aggregation of an attribute (or a group of attributes) with an indirect aggregation does not influence the overall intent of a pattern implementation. Here is the example of such a transformation applied to the `observers` attribute:

```
class Observable {
    List<Observer> observers;
    ...
}
```

```
class Observable {
    ObserverList observerList;
    ...
}
class ObserverList {
    List<Observer> observers;
    ...
}
```

In terms of logic queries, the aggregation transformations lead to the following rewrite rule:

1) $isFieldOf(Field, Type) \rightarrow isFieldOfField^*(Field, Type)$

e) Method Signature Transformations: Here we present the transformations of method signatures. We have identified three such transformations:

- addition of a new (input) parameter,
- replacing the return value with an output parameter,
- passing an input value to a method via an object state instead of passing a parameter.

In terms of logic queries, the method signature transformations lead to the following rewrite rules:

1) $numberOfParameters(Method, N) \rightarrow true$
2) $isParameter(Param, Type) \wedge isParameterOf(Param, Method) \rightarrow hasInput(Method, Type)$
3) $hasReturnType(Method, Type) \rightarrow hasOutput(Method, Type)$

f) Access Transformations: These transformations relate to access modifiers of fields, methods, and classes. Basically, it means that access modifiers are not core elements of patterns and can be ignored during pattern detection. First of all, access modifiers are language-dependent feature. There are programming languages (e.g., JavaScript) that do not support access modifiers. Also, the semantics of the access modifiers may differ from one language to another (e.g., protected access in Java and C++). Therefore, we cannot rely on access modifiers and their semantics in a language-independent detection method. Moreover, there might be software requirements that limit the visibility of a field, method, or class (e.g., a Singleton class that is visible and accessible only in a package).

In terms of logic queries, the access transformations lead to the following rewrite rule:

1) $hasModifier(Element, Access) \rightarrow true$

III. Examples

This section describes the application of our method of query relaxation to the Singleton pattern.

The Singleton pattern is the most popular pattern detected by the existing detection approaches. Its canonical implementation is simple, and the intent seems straightforward. However, by carefully analyzing the structure of this pattern, we can identify some corner cases among its implementation variants as well as in the usage context. The standard variant of the Singleton is shown in Figure 4.

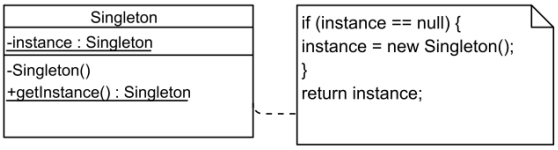


Fig. 4. The standard variant of the Singleton pattern.

A. Step 1: UML to Logic Query Transformation

Below we present the query that reflects the canonical implementation of the Singleton. As such, it can be used to detect orthodox implementations of this design pattern.

A singleton class:

$$isClass(S) \wedge hasModifier(S, 'public')$$

A singleton constructor:

$$isConstructor(SCtr) \wedge isConstructorOf(S) \wedge numberOfParameters(SCtr, 0) \wedge hasModifier(SCtr, 'private')$$

A singleton getter (access point):

$$isMethod(Get) \wedge isMethodOf(Get, S) \wedge numberOfParameters(Get, 0) \wedge hasReturnType(Get, S) \wedge hasModifier(Get, 'static') \wedge hasModifier(Get, 'public')$$

A singleton field:

$$isField(Field) \wedge isFieldOf(Field, S) \wedge isTypeOf(Field, S) \wedge hasModifier(Field, 'static') \wedge hasModifier(Field, 'private')$$

An approximation of a lazy instantiation block:

$$isCodeBlock(Code) \wedge isCodeBlockOf(Get) \wedge instantiatesOptionally(Code, Instance) \wedge isTypeOf(Instance, S) \wedge writesOptionally(Code, Field, Instance)$$

B. Step 2: Specific Query Transformations

In [18], we presented the logical fragments of the Singleton pattern along with the Singleton specific pattern-preserving code transformations. Here, we present the corresponding Singleton-preserving query transformations:

1) Instantiation: To Eager (A) – the lazy instantiation is replaced with an eager instantiation. The rewrite rule is as follows:
$$isCodeBlock(Code) \wedge isCodeBlockOf(Get) \wedge instantiatesOptionally(Code, Instance) \wedge isTypeOf(Instance, S) \wedge writesOptionally(Code, Field, Instance) \rightarrow isInitBlock(Code) \wedge instantiates(Code, Instance) \wedge isTypeOf(Instance, S) \wedge writes(Code, Field, Instance)$$
2) Placeholder: Inner Class (B) – the singleton instance is held as a static attribute of an inner class. The rewrite rule is as follows:
$$isFieldOf(Field, S) \rightarrow isFieldOf(Field, Inner) \wedge isClass(Inner) \wedge isMemberOf(Inner, S)$$

3) Placeholder: External Class (C) – the singleton instance is held as a static attribute of a class from within the same package. The rewrite rule is as follows:
$isFieldOf(Field, S) \rightarrow isFieldOf(Field, Outer) \wedge isClass(Outer) \wedge inPackage(S, P) \wedge inPackage(Outer, P)$
4) Access Point: Inner Class (D) – the public static method is moved to an inner class of a singleton. The rewrite rule is as follows:
$isMethodOf(Get, S) \rightarrow isMethodOf(Get, Inner) \wedge isClass(Inner) \wedge isMemberOf(Inner, S)$
5) Access Point: External Class (E) – the public static method is added to a newly created class in the same package. The rewrite rule is as follows:
$isMethodOf(Get, S) \rightarrow isMethodOf(Get, Outer) \wedge isClass(Outer) \wedge inPackage(S, P) \wedge inPackage(Outer, P)$
6) Access Point: Attribute (F) – the static singleton instance is made public (eagerly instantiated, with the access method removed). The rewrite rule is as follows:
$isMethod(Get) \wedge isMethodOf(Get, S) \wedge numberOfParameters(Get, 0) \wedge hasReturnType(Get, S) \wedge hasModifier(Get, 'static') \wedge hasModifier(Get, 'public') \rightarrow hasModifier(Field, 'public')$
7) Access Point: Protected (G) – the visibility of the access method is changed to protected. The rewrite rule is as follows:
$hasModifier(Get, 'public') \rightarrow hasModifier(Get, 'protected')$
8) Finality: Abstractness with Subclassing (H) – the Singleton class is made abstract and a concrete subclass is provided (the visibility of the Singleton constructor is changed to protected). The rewrite rule is as follows:
$isTypeOf(Instance, S) \rightarrow isTypeOf(Instance, CS) \wedge isClass(CS) \wedge isSubtypeOf(CS, S) \wedge hasModifier(S, 'abstract')$

C. Step 3: Generic Query Transformations

As in this paper it is infeasible to list all generated queries, we present the resulting query obtained after the application of the generic query transformations to the query being the result of the application of the specific query transformations $(A, NONE, NONE, NONE)$ (i.e., lazy instantiation changed to eager instantiation):

A singleton class:
$isClass(S)$
A singleton constructor:
$isConstructor(SCtr) \wedge isConstructorOf(S)$
A singleton getter (access point):
$isMethod(Get) \wedge isMethodOf(Get, S) \wedge hasOutput(Get, S) \wedge hasModifier(Get, 'static')$
A singleton field:
$isField(Field) \wedge isFieldOf(Field, S) \wedge isTypeOf(Field, S) \wedge hasModifier(Field, 'static')$
An approximation of an eager instantiation block:
$isInitBlock(Code) \wedge instantiates(Code, Instance) \wedge isTypeOf(Instance, S) \wedge writes(Code, Field, Instance)$

IV. Evaluation

We compared our prototype tool D-CUBED [20] with two state-of-the-art pattern detection tools: PINOT [21] and FUJABA 4.3.1 [2].

PINOT is a command-line tool written in C++ and based on jikes (the IBM Java compiler). PINOT is available as open source for custom compilation. PINOT ran smoothly, offering high performance in pattern detection tasks. The detection algorithms are hard-coded in PINOT, thus it is hard to experiment by modifying the detection approach. PINOT produces a useful, verbose report summarizing detected pattern instances.

FUJABA is a visually appealing graphic tool suite that provides pattern inference facilities as a plug-in (Inference Engine). There was no problem in launching FUJABA. It provides a UML-like language for user-defined patterns, and presents detected pattern instances as oval annotations on class diagrams. Even though this visual presentation helps in better understanding of diagrams, a summary report might be useful as well.

Similarly to PINOT, D-CUBED is a command line tool written in Java. However, contrary to PINOT, its detection queries are not hard-coded. Instead, we use XSB Prolog, a deductive database, as the data store for our program metamodel and Prolog as the query language. To generate a program metamodel from source code, we use Recoder (a front-end Java compiler) and a set of custom analyses to inspect in detail the call-flow and data-flow of a program.

We have tested these three tools against the source code of JHotDraw60b1 [22]. JHotDraw is a Java GUI framework for technical and structured graphics. It has originally been developed as a design exercise by Erich Gamma and Thomas Eggenschwiler.

Table I presents the results of the tests against JHotDraw. Unfortunately, FUJABA threw an exception during its static analysis. PINOT and D-CUBED performed their detection without any problems, however they produced significantly different results. PINOT did not report any Singleton instances, whereas D-CUBED recognized seven Singleton candidates:

1) *Clipboard* — a true positive; a singleton documented in source code.
2) *DisposableResourceManager* — a true positive; a singleton documented in source code; different placeholder;
3) *ResourceDisposabilityStrategy* — a true positive; analogous to *DisposableResourceManager*; different placeholder;
4) *CollectionsFactory* — a true positive; a singleton with subclassing and delegated construction;
5) *Alignment* — a true positive; a singleton with 6 instances;
6) *FigureEnumerator* — a false positive; there is a single static enumerator representing an empty enumerator,

though other instances are created as well;

7) *HandleEnumerator* — a false positive; there is a single static enumerator representing an empty enumerator, though other instances are created as well;

PINOT did not detect any singleton instance because its detection algorithm relies on the presence of the standard structure and a lazy instantiation block. *Clipboard* uses eager initialization (variant A), whereas two next singletons represent a different placeholder variant (variant C). *CollectionsFactory* represents a singleton with subclassing (variant H). As we did not impose an exactly one instance constraint, we also found the variant with several instances available (*Alignment*).

Unfortunately, query relaxation lead to two false positives. Therefore, the current method turned out to be too flexible. Our current goal, i.e. improving recall, decreased precision of the method. To achieve high precision, we need to filter out false positives. For a design pattern detection method, this minimization of false positives translates into understanding code constructs that violate the principles of a design patterns, even though the overall structure of a given piece of code resembles that of the design pattern. This is part of our undergoing research.

TABLE I
THE RESULTS OF SINGLETON DETECTION ON JHOTDRAW

	PINOT	FUJABA	D-CUBED
Singleton	0	×	7 (5 true positives)

× the tool raised an exception

V. RELATED WORK

There exists a number of proposed approaches to design pattern recognition, yet these approaches often lack in terms of accuracy, flexibility, or performance. A large number of approaches uses only structural information in order to detect design pattern instances (e.g. [23], [24]), but there also exist several approaches that exploit behavioural information contained in source code (e.g. FUJABA [2]–[4], Hedgehog [5], PINOT [6], [7]). We compare the existing detection methods to ours in terms of the concept of an approach, the architecture of a solution, and the mechanisms used to search for patterns.

A metamodel-based approach is not new. There exist several approaches that make use of a metamodel. Ptidej [10] is based on the PADL metamodel (the ancestor of PDL). However, a pattern in PADL is defined as a list of the required entities (a simple conjunction). Thus, it is hard to express more complex logic conditions. Moreover, it does not provide any support to data flow. SPQR [9] uses denotational semantics known as the ρ-calculus together with the set of elemental design patterns that capture call flow information. Similarly to PADL, there is no support to data flow. Hedgehog [5] utilises a Prolog-like language (Spine) to construct a pattern definition. Again, Spine's support to data flow is limited, but additional rules can be introduced. Also MAISA uses a metamodel. Its metamodel is UML and with its help it defines the structural patterns.

Current detection approaches utilise significantly different techniques to identify the instances of design patterns in source code. The most popular technique is the use of a logic inference system. This idea has been applied in Pat [25], where each structural pattern has been associated with a separated set of rules and Prolog interpreter has been used to search for patterns. Also [8] utilises a logic inference system to detect patterns in Java and Smalltalk based on a language-specific naming and coding conventions. SPQR and FUJABA also employ a logic inference engine to reason about the pattern instances.

Ptidej [10] uses a constraints solver with automatic constraint relaxation to detect sets of entities similar to a design pattern. A related work of Ptidej [26] utilises program metrics and a machine learning algorithm to fingerprint design motifs roles. One more approach that uses machine learning techniques is [27]. It enhances a pattern-matching system [11], [12] by filtering out false positives.

Numerous approaches simply navigate over a program abstract syntax tree to find the instances of patterns. They perform static or dynamic analyses to capture the behaviour of a program. Usually their detection algorithms are hard coded and tailored to a particular programming language. PINOT [6] performs static analysis to identify pattern-specific code-blocks. It is a lightweight solution performing recognition in an efficient manner. The approach described in [7] has two phases. In the first phase (the static analysis of the code structure) the abstract syntax tree is analysed in order to select the set of candidates to be pattern instances. In the second phase (the dynamic analysis of a program run) the messages passed are examined to check whether a candidate instance from the first phase is rejected or accepted.

Our approach and the prototype D-CUBED utilises structural information, but it extends the structure-driven approaches by targeting behaviour of the patterns. D-CUBED seems similar to the database driven approaches like DP++ [28] or SPOOL [29], but while these approaches address only the structural patterns, D-CUBED addresses the creational and behavioural patterns utilising the elemental relations to capture code intent.

A completely different approach is taken by the authors of DPJF [30]. They run a number of detection tools and the fuse their results. Such an approach may boost both precision and recall depending on the tuning of parameters. In this paper we focus on our specific method that can improve DPJF when their maintainers upload new version of D-CUBED.

VI. CONCLUSION

In this paper we proposed a general method to generate queries (logic programs) that detect disparate implementation variants of design patterns. First, we produce a strict query that reflect only the canonical textbook version of a design pattern. Then, we relax this query in order to allow multiple variants.

We implemented this approach in our prototype tool D-CUBED. We experimentally verified it with respect to state-of-

the-art detectors. The results are promising, since D-CUBED has detected several non-trivial variants of the Singleton that have not been revealed by other tools.

Apparently, our approach increases the recall of the detection process. However, higher recall may possibly imply lower precision. As the next step in our research, we plan to limit the number of false positives, i.e. detected instances that are not real incarnations of the design patterns. We will attempt tightening the detection queries by *adding* conditions that detect features that actually *violate* the intent of the given design pattern.

REFERENCES

[1] E. Gamma, R. Helm, R. E. Johnson, and J. M. Vlissides, *Design Patterns*. Addison-Wesley, 1994.

[2] J. Niere and L. Wendehals, "An interactive and scalable approach to design pattern recovery," Tech. Rep., 2003.

[3] J. Niere, W. Schäfer, J. P. Wadsack, L. Wendehals, and J. Welsh, "Towards pattern-based design recovery," in *ICSE '02: Proceedings of the 24th International Conference on Software Engineering*. New York, NY, USA: ACM, 2002, pp. 338–348.

[4] J. Niere, J. P. Wadsack, and L. Wendehals, "Handling large search space in pattern-based reverse engineering," in *IWPC '03: Proceedings of the 11th IEEE International Workshop on Program Comprehension*. Washington, DC, USA: IEEE Computer Society, 2003, p. 274.

[5] A. Blewitt, A. Bundy, and I. Stark, "Automatic verification of design patterns in java," in *ASE '05: Proceedings of the 20th IEEE/ACM international Conference on Automated software engineering*. New York, NY, USA: ACM, 2005, pp. 224–232.

[6] N. Shi and R. A. Olsson, "Reverse engineering of design patterns from java source code," in *ASE '06: Proceedings of the 21st IEEE/ACM International Conference on Automated Software Engineering*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 123–134.

[7] D. Heuzeroth, T. Holl, G. Högström, and W. Löwe, "Automatic design pattern detection," in *IWPC*. IEEE Computer Society, 2003, pp. 94–104.

[8] J. Fabry and T. Mens, "Language independent detection of object-oriented design patterns," *Computer Languages, Systems and Structures*, vol. 30, no. 1–2, pp. 21–33, 2004.

[9] J. Smith and D. Stotts, "Formalized design pattern detection and software achitecture analysis," Dept. of Computer Science, University of North Carolina, Tech. Rep. TR05-012, 2005.

[10] H. Albin-Amiot, P. Cointe, Y.-G. Guéhéneuc, and N. Jussien, "Instantiating and detecting design patterns: Putting bits and pieces together," in *ASE*. IEEE Computer Society, 2001, pp. 166–173.

[11] Z. Balanyi and R. Ferenc, "Mining design patterns from C++ source code," in *ICSM '03: Proceedings of the International Conference on Software Maintenance*. Washington, DC, USA: IEEE Computer Society, 2003, p. 305.

[12] R. Ferenc, J. Gustafsson, L. Müller, and J. Paakki, "Recognizing design patterns in C++ programs with integration of columbus and maisa," *Acta Cybern.*, vol. 15, no. 4, pp. 669–682, 2002.

[13] K. Stencel and P. Węgrzynowicz, "Detection of diverse design pattern variants," in *APSEC '08: Proceedings of the 2008 15th Asia-Pacific Software Engineering Conference*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 25–32.

[14] N. Pettersson, W. Löwe, and J. Nivre, "On evaluation of accuracy in pattern detection," in *First International Workshop on Design Pattern Detection for Reverse Engineering (DPD4RE'06)*, October 2006. [Online]. Available: http://cs.msi.vxu.se/papers/PLN2006a.pdf

[15] L. J. Fulop, R. Ferenc, and T. Gyimothy, "Towards a benchmark for evaluating design pattern miner tools," *Software Maintenance and Reengineering, European Conference on*, vol. 0, pp. 143–152, 2008.

[16] K. Stencel and P. Wegrzynowicz, "Implementation variants of the singleton design pattern," in *OTM Workshops*, ser. Lecture Notes in Computer Science, R. Meersman, Z. Tari, and P. Herrero, Eds., vol. 5333. Springer, 2008, pp. 396–406.

[17] K. Stencel and P. Węgrzynowicz, "Visitor pattern revisited for recognition," in *ADBIS (local proceedings)*, P. Atzeni, A. Caplinskas, and H. Jaakkola, Eds. Tampere University of Technology. Pori. Publication, 2008, pp. 154–166.

[18] P. Węgrzynowicz and K. Stencel, "Towards a comprehensive test suite for detectors of design patterns," in *ASE*. IEEE Computer Society, 2009, pp. 103–110.

[19] A. Horn, "On sentences which are true of direct unions of algebras," *J. Symb. Log.*, vol. 16, no. 1, pp. 14–21, 1951.

[20] P. Węgrzynowicz and K. Stencel, "The good, the bad, and the ugly: three ways to use a semantic code query system," in *OOPSLA Companion*, S. Arora and G. T. Leavens, Eds. ACM, 2009, pp. 821–822.

[21] N. Shi and R. A. Olsson, "Reverse engineering of design patterns from java source code," in *ASE*. IEEE Computer Society, 2006, pp. 123–134.

[22] E. Gamma and T. Eggenschwiler, "JHotDraw," http://www.jhotdraw.org/, 1996–2008.

[23] N. Tsantalis, A. Chatzigeorgiou, G. Stephanides, and S. T. Halkidis, "Design pattern detection using similarity scoring," *IEEE Trans. Software Eng.*, vol. 32, no. 11, pp. 896–909, 2006.

[24] K. Brown, "Design reverse-engineering and automated design pattern detection in Smalltalk," Master's thesis, University of Illinois at Urbana Campaign, 1997.

[25] L. Prechelt and C. Krämer, "Functionality versus practicality: Employing existing tools for recovering structural design patterns," *J. UCS*, vol. 4, no. 11, pp. 866–882, 1998.

[26] Y.-G. Gueheneuc, H. Sahraoui, and F. Zaidi, "Fingerprinting design patterns," in *WCRE '04: Proceedings of the 11th Working Conference on Reverse Engineering*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 172–181.

[27] R. Ferenc, A. Beszedes, L. Fulop, and J. Lele, "Design pattern mining enhanced by machine learning," in *ICSM '05: Proceedings of the 21st IEEE International Conference on Software Maintenance*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 295–304.

[28] J. Bansiya, "Automating design-pattern identification," *Dr. Dobbs Journal*, 1998.

[29] R. K. Keller, R. Schauer, S. Robitaille, and P. Pagé, "Pattern-based reverse-engineering of design components," in *ICSE '99: Proceedings of the 21st international conference on Software engineering*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1999, pp. 226–235. [Online]. Available: http://portal.acm.org/citation.cfm?id=302622

[30] A. Binun and G. Kniesel, "DPJF - design pattern detection with high accuracy," in *CSMR*, T. Mens, A. Cleve, and R. Ferenc, Eds. IEEE, 2012, pp. 245–254.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1567–1574

FAL: A Forensics Aware Language for Secure Logging

Shams Zawoad
zawoad@cis.uab.edu
University of Alabama at Birmingham

Marjan Mernik
marjan.mernik@uni-mb.si
University of Maribor

Ragib Hasan
ragib@cis.uab.edu
University of Alabama at Birmingham

***Abstract*—Trustworthy system logs and application logs are crucial for digital forensics. Researchers have proposed different security mechanisms to ensure the integrity and confidentiality of logs. However, applying current secure logging schemes on heterogeneous formats of logs is tedious. Here, we propose FAL, a domain-specific language (DSL) through which we can apply a secure logging mechanism on any format of logs. Using FAL, we can define log structure, which represents the format of logs and ensures the security properties of a chosen secure logging scheme. This log structure can be later used by FAL to serve two purposes: it can be used to store system logs securely, and it will help application developers for secure application logging by generating required source code.**

***Keywords*-DSL, Secure Logging, Audit Trail, Digital Forensics**

I. Introduction

IN RECENT years, digital crime case has increased tremendously. An annual report of the Federal Bureau of Investigation (FBI) states that the size of average digital forensic case is growing 35% per year in the United States. From 2003 to 2007, it increased from 83 GB to 277 GB [1]. Various logs, e.g., network log, process log, file access logs, audit trail of application play vital role in a successful digital forensics investigation. System and application logs record crucial events, such as, user activity, program execution status, system resource usage, network usage, and data changes through which some important attacks can be identified, e.g., network intrusion, malicious software, unauthorized access to software, and many more. Log is also important to ensure the auditability of a system and auditability is a vital issue to make a system compliant with the regulatory acts, e.g., Sarbanes-Oxley (SOX) [2] or The Health Insurance Portability and Accountability Act (HIPAA) [3]. Keeping system audit trails and reviewing them in a consistent manner is recommended by NIST as one of the good principles and practices for securing computer systems [4].

While the necessity of logs and application audit trail are indisputable, the trustworthiness of this evidence will remain questionable if we do not take proper measures to secure them. In many real-world applications, sensitive information is kept in log files on an untrusted machine. As logs are crucial for identifying an attacker, attackers often attack the logging system to hide the trace of their presence in the attack or to frame an honest user. Very often, experienced attackers first attack the logging system [5], [6]. Malicious insider users colluding with the attacker can also tamper with logs. Moreover, forensics investigators can also alter evidence before presenting to court. To protect logs from these possible attacks, we must need a secure logging mechanism. Researchers have already proposed several secure logging schemes [7]–[9], which are designed to defend such attacks.

However, ensuring the privacy and integrity of the logs is costly given that it requires special knowledge and skill of developers. To implement a secure logging scheme, we need to give complete access of the logs to application developers. Providing full access of sensitive logs to developers definitely increases the attack surface. They can violate the privacy, sell sensitive business or personal information, and most importantly can keep a back door for future attack. Adding secure application audit trail can also be burdensome for developers, and increases the application development cost. On the other hand, system admins, who have access to network logs, process logs may not have sufficient knowledge for developing a securing logging scheme.

In this paper, we propose a DSL [10] to assist system admins and application developers for maintaining system logs and application audit trail securely, which is crucial for digital forensics investigation. A DSL is designed for a particular domain and has great advantages over general-purpose language for that specific domain. DSLs provide higher productivity by its greater expressive power, the ease of use, easier verification and optimization [10]–[12]. Using our proposed DSL *FAL*, system admins can define log structure and parse a log file according to the structure. They can also define the security parameters to preserve the integrity and confidentiality of logs. To accomplish this, they only need their domain knowledge related with system logs. Using FAL, a software security analyst can define the required audit trail structure and can generate code for a generic purpose language (GPL), e.g., Java, C# to store the audit logs securely.

Contribution. The contribution of this work is two-fold:

- We propose the first domain-specific language FAL, which can be used to ensure the security of system logs, and application audit logs.
- We show all the DSL development processes, which can be served as a guideline for future DSL development.

978-83-60810-53-8/$25.00

II. Background and Motivation

In this section, we present the necessity of secure logging scheme, common approaches for secure logging, and how a DSL can help to mitigate some challenges of secure logging.

A. Secure Logging

As logs are crucial for digital forensics investigation, this is often become the target of attacker. There can be two types of attacks on logs:

- **Integrity**: Integrity of logs can be violated in three ways – an attacker can remove log information, can re-order the log entries, and can add fake logs. A malicious user can launch these attacks to hide the trace of illegal activities from forensics investigation, or to frame an honest user. Timing of an incident is crucial for forensics investigation. Hence, re-ordering the log entries can be important for an attacker, which can give him a chance to produce some alibi.
- **Confidentiality**: From various system logs and application logs, we can identify the activity of users as well as sensitive private information about the users. From the application logs of a business organization, we can also trace out very sensitive business information. This information has high value to attacker. Hence, attack on the confidentiality of logs can be highly beneficial to attacker.

The above attacks can come from different types of attackers:

- **External Attackers**: An external attacker can be a malicious user intending to attack users' privacy from the logs, or try to modify logs to hide the trace of any attack (e.g., network intrusion, malware, spyware). A dishonest forensic investigator can also be an external attacker, as the investigator can alter the logs before presenting to court.
- **Internal Attackers**: A more crucial attack can come from insider attackers colluding with malicious users. A dishonest insider can be a system admin, database admin, or application developer. As system admins have access to all system logs, they can always tamper with logs. Application logs and some of the system logs can be stored in database. In this case, threats can come from database admin. A malicious database admin can modify logs without leaving any trace of the modification. Application developers can modify application logs, or can create a backdoor to collect the application logs. Besides tampering the logs, these insiders can also attack on the privacy of users. They can collect and sell sensitive business and personal information derived from the logs.

To defend the confidentiality and integrity of logs, researchers have proposed several secure logging schemes [7]–[9], [13]. The commonalities among these secure logging schemes are: encrypting sensitive fields to protect the confidentiality, and maintain a hash-chain of the logs to protect the integrity of logs. Hash-chain maintains the chronological information of data. Hence, if any log is missing from the chain or if there is a reordering of the logs then this alteration can be detected from the hash-chain. Hash-chain of one log entry is calculated using the hash of its previous entry. In this way, it preserves the sequence information.

B. Motivation

Though there are some proven secure logging schemes, developing and maintaining a scheme is always challenging because of the following reasons:

1) The first problem is logs are in heterogeneous format. Unfortunately, there is no standard of logs format. Hence, two types of systems logs can look completely different. Moreover, same log can vary by operating systems. For example, format of a process log entry is different in MacOS and Debian.
2) To build a secure logging scheme, we need to permit the logging scheme developers to access the logs. Developers' accessibility to crucial log information certainly increases the attack surface. Earlier, we only need to trust system admins; adding developers in the loop adds an extra level of trust. The developer might place a back door to collect plain log information and can violate the privacy of users.
3) For application logging, application developers need to add secure application logging code for every scenario. Most of the cases, we need to log the database operations – Add, Update, Delete. Through these logs, we can get who has done some specific operations on a specific data. Writing code for all of the possible scenarios is burdensome for developers, and skipping one important logging method may turn out to be crucial.

To resolve the above challenges, we suggest that a well-defined DSL should help. For system logs, with the help of a DSL, we can shift the responsibility of developing a secure logging scheme from programmers to system admins. Because systems admins already have the domain knowledge about system logs, and with the help of a DSL, they can easily define the required security parameters. In this way, we can minimize one level of attack surface. The DSL should also deal with the heterogeneous formats of logs. Hence, we do not need to re-implement a scheme when the log format changes because of any system migration. For application logs, a DSL can generate required application logging code to ease the life of application developers. However, using proprietary encryption and hashing algorithm cannot be adopted by a DSL. Hence, our proposed DSL can only handle established encryption and hashing algorithms.

III. The domain-specific-language FAL

A. Domain Analysis

The very first step of designing a DSL is the detailed analysis and structuring of the application domain [14], which is provided by domain analysis. Output of domain analysis is a Domain Model, which gives us commonalities and variabilities, semantics of concepts, and dependencies between properties. Among various schemes of domain analysis, we choose FODA

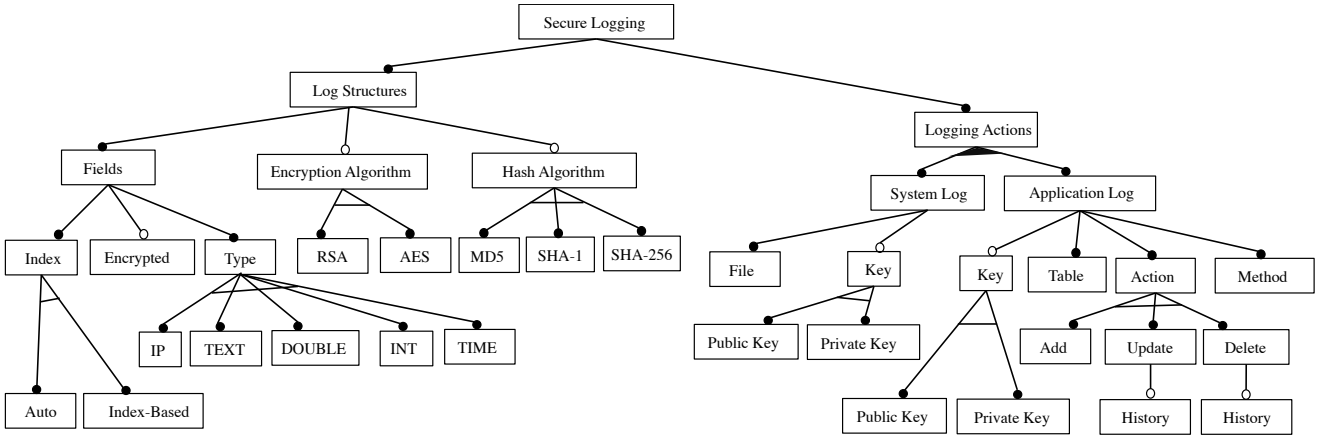


Fig. 1: The Feature Diagram of FAL

(Feature Oriented Domain Analysis). In FODA, the results of the domain analysis are obtained in a feature model [15]. One of the most prominent ways of describing feature model is by feature diagram (FD). The FD is represented as a tree with nodes as rectangles and arcs connecting the nodes. Nodes determine the features, while arcs determine the dependency between the features. The nodes can be mandatory or optional, which are denoted by closed dots, or open dots respectively. The FD of FAL is illustrated in Figure 1.

From Figure 1, it is clear that a secure logging scheme constitutes of log structure and logging action. Every log structure must have fields. Every field must have a type. According to the chosen secure logging scheme, a field can be encrypted or not. Fields may have an index attribute, which will be used to specify the location of a field in an input. The type of a field can be IP, Text, Double, Integer, or Time. Time can be auto-generated, i.e. current system time, or can be index-based. For index-based field, value will be extracted from input file or argument list according to the position defined by the index. For encryption, encryption algorithms, such as, RSA [16], AES [17] can be used. Some secure logging mechanisms use hashing, and hash-chain to ensure the integrity of the logs. Hence hashing algorithms, e.g., SHA-1[1], SHA-256[1], or MD5[2] can be used.

After defining a secure log structure, we need to use the structure for system or application logging. There can be two types of actions. First, for system logs, we need to parse the log files according to a pre-defined structure, and apply the security features while storing. Second, for application log, we need to generate GPL code. For system logs, we must have a file name, and we may have public or private key file. By encrypting with public key, we can ensure that only the private key owner can decrypt certain information. Private key is also needed to create a signature on certain data. Using the public key we can verify the signature. For application logging, we must have a table name, action, method, and may have public or private key file. Method is actually a method name of a GPL program, from where the action is called. An action can be adding a new record, update, or delete a record. For update and delete, we may want to save the history of previous records.

The FDs represent the common features, which always exist in a system (commonalities) and optional features, which may or may not exist in a system (variabilities). Some of the commonalities identified from the FD of FAL are *Fields*, *Type*, etc., and some variabilities are *Encryption Algorithm*, *Key*, etc. From FD, the variation points can be easily identified (optional, one-of and more-of features). After the domain analysis, we can gather the following information – terminology, concepts, and common and variable properties of concepts and their interdependencies.

B. The Abstract Syntax

After the domain analysis, the next step is to design the DSL, from which we will get syntax and semantics of the language. During the domain analysis using FODA, we identified several concepts in the application domain that needed to be mapped into DSL syntax and semantics. From the FD, we can notice the relationship between concepts/features in an application domain and non-terminals in a context-free grammar (CFG). Table I represents the mapping between application domain concepts and non-terminals in context-free grammars, which appears on the left hand side (LHS) and right-hand side (RHS) of CFG production.

Based on Table I, we define the abstract syntax of FAL, which is presented in Table II. The syntactic domains of variables are presented in Table III. A FAL program consists of Log structures LS, and logging actions LA. Log structure LS defines field description F and security parameter S. There can be one or more LS. The field description F specifies field type, id, index I, and encrypted status. There can be one or more fields in a log structure. Index I is either an integer number, or auto. Security parameter S defines encryption and hashing algorithm. Logging action LA can be either System logging action SLA or Application logging action ALA. There can be one or more logging actions. SLA specifies the system log file

[1]http://www.itl.nist.gov/fipspubs/fip180-1.htm
[2]http://tools.ietf.org/html/rfc1321

TABLE I: Translation of the application domain concepts to a context-free grammar

Application domain concepts	LHS non-terminal	RHS structure
Secure Logging	P	Description of Log structure, and logging action
Log Structure	LS	Description of fields and security parameters
Fields	F	Field id, type (IP, Text, Double, Integer, Time), indexing feature, encrypted (or not encrypted)
Index	I	Position of a field in input, or auto.
Security Parameters	S	Description of encryption and hashing algorithm
Logging Action	LA	Description of system logging, or application logging statement.
System logging	SLA	File name to be parsed to store securely, and encryption key.
Application log	ALA	Database operation, table id, GPL method name, encryption key, and history preservation option.

name and encryption key. ALA specifies the database action name, database table name, GPL method name, encryption key, and history preservation option.

TABLE II: Abstract syntax of FAL

P ::= LS LA
LS ::= lid F S |LS1; LS2
F ::= type fid I encrypted |type fid I |F1; F2
S ::= encAlg hashAlg |encAlg |hashAlg |ϵ
I ::= n |Auto
LA ::= SLA |ALA |LA1; LA2
SLA ::= file key |file
ALA ::= action tid key m withhistory |action tid m key |
action tid m history |action tid m

TABLE III: Syntactic Domains

P ∈ Pgm	LS ∈ LogStructure
F ∈ Field	LA ∈ LogAction
I ∈ Index	S ∈ SecAttrs
SLA ∈ SystemLog	ALA ∈ AppLog
n ∈ Num	file ∈ FileSpec
type ∈ {IP, Text, Double, Integer, Time}	fid ∈ FileIdentifier
tid ∈ TableIdentifier	m ∈ MethodName
action ∈ {Add,Update,Delete}	key ∈ KeyFileSpec
lid ∈ LogStructureIdentifier	encAlg ∈ {RSA,AES}
hashAlg ∈ {MD5, SHA-1,SHA-256}	

C. The Concrete Syntax

After defining the abstract syntax, we experimented with various forms of concrete syntaxes to see how various constructs might look. For example, a log structure with two field *fromip* and *user* can be defined using the concrete syntax as described in Listing 1.

Listing 1: FAL Log Structure

```
1: Define netlog {
2:     IP fromip Index 0 Encrypted;
3:     TEXT user Index 1;
4:     Use Encryption With RSA;
5:     Use Logchain With SHA_1;
6: };
```

Here, *fromip* field has data type IP, and *user* is of TEXT data type. The *Index* attribute represents the position of a field in the network log file. The *Encrypted* attribute states that the field will be encrypted according to the encryption algorithm defined in line 4. If there are multiple encrypted fields, all the fields will be encrypted using the same encryption algorithm. Line 5 adds the flexibility of choosing any hash function.

After defining a log structure a log action will be defined, which uses the pre-defined log structure. A concrete example of storing a network log file securely can be defined as follows (Listing 2):

Listing 2: FAL Logging Action

```
1: Watchfile network.log Using netlog
2: {
3:     Privatekey private.key;
4: }
```

The *Watchfile* statement uses the predefined 'netlog' structure to parse the 'network.log' file and provides the required encryption key to start the process of preserving logs securely.

Listing 3: FAL Program for System and Application Log

```
 1: SampleProgram[
 2:     Define netlog {
 3:         IP fromip Index 0 Encrypted;
 4:         TEXT user Index 1;
 5:         Use Encryption With RSA;
 6:         Use Logchain With SHA_1;
 7:     }
 8:     Define patientlog{
 9:         TIME logtime Auto;
10:         TEXT user Index 0 Encrypted;
11:         INT refid Index 1;
12:         TEXT message Index 2 Encrypted;
13:         Use Logchain With SHA_256;
14:     }
15:     Watchfile network.log Using netlog {
16:         Privatekey private.key;
17:     }
18:     Watchtable Patient Using patientlog {
19:         Action Edit Withhistory;
20:         Method updatepatient;
21:         Publickey public.key;
22:     }
23: ]
```

When a language designer is satisfied with the look and feel of the language's syntax, and possible additional constraints from domain experts or language end-users are fulfilled, the concrete syntax can be finalized. In Listing 3, a complete example of FAL program for secured system and application logs is described. We finalized the concrete syntax on the basis of several example programs. Finalizing the concrete syntax process can be executed in parallel with defining language semantics. In Table IV, the FAL concrete syntax is given.

TABLE IV: The concrete syntax of FAL

```
Program := #CCStart [LOG_STRUCT LOG_ACTION]
LOG_STRUCTS := LG_STRUCTS
LG_STRUCTS := LG_STRUCTS LG_STRUCT |LG_STRUCT
LG_STRUCT:= Define #Id {DEF}
DEF := FIELDS SEC_ATTRS
FIELDS := FIELDS FIELD |FIELD
FIELD := #Type #Id IND_BASE ENC ;
IND_BASE := Index #Number |Auto
ENC := Encrypted |ϵ
SEC_ATTRS := SEC_ATTRS SEC_ATTR |ϵ
SEC_ATTR := Use SEC_STMT ;
SEC_STMT := ENC_STMT |HASH_STMT
ENC_STMT := Encryption With #EncAlgorithm
HASH_STMT := Logchain With #HashAlgorithm
LOG_ACTION := LG_ACTIONS
LG_ACTIONS := LG_ACTIONS LG_ACTION |LG_ACTION
LG_ACTION := SYS_ACT |APP_ACT
SYS_ACT := Watchfile #FileName Using #Id {ENC_KEY}
ENC_KEY := PUB_KEY |PRIV_KEY |ϵ
PUB_KEY := Publickey #FileName;
PRIV_KEY := Privatekey #FileName;
APP_ACT := Watchtable #CCStart Using #Id {PARAM}
PARAM := DB_ACTION GPL_MTHD ENC_KEY
DB_ACTION := Action ACT_NAME ;
ACT_NAME := Add |ACT_HSTRY
ACT_HSTRY := ACT_HSTRY_NAME HISTRY_STMT
ACT_HSTRY_NAME := Edit |Delete
HISTRY_STMT := Withhistory |ϵ
GPL_MTHD := Method #Id ;
```

D. Translational Semantics

The advantages of using formal description for semantics of DSL (e.g., attribute grammars, denotational semantics, operational semantics) have been previously discussed in [10], where an ability to find problems in semantics before a DSL is actually implemented was exposed. In this work, we used translational semantics, which is simpler to define then denotational and operational semantics, and it is often used for defining semantics of domain-specific modeling languages [18]. Due to space considerations, only the translational semantics for log structures is presented in Listing 4 (translational semantics for logging actions is omitted from this paper). For each non-terminal in CFG, (Table II) a translational function is defined, which maps syntactic domains (Table III) to their meanings – generated code in Java using a specialized API for secure logging. In particular, the meaning of non-terminal LS is defined by translational function TLS, which maps $LogStructure$ to $code$. Two different forms of LS exist (see abstract syntax in Table II). Hence, two translational functions TLS are defined (lines 4 and 5 in Listing 4). The first translational function TLS (line 4 in Listing 4) maps syntactic structure $lid\ F\ S$ into several Java statements: declaration of new object as an instance of class $LogStructure$, setting a name to the newly created object by calling $setName$ method, and additional Java statements, which will be generated by applying translational functions TF and TS on non-terminals F and S representing fields and security attributes, respectively. Whilst, the second translational function TLS (line 5 in Listing 4) define the meaning of sequence of log structures ($LS1; LS2$). The generated code for $LS1$ is simply concatenated with generated code for $LS2$ (line 5 in Listing 4). In similar manner, other translational functions are defined.

Listing 4: Translational Semantics

```
1:  TP : Pgm → Code
2:  TP⟦LS LA⟧ = TLS⟦LS⟧ + TLA⟦LA⟧
3:  TLS : LogStructure → Code
4:  TLS⟦lid F S⟧ = "LogStructure " + lid + " = new LogStructure();" +
    lid+".setName(" + lid + ");" + TF⟦F⟧ lid + TS⟦S⟧ lid
5:  TLS⟦LS1; LS2⟧ = TLS⟦LS1⟧ + TLS⟦LS2⟧
6:  TF : Field → lid → Code
7:  TF⟦type fid I encrypted⟧ = lid+".addField(FieldType." + type + "," +
    fid + "," + TI⟦I⟧ + ", true);"
8:  TF⟦type fid I⟧ = lid+".addField(FieldType." + type + "," + fid + "," +
    TI⟦I⟧ + ", false);"
9:  TF⟦F1; F2⟧ = TF⟦F1⟧ + TF⟦F2⟧
10: TI : Index → Code
11: TI⟦n⟧ = "true, " + n
12: TI⟦Auto⟧ = "false, INTEGER.MAX_VALUE"
13: TS : SecAttrs → lid → Code
14: TS ⟦encAlg hashAlg⟧ = lid + ".setEncryptionAlgorithm("+ encAlg
15:     + ");" + lid + ".setHashingAlgorithm("+ hashAlg + ");"
16: TS ⟦encAlg⟧ = lid + ".setEncryptionAlgorithm("+ encAlg +");"
17: TS ⟦hashAlg⟧ = lid + ".setHashingAlgorithm("+ hashAlg + ");"
18: TLA : LogAction → Code
19: ...
```

E. Implementation

Various implementation techniques to implement a DSL exist, such as preprocessing, embedding, compiler/interpreter, compiler generator, extensible compiler/interpreter, commercial off-the-shelf, and hybrid approaches [10]. Kosar et al. [19] suggested focusing end-user usability while implementing a DSL. One implementation approach can be good in terms of effort needed to implement a DSL. However, the same approach may not be suitable for end-users. End-users may need extra effort to rapidly write correct programs using the DSL. If only DSL implementation effort is taken into consideration, then the most efficient implementation technique is embedding. However, the embedding approach might have significant penalties when end-user effort is taken into account (e.g., DSL program size, closeness to original notation, debugging and error reporting). To minimize end-user effort, building a DSL compiler [19] is most often a good solution, but this process costs most from an implementation point of view. However, the implementation effort can be greatly reduced, but not as much as with embedding, especially if compiler generators (e.g., LISA [20], ANTLR [21], Silver [22]) are used.

To implement FAL, we depend on source-to-source transformation technique. To transform a FAL program into an intermediate Java program, we build a FAL compiler using LISA, which has proven itself useful in many other DSL projects [23], [24]. The intermediate program uses a pre-build Java API. The design of the API is illustrated in Figure 2.

Fields are represented by *Field* class. The *LogStructue* has a list of Field object, and the security attributes. The name field of *LogStructure* is used to map with the database table name. *LogAction* is an Abstract class with the abstract method *execute*, and it also has an instance of *LogStructure*. *FileWatcher* extends the *LogAction* class and implements the execute method. The execute method is responsible to parse a log file and store it to database with the help of LogStructure and Field. *TableWatcher*

also extends the *LogAction* class and implements the execute method, which generates application logging code for developer. The *SecurityUtil* class defines all the required encryption and hashing methods.

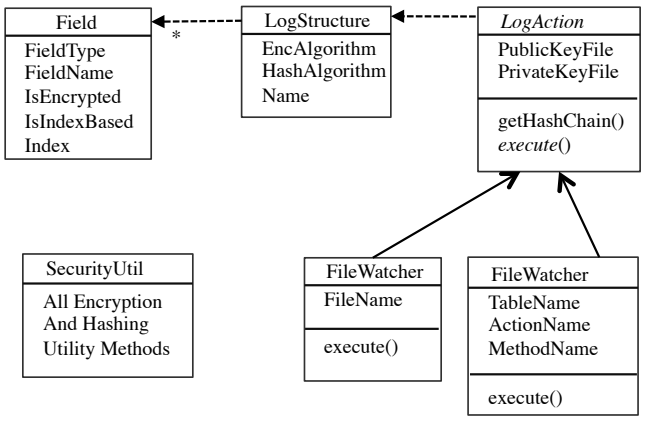


Fig. 2: Design of the API for FAL

After finalizing the Java API, we now know what the intermediate program will be. The FAL compiler will generate this intermediate program from a FAL program. To transform the FAL program to Java program correctly, we use the attribute grammar based approach and LISA specifications are based on attribute grammars [25], [26]. It is capable to generate the compiler from formal attribute grammar-based language specifications.

The first task to implement the compiler is to define the lexicon. Defining the lexicon in Lisa is straightforward. It is showed in Listing 5.

Listing 5: Lexical specification for FAL in LISA.

```
1:  lexicon {
2:      Number [0-9]+
3:      Id [a-z][a-z0-9_]*
4:      Type IP |TEXT |INT |TIME |DOUBLE
5:      EncAlgorithm RSA |AES
6:      HashAlgorithm MD5 |SHA_1 |SHA_256
7:      keywords Define |Use |Encryption |With |Logchain |Index |
8:          Auto |Encrypted |Watchfile |Using |Publickey |Privatekey |
9:          Watchtable |Action |Withhistory |Method |Parameter
10:     FileName [a-z][a-z0-9_]*.[a-z]*
11:     CCStart [A-Z][a-z0-9_]*
12:     ActionName Add |Edit |Delete
13:     Separator \; |\{ |\} |\, |\[ |\]
14:     ignore [ \0x09\0x0A\0x0D\]+
15: }
```

To write the attribute-based semantic rules, first, we need to identify the required attributes for proper semantic analysis. Listing 6 presents the attributes that we used. *code* is the main synthesized attribute that produces the targeted GPL program. *ivar* is an inherited attribute that is used to propagate the variable name down the parse tree. *envs* is a synthesized attribute and *envi* is an inherited attribute; both were needed to maintain a HashSet of already defined variables. *errorMsg* is a synthesized attribute required to report FAL error message to users. *ok* is a synthesized attribute that indicates whether a FAL program is correct or not. Finally, *PROGRAM.file* attribute is used to write the generated GPL program in a file.

Listing 6: Attributes for FAL in LISA.

```
1:  attributes String *.code;
2:      String *.ivar;
3:      String *.errorMsg;
4:      HashSet *.envs;
5:      HashSet *.envi;
6:      boolean *.ok;
7:      BufferedWriter PROGRAM.file;
```

An implementation of translational semantics (Listing 4) using LISA is a straightforward task. The implementation of translational function TF (Line 7 in Listing 4) is presented in Listing 7. Note, how closed both notations are.

Listing 7: Semantic Rules in LISA.

```
1:  rule field {
2:      FIELD ::= #Type #Id IND_BASE ENC \; compute {
3:          FIELD.code = FIELD.ivar + ".addField( FieldType." +
4:              #Type.value() + ",\"" + #Id.value()+"\", " +
5:              IND_BASE.code + "," + ENC.code+");";
6:      };
7:  }
8:  rule ind_base {
9:      IND_BASE ::= Index #Number compute {
10:         IND_BASE.code = "true,"+ #Number.value();
11:     }
12:     |Auto compute {
13:         IND_BASE.code = "false,Integer.MAX_VALUE";
14:     };
15: }
16: rule enc {
17:     ENC ::= Encrypted compute {
18:         ENC.code = "true";
19:     }
20:     |epsilon compute {
21:         ENC.code = "false";
22:     };
23: }
```

After compiling a FAL program, a required Java code is automatically generated, which uses previously defined APIs to store logs, and generate audit trail code for ensuring the integrity and confidentiality of the logs.

IV. Practical Experience

The goal of this section is to acquaint the reader with the practical experiences that were obtained by using FAL. We have therefore selected two case studies of FAL applications:

- Preserve snort log securely using FAL.
- Generate application logging code for a patient information update method in Java.

A. Preserve Snort log

Snort[3] is a free lightweight network intrusion detection system. The network logs generated by Snort plays vital role in network forensics. Hence, preserving the confidentiality and integrity of Snort logs is crucial from digital forensics perspective. Here is a sample Snort log:

```
11/19-13:43:43.222391 11.1.0.5:51215 ->
74.125.130.106:80  TCP TTL:64 TOS:0x0 ID:22101
```

[3]http://www.snort.org

```
IpLen:20 DgmLen:40 DF  ***A***F Seq: 0x3EA405D9
Ack: 0x89DE7D Win: 0x7210 TcpLen: 20''
```

This log tells that the machine with IP 11.1.0.5 performed an http request to machine 74.125.130.160 at time 11/19-13:43:43.222391. Hence, when a machine attacks another machine, we can identify the attacker machine IP from the snort log. Let's assume that a system admin decides to store the 'from IP', 'to IP', and time of network request securely. To protect the confidentiality of logs, among these three fields, the admin decides to encrypt 'from IP', and 'to IP' by the public key of law enforcement agencies using RSA algorithm. To protect the integrity of the logs, the system maintains hash-chain of the logs using SHA-256 hash function. The FAL program described in Listing 8 can be used to ensure all the properties.

Listing 8: FAL Program for Snort Log

```
SnortParser[
    Define snortlog {
        IP fromip Index 1 Encrypted;
        IP toip Index 3 Encrypted;
        Time logtime Index 0;
        Use Encryption With RSA;
        Use Logchain With SHA_256;
    };
    Watchfile snortnetwork.log Using snortlog {
        Publickey lawpublic.key;
    }
]
```

The above FAL program will generate Java code as follows (Listing 9):

Listing 9: Translated Java Code from FAL

```
LogStructure snortlog = new LogStructure();
snortlog.setName("snortlog");
snortlog.addField(FieldType.IP,"fromip",true,1,true);
snortlog.addField(FieldType.IP,"toip",true,2,true);
snortlog.addField(FieldType.TIME,"logtime",true,0,false);
snortlog.setEncryptionAlgorithm("RSA");
snortlog.setHashingAlgorithm("SHA_256");
FileWatcher snortlogFileWatcher = new FileWatcher();
snortlogFileWatcher.setLogStructure(snortlog);
snortlogFileWatcher.setFileName("snortnetwork.log");
snortlogFileWatcher.setPubicKeyFile("public.key");
snortlogFileWatcher.execute();
```

Executing the Java code (Listing 9) will parse the snort log file and store them with the security parameter. However, FAL users do not need to understand the underlying API or the intermediate Java code generated by FAL.

B. Application Logging

Application log is crucial for many applications including business and health care sector. The methods that directly communicate with database need to be logged. From these logs, later we can identify who added a new record, or who updated or deleted some record, etc. Application developer needs to integrate this logging feature with every method that updates database. FAL can generate the necessary logging code for application developer.

Here, we present a hypothetical scenario of a health care application, where we can use FAL for secure application logging. In the application, there is a Patient table, and we want to store logs whenever any update is operated on patient's record. The log will include the user name of the application, patient id is being updated, a description of the operation, and time of operation. The security analyst of the application decides to encrypt user name, and the operation description using AES encryption algorithm and SHA-1 hash function to maintain the hash-chain of logs. The FAL program described in Listing 10 can be used to generate necessary application logging code.

Listing 10: FAL Program for Application Logging

```
PatientAppLog [
    Define useraudit {
        TIME logtime Auto;
        TEXT username Index 0 Encrypted;
        INT refid Index 1;
        TEXT message Index 2 Encrypted;
        Use Encryption With AES;
        Use Logchain With SHA_1;
    };
    Watchtable Patient Using useraudit {
        Action Edit Withhistory;
        Method updatepatient;
        Privatekey serveraes.key;
    }
]
```

The translated Java code from FAL program (Listing 10) will generate the application logging method as described in Listing 11.

Listing 11: Generated Code For Application Logging

```
public void auditPatientEdit(String username, int refid, String message,
String logtime)
{
    try {
        String rowValue = username + refid + message + logtime;
        String currHashs = getHashChain("useraudit","id",rowValue,
            "SHA-1");
        String aesKey = HandleKey.readAESKey("serveraes.key");
        username = HandleKey.aesEncrypt(username +"", aesKey);
        message = HandleKey.aesEncrypt(message +"", aesKey);
        String query = "insert into useraudit( username, refid,
            message, logtime, tablename, actionname, methodname,
            logchain, withhistory) values("' + username
            + "'," + refid + ",'" + message + "','" + logtime +
            "','Patient','Edit','updatepatient','"+currHashs+"',true)";
        DBHandler dbHandler = new DBHandler();
        dbHandler.insertData(query);
    }catch (Exception e) { e.printStackTrace();}
}
```

V. Related Work

There has been a lot of prior research on secure logging, because of its vital role in digital forensics. Schneier et al. proposed a secure audit logging scheme, which can detect any modification of logs after a machine got compromised [27]. It also preserves the confidentiality of logs. Zawoad et al. proposes a secure logging scheme for cloud computing environment where the cloud service provider itself can be dishonest and can try to alter original logs [9]. Though there are no DSL for secure logging, there are some DSLs for providing access control facility on the audit logs or provenance

record and also for general-purpose access control. Ni et al. provided a XML-based access control language for general provenance model [28]. Using this language, users can define and evaluate access control policies on application audit logs. It also supports specifying policies to a particular record and its fields. Weissmann proposed ACS [29], an access control language for specifying access control policy, which especially resolves undecidability of granting or denying access, and incapability of editing control policy without changing the model. Ribeiro et al. provided SPL, an access control language for security policies with complex constraints [30]. SPL supports simultaneously multiple complex policies by resolving conflicts between two active policies. Beyond the permission / prohibition, they also showed how to express and implement the obligation concept.

VI. Conclusion and Future Work

For proper digital forensics investigation, maintaining the trustworthiness of logs is compulsory, and for this, we need a proper secure logging mechanism. To address the problem of secure logging mechanism, we have designed and implemented the domain-specific language FAL with the following benefits:

- Shifting the responsibility of developing a secure logging schemes from application programmers to security experts, which in turn increases trustworthiness.
- Required code to use specialized API for secure application logging is automatically generated. Hence, the effort and cost for developing secure logging scheme is lower.
- Heterogeneous formats of logs with any secure logging schemes can be easily handled.
- Detail understanding of specialized API for secure logging is not needed for FAL users.

We are working to add user specified delimiter feature with FAL. In future, we will add a user-friendly error reporting. Another important future feature is to incorporate timing option with system logging action. With this feature, users can define for how long they want to run system logging option. We will also work towards making FAL more robust so that it can generate audit-trailing code for all popular GPLs. Finally, FAL's design needs to be validated by end-users by performing usability studies and control experiments.

References

[1] FBI, "Annual report for fiscal year 2007," 2008 Regional Computer Forensics Laboratory Program, 2008, [Accessed July 5th, 2012].

[2] Congress of the United States, "Sarbanes-Oxley Act," http://thomas.loc.gov, 2002, [Accessed May 5th, 2013].

[3] U.S. Department of Health and Human Service, "Health information privacy," http://www.hhs.gov/ocr/privacy/, [Accessed May 5th, 2013].

[4] M. Swanson and B. Guttman, *Generally Accepted Principles and Practices for Securing Information Technology Systems*. National Institute of Standards and Technology (NIST), Technology Administration, US Department of Commerce, 1996.

[5] M. Bellare and B. Yee, "Forward-security in private-key cryptography," *Topics in Cryptology, CT-RSA 2003*, pp. 1–18, 2003.

[6] ——, "Forward integrity for secure audit logs," Technical report, Computer Science and Engineering Department, University of California at San Diego, Tech. Rep., 1997.

[7] D. Ma and G. Tsudik, "A new approach to secure logging," *Transaction of Storage (TOS)*, vol. 5, no. 1, pp. 2:1–2:21, Mar. 2009.

[8] B. Schneier and J. Kelsey, "Secure audit logs to support computer forensics," *ACM Transactions on Information and System Security (TISSEC)*, vol. 2, no. 2, pp. 159–176, May 1999.

[9] S. Zawoad, A. Dutta, and R. Hasan, "SecLaaS: Secure logging-as-a-service for cloud forensics," in *Proceedings of 8th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, May 2013.

[10] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM computing surveys (CSUR)*, vol. 37, no. 4, pp. 316–344, 2005.

[11] A. Van Deursen and P. Klint, "Little languages: Little maintenance?" *Journal of software maintenance*, vol. 10, pp. 75–92, 1998.

[12] T. Kosar, N. Oliveira, M. Mernik, V. J. M. Pereira, M. Črepinšek, C. D. Da, and R. P. Henriques, "Comparing general-purpose and domain-specific languages: An empirical study," *Computer Science and Information Systems*, vol. 7, no. 2, pp. 247–264, 2010.

[13] R. Accorsi, "On the relationship of privacy and secure remote logging in dynamic systems," in *Security and Privacy in Dynamic Environments*. Springer US, 2006, vol. 201, pp. 329–339. [Online]. Available: http://dx.doi.org/10.1007/0-387-33406-8_28

[14] A. Van Deursen and P. Klint, "Domain-specific language design requires feature descriptions," *Journal of Computing and Information Technology*, vol. 10, no. 1, pp. 1–17, 2004.

[15] P.-Y. Schobbens, P. Heymans, J.-C. Trigaux, and Y. Bontemps, "Generic semantics of feature diagrams," *Computer Networks*, vol. 51, no. 2, pp. 456–479, 2007.

[16] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[17] N. F. Pub, "197: Advanced encryption standard (AES)," *Federal Information Processing Standards Publication*, vol. 197, pp. 441–0311, 2001.

[18] B. Bryant, J. Gray, M. Mernik, P. Clarke, R. France, and G. Karsai, "Challenges and directions in formalizing the semantics of modeling languages," *Computer Science and Information Systems*, vol. 8, no. 2, pp. 225–253, 2011.

[19] T. Kosar, P. A. Barrientos, M. Mernik *et al.*, "A preliminary study on various implementation approaches of domain-specific language," *Information and Software Technology*, vol. 50, no. 5, pp. 390–405, 2008.

[20] M. Mernik and V. Žumer, "Incremental programming language development," *Computer Languages, Systems & Structures*, vol. 31, no. 1, pp. 1–16, 2005.

[21] T. Parr, "The definitive ANTLR reference: Building domain-specific languages (pragmatic programmers)," *Pragmatic Bookshelf, May*, 2007.

[22] E. Van Wyk, D. Bodin, J. Gao, and L. Krishnan, "Silver: an extensible attribute grammar system," *Electronic Notes in Theoretical Computer Science*, vol. 203, no. 2, pp. 103–116, 2008.

[23] P. R. Henriques, M. V. Pereira, M. Mernik, M. Lenic, J. Gray, and H. Wu, "Automatic generation of language-based tools using the LISA system," in *Software, IEE Proceedings-*, vol. 152, no. 2. IET, 2005, pp. 54–69.

[24] I. Fister, M. Mernik, and J. Brest, "Design and implementation of domain-specific language Easytime," *Computer Languages, Systems & Structures*, vol. 37, no. 4, pp. 151–167, 2011.

[25] D. E. Knuth, "Semantics of context-free languages," *Mathematical systems theory*, vol. 2, no. 2, pp. 127–145, 1968.

[26] J. Paakki, "Attribute grammar paradigms: a high-level methodology in language implementation," *ACM Computing Surveys (CSUR)*, vol. 27, no. 2, pp. 196–255, 1995.

[27] B. Schneier and J. Kelsey, "Secure audit logs to support computer forensics," *ACM Transactions on Information and System Security (TISSEC)*, vol. 2, no. 2, pp. 159–176, 1999.

[28] Q. Ni, S. Xu, E. Bertino, R. Sandhu, and W. Han, "An access control language for a general provenance model," *Secure Data Management*, pp. 68–88, 2009.

[29] M. Weißmann, "Domain specific language for specifying access controls," Ph.D. dissertation, Georg Simon Ohm University of Applied Sciences, Nuernberg, Germany, 2007.

[30] C. Ribeiro, A. Zuquete, P. Ferreira, and P. Guedes, "SPL: An access control language for security policies with complex constraints," in *Proceedings of the Network and Distributed System Security Symposium*, 2001, pp. 89–107.

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1575–1582

Dynamic loop reversal - the new code transformation technique

I. Šimeček, P. Tvrdík
Department of Computer Systems, Faculty of Information Technology,
Czech Technical University in Prague Prague, Czech Republic Email: xsimecek,pavel.tvrdik@fit.cvut.cz

***Abstract*—In this paper, we describe a new source code transformation called *dynamic loop reversal* that can increase temporal and spatial locality. We also describe a formal method for predicting the cache behaviour and evaluation results of the accuracy of the model by measurements on a cache monitor. The comparisons of the numbers of measured cache misses and the numbers of cache misses estimated by the model indicate that model is relatively accurate and can be used in practice.**

I. Introduction

LINEAR codes for dense linear algebra consist mainly of loops. A number of source code transformations techniques have been developed and used in the state-of-the-art compilers. In this paper, we consider the following standard techniques: *loop unrolling*, *loop blocking*, *loop fusion*, and *loop reversal* [1], [2], [3]. The main result of this paper is a description of a new transformation technique, called *dynamic loop reversal*, shortly *DLR*, to improve temporal and spatial locality.

Models for predicting the number of cache misses have also been developed for standard source code transformations [4], [5], [6], [7]. In order to incorporate the DLR into compilers, we propose such a model for the DLR in Section IV-B .

II. Terminology

Throughout the paper, we assume that indexes of vectors and matrices start from 1, all elements of vectors and matrices are of type *double* and that all matrices are stored in the row-major format.

A. The cache architecture model

We consider a *set-associative cache*. The number of sets is denoted by h. One set consists of s independent *blocks*. The size of the data part of a cache in bytes is denoted by DC_{S}. The cache block size in bytes is denoted by B_{S}. Then $DC_{\mathrm{S}} = s \cdot B_{\mathrm{S}} \cdot h$. The size of type `double` is denoted by S_{D}. We consider only *write-back* caches with *LRU block replacement* strategy.

B. The compressed sparse row (CSR) format

A matrix A is *dense* if it contains $\Theta(n^2)$ nonzero elements and it is *sparse* otherwise. In practice, a matrix is considered sparse if the ratio of nonzero elements drops bellow some threshold. The most common format (see [8], [9], [10]) for storing sparse matrices is the *compressed sparse row* (CSR) format. The number of nonzero elements is denoted by NZ_A. A matrix A stored in the CSR format is represented by three linear arrays $Elem_A$, $Addr_A$, and Ci_A. Array $Elem_A[1,\ldots,NZ_A]$ stores the nonzero elements of A, array $Addr_A[1,\ldots,n]$ contains indexes of initial nonzero elements of rows of A, and array $Ci_A[1,\ldots,NZ_A]$ contains column indexes of nonzero elements of A. Hence, the first nonzero element of row j is stored at index $Addr_A[j]$ in array $Elem_A$. The density of the matrix A (denoted by $density(A)$) is the ratio between NZ_A and n^2.

III. Code restructuring

In this section, we propose a new optimization technique called *dynamic loop reversal* (or alternatively *outer-loop-controlled loop reversal*).

A. Standard static loop reversal

In the standard loop reversal, the sense of the passage through the interval of a loop iteration variable is reversed. This rearrangement changes the sequence of memory requirements and reverses data dependencies. Therefore, it allows further loop optimizations in general.

Example code 1

```
1: for i ← n, 2 do
2:     B[i]+ = B[i − 1];
3: for i ← 2, n do
4:     A[i]+ = B[i];
```

Example code 1 represents a typical combination of data-dependent loops whose data dependency can be recognized automatically by common compiler optimization techniques. However, the first loop is *reversible* (it means that it is possible to alternate the sense of the passage). The reversal of the second loop and loop fusion can be applied and the reuse distances (see Section IV-A for the definition of the reuse distance) for memory transactions on array B are decreased.

Example code 2 Loop reversal and loop fusion applied to Example code 1

```
1: for i ← n, 2 do
2:     B[i]+ = B[i − 1];
3:     A[i]+ = B[i];
```

In Example code 3, data-dependency analysis reveals that the two loops are also reversible.

978-83-60810-53-8/$25.00

Example code 3

```
1: for i ← 1, n do                    ▷ Loop 1
2:     s+ = A[i] ∗ A[i];
3: norm = √s;
4: for i ← 1, n do                    ▷ Loop 2
5:     A[i]/ = norm;
```

However, the application of the loop reversal to the second loop decreases the reuse distances.

Example code 4 Loop reversal applied to Example code 3

```
1: for i ← 1, n do                    ▷ Loop 1
2:     s+ = A[i] ∗ A[i];
3: norm = √s;
4: for i ← n, 1 do                    ▷ Loop 2
5:     A[i]/ = norm;
```

The problem is that in this case (and in other similar cases), the compiler heuristics for the decision which loop to reverse to minimize reuse distances is complicated.

B. The effect of the static loop reversal on cache behaviour

If the size of array A is less than the cache size ($nS_{\mathrm{D}} \leq DC_{\mathrm{S}}$), then Example codes 3 and 4 are equivalent as to the cache utilization. However, if the size of array A exceeds the cache size, then no elements of A are reused in Example code 3,whereas the last $k = \frac{B_{\mathrm{S}}}{S_{\mathrm{D}}}$ elements of A are reused within the second reversed loop in Example code 4. So, the loop reversal improves the temporal locality (see Figures 1 and 2).

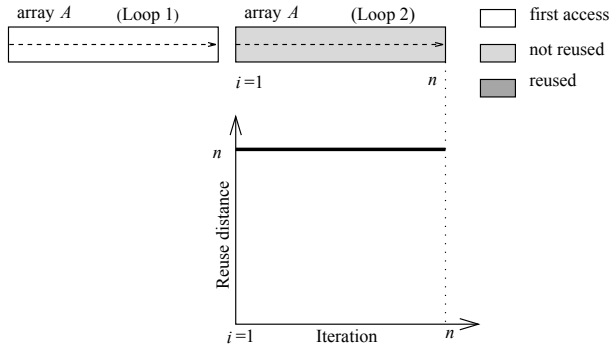


Fig. 1. The reuse distances in Example code 3.

C. Dynamic loop reversal

The static loop reversal is used to reverse data-dependency in one dimension. This has motivated us to generalize this idea and we have designed another optimization for nested reversible loops based on loop reversal. Consider the following code:

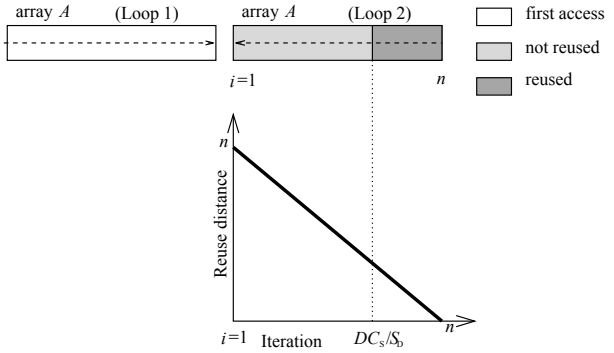


Fig. 2. The reuse distances in Example code 4.

Example code 5

```
1: for i ← 1, n do
2:     s = 0;
3:     for j ← 1, n do
4:         s+ = A[i][j] ∗ x[j];
```

The direction of the inner loop can be alternated forward and backward in even and odd iterations of the closest outer loop. In this way, we can use the positive effect of a loop reversal in every iteration of the outer loop. This is why we call it a *dynamic loop reversal*, or *DLR* for short. Example code 5 is a candidate for such a transformation.

Example code 6 DLR applied to Example code 5

```
1: for i ← 1, n do
2:     s = 0;
3:     if i is odd then
4:         for j ← 1, n do
5:             s+ = A[i][j] ∗ x[j];
6:     else
7:         for j ← n, 1 do
8:             s+ = A[i][j] ∗ x[j];
```

We will denote this transformation by DLR($i \rightarrow j$). For large arrays, this transformation leads to even better temporal locality than the original Example code 5, because it reduces the reuse distances and data reside in the cache from the previous iteration. On the other hand, the saving of the number of cache misses in one iteration of the outer loop is bounded by $DC_{\mathrm{S}}/B_{\mathrm{S}}$. So, the *DLR* has a significant effect if the cache size is comparable to the sum of affected arrays sizes in one iteration of the outer loop. The necessary condition for applying the *DLR* is that the inner loop must be reversible.

D. The application of DLR on triple-nested loops

In the previous text, the DLR was applied to double-nested loops, but it can also be applied to triple-nested loops. Consider the following code skeleton:

Example code 7

```
1: for i ← 1, n do
2:     for j ← 1, n do
3:         for k ← 1, n do
4:             (* loop body *)
```

In Example code 7, there are three options how the DLR can be applied:

- on the i-loop: DLR($i \to j$),
- on the j-loop: DLR($j \to k$),
- both transformations: DLR($i \to j$) and DLR($j \to k$).

The last option means composition of two transformations DLR($i \to j$) and DLR($j \to k$). This composition we will denoted by DLR($i \to j \to k$). In this case, the effect of DLR is twofold: DLR($i \to j$) (on the outer pair of loops) can improve temporal locality inside the L2 cache and DLR($j \to k$) (on the inner pair of loops) can improve temporal locality inside the L1 cache.

E. Comparison and possible combinations of DLR and other loop restructuring techniques

In this section, we describe some loop restructuring techniques (for details see [11], [12], [13], [14], [15]), compare them with DLR, and discuss their possible combinations with DLR.

1) Loop unrolling: Loop unrolling has two main effects. Firstly, it makes the sequential code longer, so it may improve data throughput, because the instructions could be better scheduled and the internal pipeline could be better utilized. Secondly, the number of test condition evaluations drops according to the unrolling factor. In general, the loop unrolling concentrates on maximizing the machine throughput, not on improving the cache behaviour.

2) Loop tiling (blocking): Loop tiling (sometimes called loop blocking or iteration space tiling) is one of advanced loop restructuring techniques. A compiler can use it to increase the cache hit rate. One possible motivation for using this technique is that the *loop range* (e.g., the size of the array traversed repeatedly within the loop) is too big and exceeds the data cache size DC_{S}. Thus, the loop should be split into two loops: the outer loop is the out-of-cache loop and the inner one is the in-cache loop. The value B_f is called the *tiling or block factor* and its optimal value depends on the size of the cache.

The loop tiling and DLR can be easily combined. DLR can be applied on every pair of immediately nested loop, but its useless to apply it for in-cache loops (i-loop, j-loop, and k-loop). We consider loop tiling as a competitor for DLR and we have performed experiments with both. These quantitative measurements of effects of these techniques are presented in Section VI-D.

IV. An Analytical Model of the Cache Behavior for the DLR

The *polytope model* (for details see [3], [6]) is used by modern compilers for an estimation of the parameters for loop restructuring techniques. We will present two cache behaviour models based on reuse distances (shortly RD).

A. A cache miss model with reuse distances

This model is inspired by the model introduced in [16]. We will call it the *basic RD model*.

Definition Consider an execution of an algorithm on the computer with load/store architecture and assume that addresses of memory transactions during this execution form a sequence $P[1,\ldots,n] = [addr_1,\ldots,addr_n]$. Then P is called a **sequence of memory access addresses** and $P[i] = addr_i$ is the i-th transaction with memory address $addr_i$. The **reuse distance** $RD(t)$, where $t \in \langle 1, n\rangle$, is the number of **different** memory addresses accessed between two uses of the address $P[t]$. Formally, if $P[t] = addr_t$ and $\epsilon(t) > 0$ is the minimal integer number such that $P[t-\epsilon(t)] = addr_t$, then $RD(t) = |\{P[t-\epsilon(t)],\ldots,P[t-1]\}|$. If such an $\epsilon(t)$ does not exist, then $RD(t) = \infty$, otherwise $RD(t) \le \epsilon(t)$.

The notion of reuse distances can be used for developing a simple cache miss model based on estimating the numbers of thrashing misses in fully-associative ($h = 1$) caches. If $RD(t) > DC_{\mathrm{S}}/S_{\mathrm{D}}$, then the content of the cache block from the memory address $P(t)$ is replaced by some new value and a cache miss occurs. If $RD(t) = \infty$, then a compulsory miss occurs, otherwise a thrashing miss occurs. Recall that we assume only caches with LRU block replacement strategy.

In this basic RD model, the spatial locality of the cache memory is not considered, i.e., it is assumed that a cache block contains exactly one array element ($B_{\mathrm{S}} = S_{\mathrm{D}}$). However, $B_{\mathrm{S}} = c \cdot S_{\mathrm{D}}$, where c is typically 4 or 8 in modern processors, and therefore, spatial locality must be taken into account in order to have a more realistic model.

B. A simplified cache miss model for the DLR

Even the basic RD model is too complicated for modelling the cache behavior of DLR in real applications. Hence, we introduce another model that is even more simplified. We call this model *simplified RD model*. We use this model for enumeration cache misses saved by DLR. To derive an analytical model of the effect of the DLR on the cache behaviour, consider the following code skeleton representing most often memory access patterns during a matrix computation:

Example code 8

```
1: statement_1;
2: for i ← i_1, i_2 do
3:     statement_2;
4:     for j ← j_1, j_2 do
5:         statement_3;
6:         = B[j];          ▷ Memory operation of type α
7:         = B[i];          ▷ Memory operation of type β
8:         = A[i][j];       ▷ Memory operation of type γ
9:         = A[j][i];       ▷ Memory operation of type δ
10:    statement_4;
11: statement_5;
```

We consider the following simplifying conditions:

A1 We assume that all matrices are stored in the row-major order.

A2 We assume that *statements*$_{1-5}$ contain only local computation with register operands. That is, we assume that *statements*$_{1-5}$ have negligible cache effects and the only memory accesses are memory operations of type $\alpha - \delta$.

A3 We assume that the reuse distances depend on the exact ordering of memory operations (inside the j-loop) only slightly and so do the number of cache misses.

A4 We do not distinguish between load and store operations.

A5 We assume that the cache memory is big enough to hold all the data for one iteration of the (inner) j-loop.

A6 We assume that the cache memory is not able to hold all the data for one iteration of the outer i-loop. Otherwise, the DLR has no effect in comparison to standard execution.

A7 This model is derived only for immediately nested loops.

Let us now analyse the effect of DLR($i \to j$) on individual memory operations.

- A memory operation of type α is affected by the DLR, because its operand (or its part) can be reused. The effect of DLR can be estimated by the RD analysis.
- A memory operation of type β is not affected by the DLR, because it returns the same value (in the j-loop). It is usually eliminated by an optimizing compiler.
- A memory operation of type γ is not affected by the DLR, because its operand cannot be reused due to the row-major matrix format assumption.
- A memory operation of type δ is affected by the DLR, due to its spatial locality.

1) Evaluation of simplified RD model: The number of cache misses during one execution of Example code 8 is denoted by X. The number of cache misses during one execution of Example code 8 with DLR($i \to j$) is denoted by Y. The reduction of the number of cache misses during one execution of Example code 8 due to the DLR($i \to j$) is denoted by μ_{saved} and it is equal to $X - Y$. The value of μ_{saved} has an upper bound

$$\mu_{\text{saved}} \leq (i_2 - i_1) \cdot DC_{\text{S}}/B_{\text{S}}.$$

This general upper bound can be reached only for loops where all memory operations are affected by the DLR. In practical cases, the reduction of the number of cache misses is smaller. To estimate the reduction of the number of cache misses during an execution of Example code 8 with the DLR, we need to count the number of iterations of the j-loop that can reside in the cache. We will denote this number by N_{iter}

$$N_{\text{iter}} = \frac{DC_{\text{S}}}{B_{\text{S}} \sum_m SCMO(m)}, \quad (1)$$

where

- m is a memory operation (of type $\alpha - \delta$) in the j-loop,
- $SCMO(m)$ is the probability that memory operation m loads data into a new cache block.

$$SCMO(m) = \begin{cases} 1 & \text{if } m \text{ is a memory operation of types } \beta \text{ or } \delta \text{ which are accessed in column-like pattern.} \\ S_{\text{D}}/B_{\text{S}} & \text{if } m \text{ is a memory operation of types } \alpha \text{ or } \gamma \text{ which are accessed in row-like pattern.} \end{cases} \quad (2)$$

If $N_{\text{iter}} < 1$, then the assumption (A5) is not satisfied and $\mu_{\text{saved}} = 0$.
If $N_{\text{iter}} \geq (j_2 - j_1)$, then the assumption (A6) is not satisfied and $\mu_{\text{saved}} = 0$.

We can also estimate probability (denoted by $PDLR(m)$) that the memory location accessed by memory operation m is reused using DLR.

$$PDLR(m) = \begin{cases} 0 & \text{if } m \text{ is a memory operations of types } \beta \text{ or } \gamma \text{ (i.e., it is not affected by the DLR;)} \\ 1 - S_{\text{D}}/B_{\text{S}} & \text{if } m \text{ is a memory operation of type } \delta \text{ (i.e., it is affected by the DLR, for column-like access, the last element in cache-line is not counted;)} \\ 1 & \text{if } m \text{ is a memory operation of type } \alpha \text{ (i.e., it is affected by the DLR,for row-like access.)} \end{cases} \quad (3)$$

Finally, the number of cache misses saved by the DLR applied to the i-loop can be approximated by

$$\mu_{\text{saved}} = (i_2 - i_1) \cdot N_{\text{iter}} \sum_m \big(PDLR(m) \cdot SCMO(m)\big), \quad (4)$$

where m is a memory operation in the j-loop.

Comparisons of the numbers of estimated and measured cache misses are presented in Section VI-C3.

V. Experimental evaluation of the DLR

A. Testing codes

For measuring of the effect of DLR (performance, cache miss rate, and so on), we use two simple codes:

- matrix-matrix multiplication (MMM for short),
- multiplication of two sparse matrices (spMMM for short).

We have deeply studied characteristics of these codes in following sections:

- For performance results, see Section VI-A.
- For cache utilization results, see Section VI-B.
- We also evaluate precision of our analytical model for MMM_STD code, see Section VI-C.
- We also combine effects of DLR and loop tiling for MMM_STD code, see Section VI-D.

1) Matrix-matrix multiplication: We consider input real square matrices A and B of order n. A standard sequential pseudocode for matrix-matrix multiplication $C = A \cdot B$ is the following:

```
1: procedure MMM_STD(in A,B;out C)
2:     for i ← 1, n do
3:         for j ← 1, n do
4:             sum = 0;
5:             for k ← 1, n do
6:                 sum+ = A[i][k] * B[k][j];
7:             C[i][j] = sum;
8:     return C;
```

2) Multiplication of two sparse matrices: We consider input real square sparse matrices A and B of order n represented in the CSR format (see Section II-B), output matrix C is a dense matrix of order n. A standard sequential pseudocode for the sparse matrix-matrix multiplication $C = A \cdot B$ can be described by the following pseudocode:

```
1: procedure SPMMM_CSR(in A,B;out C)
2:     for y ← 1, n do
3:         for i ← A.Addr[y], A.Addr[y + 1] − 1 do
4:             x = A.Ci[i];
5:             for j ← B.Addr[x], B.Addr[x + 1] − 1 do
6:                 x2 ← B.Ci[j];
7:                 C[y][x2]+ = A.Elem[i] * B.Elem[j];
8:     return C;
```

B. Configuration of the experimental system

All cache events were evaluated by our software cache emulator [17] and verified by the Intel Vtune tool. The experiments were performed on the Pentium 4 Celeron at 2.4 GHz, 512 MB, running OS Windows XP Professional, with the following cache parameters:

- L1 data cache with $DC_S = 8K$, $B_S = 32$, $s = 4$, $h = 64$, and LRU strategy.
- L2 unified cache with $DC_S = 128K$, $B_S = 32$, $s = 4$, $h = 1024$, and LRU strategy.

We used the Intel compiler version 7.1 with switches:
`-O3 -fno_alias -xK -ipo`

VI. The Results of Experimental Evaluation

A. Performance evaluation of testing codes

We count every floating point operation (multiplication, addition and so on). The performance in MFLOPS is then defined as follows:

$$\text{MFLOPS(MMM\_STD)} = \frac{2n^3}{\text{execution time } [\mu\text{s}]}$$

$$\text{MFLOPS(SPMMM\_CSR)} = \frac{2 \cdot NZ_A \cdot NZ_B}{n \cdot \text{execution time } [\mu\text{s}]}$$

The graph in Figure 3 illustrates the performance with or without DLR. These graphs illustrate that the DLR increases the code performance due to better cache utilization. There

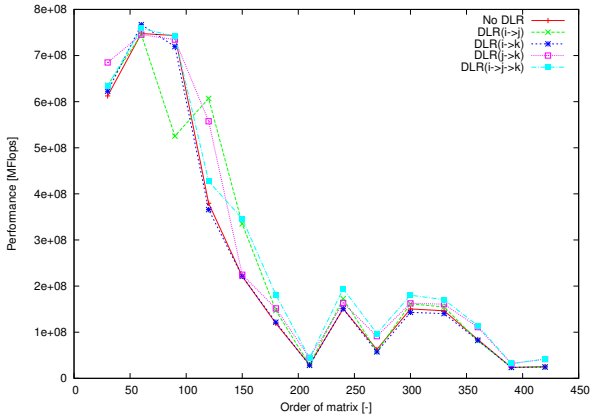


Fig. 3. Performance of MMM_STD.

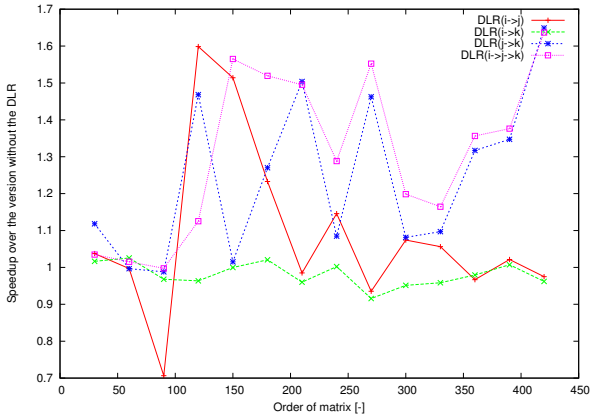


Fig. 4. Speedup of MMM_STD

is a performance gap (for example for $n = 120$ for the MMM_STD), which DLR can overcome. The graph in Figure 4 shows the speedup over the version without the DLR. We can conclude that the fastest code is the version with DLR($i \to j \to k$) for the MMM_STD code. We can also conclude that the average measured speedup is more than 20% in the measured set for the MMM_STD code.

For small matrices, a small slowdown was measured. While the DLR can improve the cache hit rate, it has more overhead due to more conditional loops. This effect becomes even more important for the DLR on triple loops.

B. Cache miss rate evaluation

The cache utilization is enumerated according to the following definitions. Let us define "relative number of cache misses" as the ratio between the number of cache misses with DLR and the number of cache misses without DLR.

The graphs on Figures 5 and 6 illustrate the number of cache misses occurring during one execution of the MMM_STD pseudocode. We can conclude that

- the DLR effect depends on the value of the parameter n and on the cache memory size (this observation proves the results of the analytical model from Section IV-B)

- except for few cases, the DLR transformation has a positive impact on cache utilization.

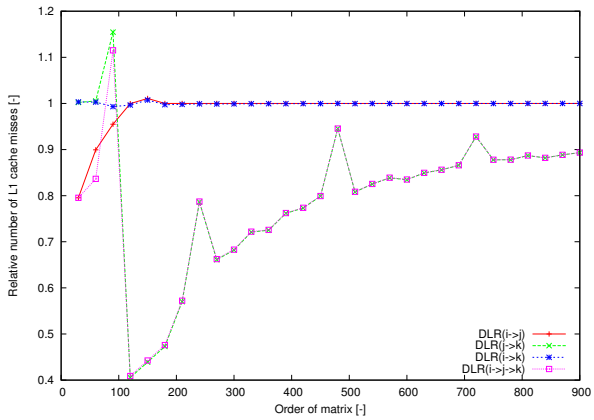


Fig. 5. Relative number of cache misses during MMM_STD for L1 cache

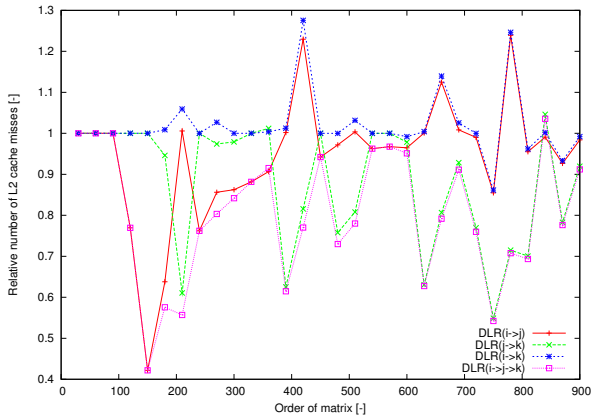


Fig. 6. Relative number of cache misses during MMM_STD for L2 cache

C. Evaluation of simplified RD model

1) Analytical cache model for MMM_STD*:* To analyse this algorithm, we omit accesses in array C at code line 7, because they are much less frequent. In this simplified model, the algorithm contains the following types of memory accesses:

- If DLR($i \to j$) is applied, then memory operations with $A[i][k]$ are of type β and memory operations with $B[k][j]$ are of type α.
- If DLR($j \to k$) is applied, then memory operations with $A[i][k]$ are of type α and memory operations with $B[k][j]$ are of type δ.

2) Analytical cache model for SPMMM_CSR*:* Analysis of cache behaviour and DLR effects for this algorithm are beyond the scope of the compiler due to its irregular memory pattern.

3) An example of evaluation of the cache analytical model: We apply DLR($j \to k$) on the MMM_STD pseudocode. In this case as we stated above, memory operations with $A[i][k]$ are of type α and memory operations with $B[k][j]$ are of type δ.

Firstly, we must count how many iterations of the j-loop can reside in the cache. From the types of memory operations (Eq. (2)), we can derive that

$$SCMO(A[i][k]) = S_{\mathrm{D}}/B_{\mathrm{S}}, \quad PDLR(A[i][k]) = 1.$$

$$SCMO(B[k][j]) = 1, \quad PDLR(B[k][j]) = 1 - S_{\mathrm{D}}/B_{\mathrm{S}}.$$

So, the number of iterations is (from cache parameters in Eq. (1)

$$N_{\mathrm{iter}} = \frac{DC_{\mathrm{S}}}{B_{\mathrm{S}}(1 + S_{\mathrm{D}}/B_{\mathrm{S}})}.$$

The number of cache misses saved by DLR(k, j) per one iteration of the j-loop (Eq. (4)) is $\mu_{\mathrm{saved}} = N_{\mathrm{iter}}$.

The total number of cache misses saved by DLR($j \to k$) during one execution of the MMM_STD pseudocode is

$$\text{total } \mu_{\mathrm{saved}} = n^2 N_{\mathrm{iter}}.$$

For the given cache configuration, it gives the following results:

- for L1 cache: $N_{\mathrm{iter}} = 228$.
- for L2 cache: $N_{\mathrm{iter}} = 3640$.

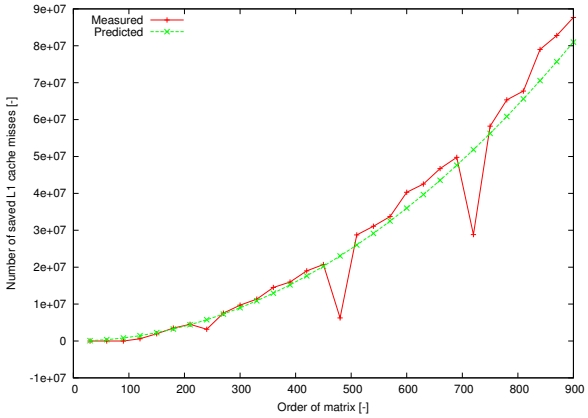


Fig. 7. Comparison of the numbers of estimated and measured cache misses (μ_{saved}) saved by the DLR during the execution of MMM_STD for L1 cache.

Comparisons of the numbers of estimated and measured cache misses are shown in Figures 7 and 8.

4) Discussion of the precision of the simplified RD model: Our analytical model is derived from the RD, which is based on fully-associative cache memory assumption. This assumption is the main source of errors in predictions. The errors are higher for L2 caches due to their lower associativity.

D. Evaluation of combination of DLR and loop tiling

We have also measured the performance and cache utilization for pseudocode MMM_STD with loop tiling and effects of the DLR transformation on this code. Graphs on Figures 9 and 10 illustrate the fact that loop tiling can greatly improve the cache utilization. On the other hand, the tiling factor must be chosen very carefully, because the number of cache misses grows quickly with the distance of the tiling factor from the

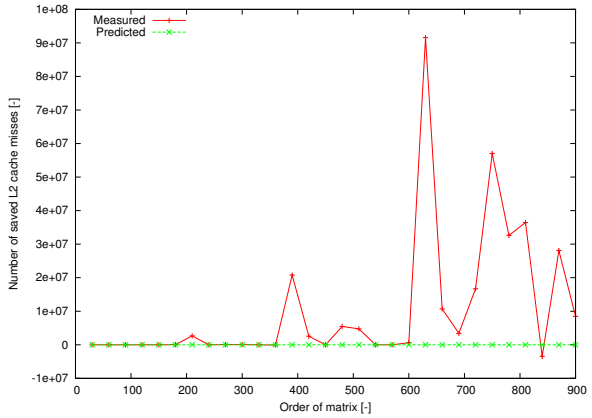


Fig. 8. Comparison of the numbers of estimated and measured cache misses (μ_{saved}) saved by the DLR during the execution of MMM_STD for L2 cache.

optimal value. When the DLR is applied, the growth is more smooth, so the code is less sensitive to the tiling factor value. Hence, the DLR technique is useful in cases when it is hard to predict a good value for the tiling factor.

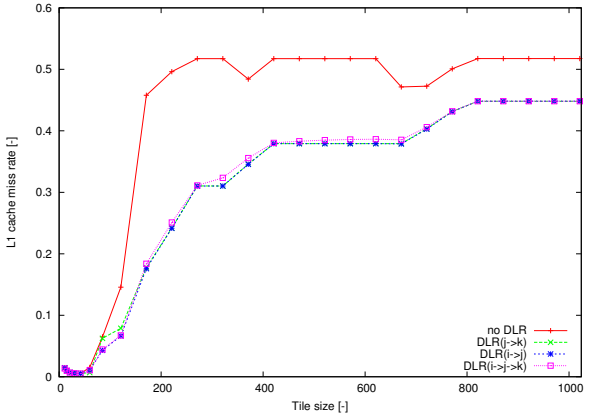


Fig. 9. The L1 miss rate for MMM_STD code with loop tiling (for n=1024) for different values of the tile size.

E. Evaluation of the DLR for the SPMMM_CSR code

The SPMMM_CSR code is a simple example of an irregular code. For the testing purposes, we always generate five sparse matrices with random locations of nonzero elements with given properties (order of matrix, number of nonzero elements or density). The average value of these five measurements were taken as a result. In this code, the memory access pattern is hard to predict on the compiler level and loop tiling is excluded. Thus the DLR is usable and the application of this technique can save reasonably large number of cache misses (see Figures 11,12, and 13).

VII. Automatic compiler support of the DLR

The DLR transformation brings new possibilities to optimize nested loops.

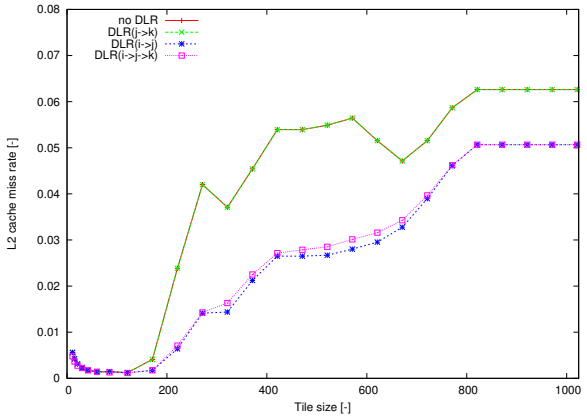


Fig. 10. The L2 miss rate for MMM_STD code with loop tiling (for n=1024) for different values of the tile size.

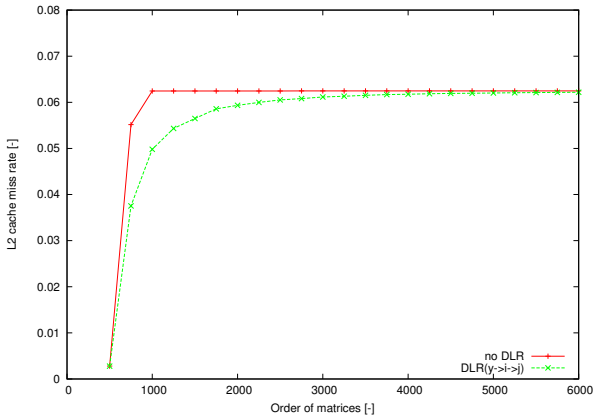


Fig. 11. The L1 and L2 miss rate for SPMMM_CSR algorithm (for $density(A) = 7\%$ and $density(B) = 21\%$)

A. A proposed algorithm of automatic compiler support of the DLR

Let $\mathcal{L}_{1\ldots b}$ represent a hierarchy of immediately nested loops ($\mathcal{L}_1$ is the outermost loop, $\mathcal{L}_b$ is the innermost loop). The control variable for the loop $\mathcal{L}_i$ is denoted by $\mathcal{C}_i$. We propose the following function that returns a list of loop numbers that can profit from the DLR application and that can be implemented into compiler to support the DLR application automatically.

1: **procedure** DLR_APPLICATION(in $b, \mathcal{L}, \mathcal{C}$)
2: $res = []$;
3: **for** $i \leftarrow 1, b-1$ **do** ▷ here we consider application of DLR($\mathcal{C}_i \rightarrow \mathcal{C}_{i+1}$)
4: **if** this DLR application is possible **then**
5: compute μ_{saved} from the proposed cache model;
6: compute overhead of this DLR application;
7: **if** this DLR application pays-off **then**
8: add i to the res;
9: **return** res;

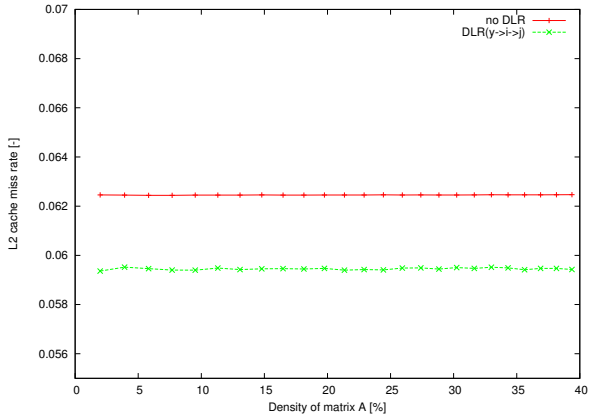


Fig. 12. The L2 miss rate for algorithm SPMMM_CSR (n=2200 and $density(B) = 17\%$)

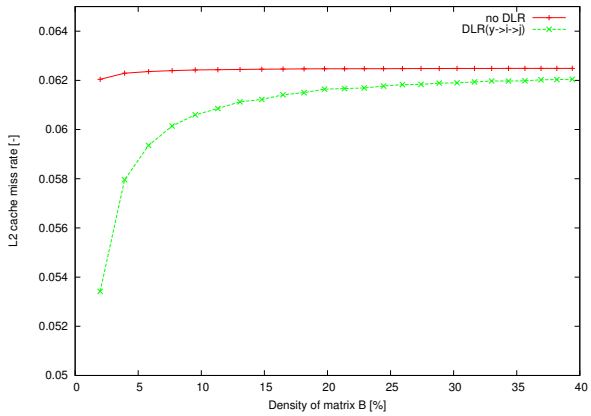


Fig. 13. The L2 miss rate for algorithm SPMMM_CSR (n=3700 and $density(A) = 10\%$)

If this function returns empty list, then DLR does not pay-off for any loop in $\mathcal{L}$. In other case, it returns a list (res) of loop number x such that the DLR should be applied to $\mathcal{L}_x$, i.e., DLR($\mathcal{C}_x \rightarrow \mathcal{C}_{x+1}$) to increase the code performance.

B. Discussion of applicability of the DLR inside compilers

The function in Section VII-A is very general. The real incorporation of the DLR into existing compilers (like GCC or LLVM) must address more issues:

- Where can be the DLR applied? The DLR can be applied on the nested reversible loops. This condition can be easily checked by the compiler.
- Where should be DLR applied? The DLR should be applied on a pair or triple of loops that causes its maximal effect (mentioned in Section III-C). This compiler decision is very similar as for loop tiling.
- Has DLR significant effect? Yes. In most cases, higher speedups are achieved by loop unrolling or loop tiling. But the DLR can be combined with these techniques (see Section VI-D) and also the DLR can be applied on some codes where loop tiling could not (for example sparse matrix operations).

VIII. Conclusions

We have described a new code transformation technique, the dynamic loop reversal, whose goal is to improve temporal locality. This transformation seems to be very useful for codes with nested loops. We have demonstrated significant performance gains for two basic algorithms from linear algebra.

We have also developed a probabilistic analytical model for this transformation and compared the numbers of measured cache misses and the numbers of cache misses estimated by the model. The inaccuracies of the model are due to some simplifying assumptions.

This work is to contribute to the development of more efficient compiler techniques.

References

[1] K. Kennedy and J. R. Allen, *Optimizing compilers for modern architectures: a dependence-based approach*. Morgan Kaufmann Publishers Inc., 2002.

[2] K. R. Wadleigh and I. L. Crawford, *Software optimization for high performance computing*. Hewlett-Packard professional books, 2000.

[3] M. Wolfe, *High-Performance Compilers for Parallel Computing*. Addison-Wesley, Reading, Massachusetts, USA, 1995.

[4] X. Vera and J. Xue, "Efficient compile-time analysis of cache behaviour for programs with IF statements," Beijing, October 2002. [Online]. Available: citeseer.ist.psu.edu/567600.html

[5] N. Ahmed, N. Mateev, and K. Pingali, "Tiling imperfectly-nested loop nests," in *Proceedings of the 2000 ACM/IEEE conference on Supercomputing (CDROM)*. IEEE Computer Society, 2000, p. 31.

[6] J. Xue, *Loop tiling for parallelism*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.

[7] P. Tvrdík and I. Šimeček, "Analytical model for analysis of cache behavior during cholesky factorization and its variants," in *Proceedings of the International Conference on Parallel Processing Workshops (ICPP 2004)*, vol. 12, Montreal, Canada, 2004, pp. 190–197. [Online]. Available: http://dl.acm.org/citation.cfm?id=1018426.1020360

[8] E. Im, *Optimizing the Performance of Sparse Matrix-Vector Multiplication - dissertation thesis*. University of Carolina at Berkeley: Dissertation thesis, 2001.

[9] D. B. Heras, J. C. Cabaleiro, and F. F. Rivera, "Modeling data locality for the sparse matrix-vector product using distance measures," *Parallel Computing*, vol. 27, no. 7, pp. 897–912, Jun. 2001.

[10] P. Tvrdík and I. Šimeček, "Analytical modeling of optimized sparse linear code," in *Parallel Processing and Applied Mathematics*, vol. 3019/2004, no. 4, Czestochova, Poland, 2003, pp. 207–216. [Online]. Available: http://www.springerlink.com/content/drwdhen7db199k05/

[11] S. Carr, K. S. McKinley, and C.-W. Tseng, "Compiler optimizations for improving data locality," in *In Proceedings of the Sixth International Conference on Architectural Support for Programming Languages and Operating Systems*, 1994, pp. 252–262.

[12] M. E. Wolf and M. S. Lam, "A data locality optimizing algorithm," *SIGPLAN Not.*, vol. 26, pp. 30–44, May 1991. [Online]. Available: http://doi.acm.org/10.1145/113446.113449

[13] S. Carr and R. Lehoucq, "Compiler blockability of dense matrix factorizations," *ACM Transactions on Mathematical Software*, vol. 23, pp. 336–361, 1996.

[14] Y. Song and Z. Li, "New tiling techniques to improve cache temporal locality," *SIGPLAN Not.*, vol. 34, pp. 215–228, May 1999. [Online]. Available: http://doi.acm.org/10.1145/301631.301668

[15] X. Vera and J. Xue, "Let's study whole-program cache behaviour analytically," in *Proceedings of the 8th International Symposium on High-Performance Computer Architecture*, ser. HPCA '02. Washington, DC, USA: IEEE Computer Society, 2002, pp. 175–. [Online]. Available: http://dl.acm.org/citation.cfm?id=874076.876456

[16] K. Beyls and E. D'Hollander, "Reuse distance as a metric for cache behavior," in *Proceedings of PDCS'01*, August 2001, pp. 617–662. [Online]. Available: citeseer.ist.psu.edu/beyls01reuse.html

[17] P. Tvrdík and I. Šimeček, "Software cache analyzer," in *Proceedings of CTU Workshop*, vol. 9, Prague, Czech Republic, Mar. 2005, pp. 180–181.

Author Index