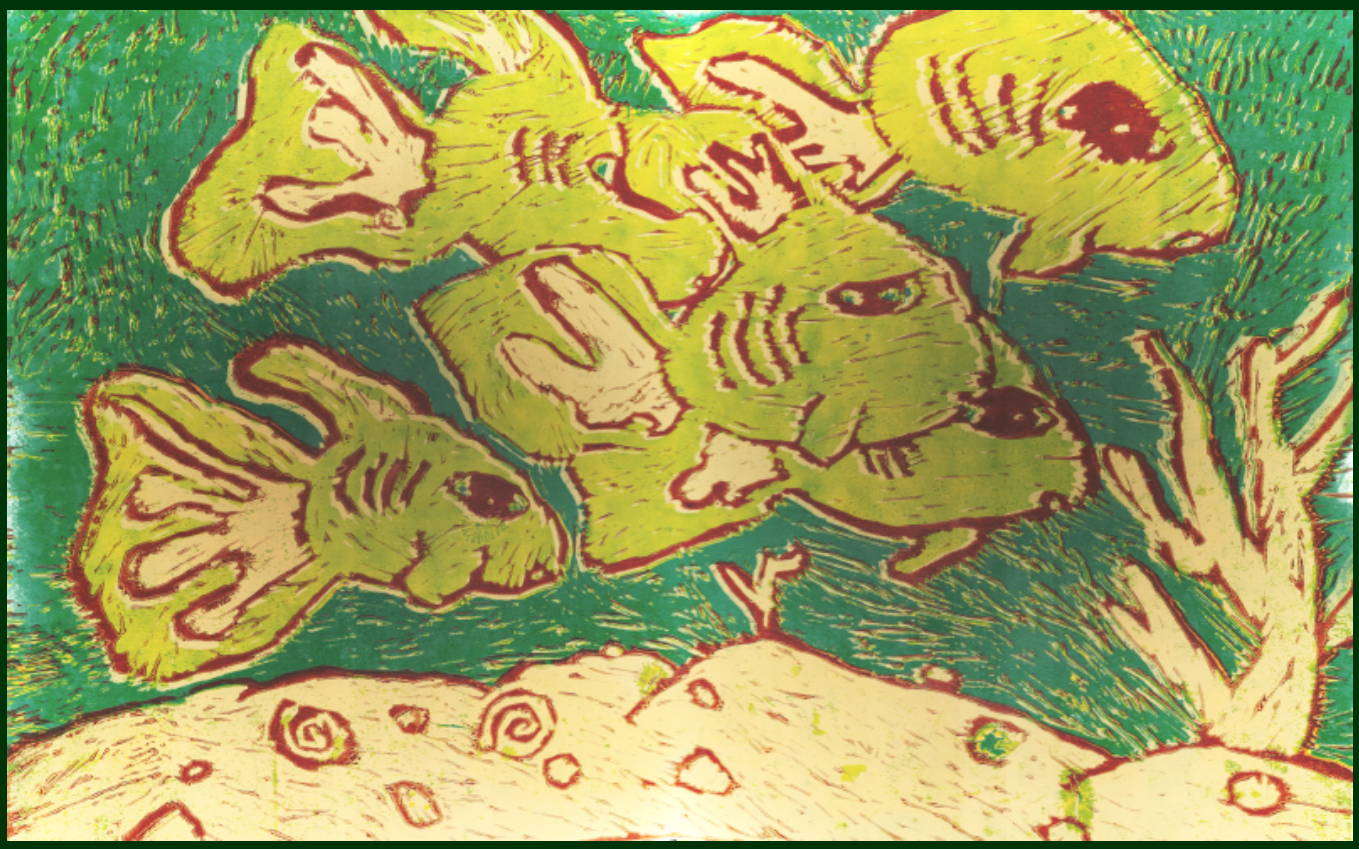
Annals of Computer Science and Information Systems
Volume 1

Maria Ganzha, Leszek Maciaszek, Marcin Paprzycki (eds.)



Position Papers of the 2013 Federated Conference on Computer Science and Information Systems

September 8–11, 2013. Kraków, Poland

Annals of Computer Science and Information Systems

Contact: secretariat@fedcsis.org

WWW site: `http://www.fedcsis.org/annals/`

ISSN: 2300-5963

Annals of Computer Science and Information Systems, Volume 1

Position Papers of the 2013 Federated Conference on Computer Science and Information Systems

Maria Ganzha
Leszek Maciaszek
Marcin Paprzycki (eds.)

2013, Warszawa, Polskie Towarzystwo Informatyczne

Annals of Computer Science and Information Systems, Volume 1

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

ISBN WEB: 978-83-60810-55-2
ISBN USB: 978-83-60810-56-9

ISSN: 2300-5963

Dear Reader, it is our pleasure to present to you Position Papers of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), which took place in Kraków, Poland, on September 8–11, 2013. This is the first year when position papers have been introduced as a separate category of contributions. This decision was grounded in the observation that, with decreasing acceptance rate of full papers (this year approximately 43%), a number of valuable contributions have to be rejected. They often represent initial stages of interesting projects, descriptions of industrial applications, etc. Therefore, in order to not to lose interesting contributions, this volume was created.

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

FedCSIS consisted of the following events:

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

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

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

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

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

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

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

September 8–11, 2013. Kraków, Poland

TABLE OF CONTENTS

Innovative Network Systems and Applications

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

2ND International Conference on Wireless Sensor Networks

Information Technology for Management, Business & Society

11TH Conference on Advanced Information Technologies for Management

Software Systems Development & Applications

4TH Workshop on Advances in Programming Languages

8th International Symposium
Advances in Artificial Intelligence and Applications

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

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

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

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 3–7

Computer Aided Material Selection in Design Process

Marcin Jaromin
Rzeszow University of Technology al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
E-mail:mjaromin@prz.edu.pl

***Abstract*— The selection of proper materials for different components is one of the most challenging tasks in engineer's activity. The traditional approach to the selection of materials required looking through the literature and material safty data sheets.With the development of information technology, engineering work has also changed. It involves not only arduous browsing catalogues and guides but also defining queries in database applications. This approach greatly accelerated the process of selection of materials, but it did not eliminate all the problems associated with obtaining information from large data sets. This paper presents the developed Web Material Selector application designed for computer-aided materials selection. The WMS allows to work with very large databases through the use of a professional tool to manage complex data structures. At the same time the application is designed to optimize the search process, what significantly reduces the number of materials whose properties a designer should carefully consider.**

I. Introduction

IT is estimated that there are more than 80 000 materials on the market, each of which is described by a series of properties: mechanical, physical, chemical, economic, etc. These numbers show how difficult task was put before an engineer who must made the best choice of the optimal material from which the given item will be made The traditional approach to the selection of materials for specific applications required looking through the literature and material safty data sheets.

Realising the potential for analysis and searching large data sets offered by modern database management systems, to some extent facilitates the work of structural engineers. However, existing solutions do not eliminate all the problems associated with obtaining information from large data sets. On the one hand, the engineer would have the greatest database, which will enable the analysis of the broadest range of materials properties. On the other hand, when searching large databases, the number of materials that meet the very stringent criteria (on the assumption that the element does not operate in extreme conditions) is very high, which hampers the decision process and extended the selection of the best material. The solution to these problems may be to design a new algorithm for the selection of materials in the design process, using the methods of artificial intelligence in the data classification and multi-criteria decision-making models. The main function of this algorithm will reduce and prioritize the search results to databases in such a way that making the decision on the selection of the material as easy as possible.

II. Methodology for the Selection of Materials

Rules for the selection of materials in engineering design has been described in detail in the works of Dobrzański [1] and Ashby [2].The authors defined a complex decision-making process that requires good knowledge of the widely understood materials science and related areas. The procedure of an engineer depends to a large extent on the features that the projected element will meet and the type of project (original, adaptive, alternative). Fig. 1 shows the general scheme of the process of selection of materials.

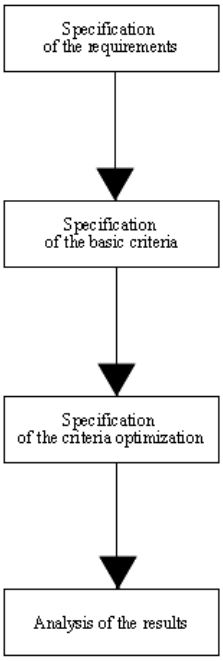


Fig. 1 Methodology for the selection of materials

In the first stage there are a reconnaissance of design and constructional assumptions and analysis of material requirements. In the next step, the basic criteria are formulated. They are strict conditions that allow us to assess whether the given material can be used in the present analyzed case or whether it should be unconditionally rejected. The result of the analysis of the material base in terms of the basic criteria is to obtain the space of solutions, ie the set of materials that may be used in a given project.

In the next step optimization criteria (functional) are described. These criteria are not strict conditions but they only define the value of the objective function, thus indicating the direction to be pursued. The analysis of the space of solutions, taking into account these criteria, allows to organize the list of pre-selected materials. The order of materials on the list is dependent on the optimization algorithms used in the process. If there are a lot of materials that meet the basic criteria, as a result of optimization, the worst materials that meet the specified criteria can also be removed from the list. The last step in the process of designing involves a detailed analysis of materials ranked as a space of solutions in terms of the possibility of using a given material in a specific project. Here assessment, apart from properties, includes so. local conditions, making the selection of the best material from dependent on the owned equipment, the availability of material on the market, confidence in providers, etc.

The vast amount of materials and a lot of properties which are used to describe them justify the attempts at working out computer systems supporting the process of material selection.

III. Systems Architecture of WMS

The few currently existing solutions of computer-aided materials selection have a number of constraints. The flagship program Cambridge Engineering Selector based on the idea of Ashby has a very general approach to the material data that only allows pre-selection. However, applications developed under the leadership of Dobrzański are equipped with much more extensive and accurate database but supporting the process of selection is limited only by the multi-criteria search of these databases.

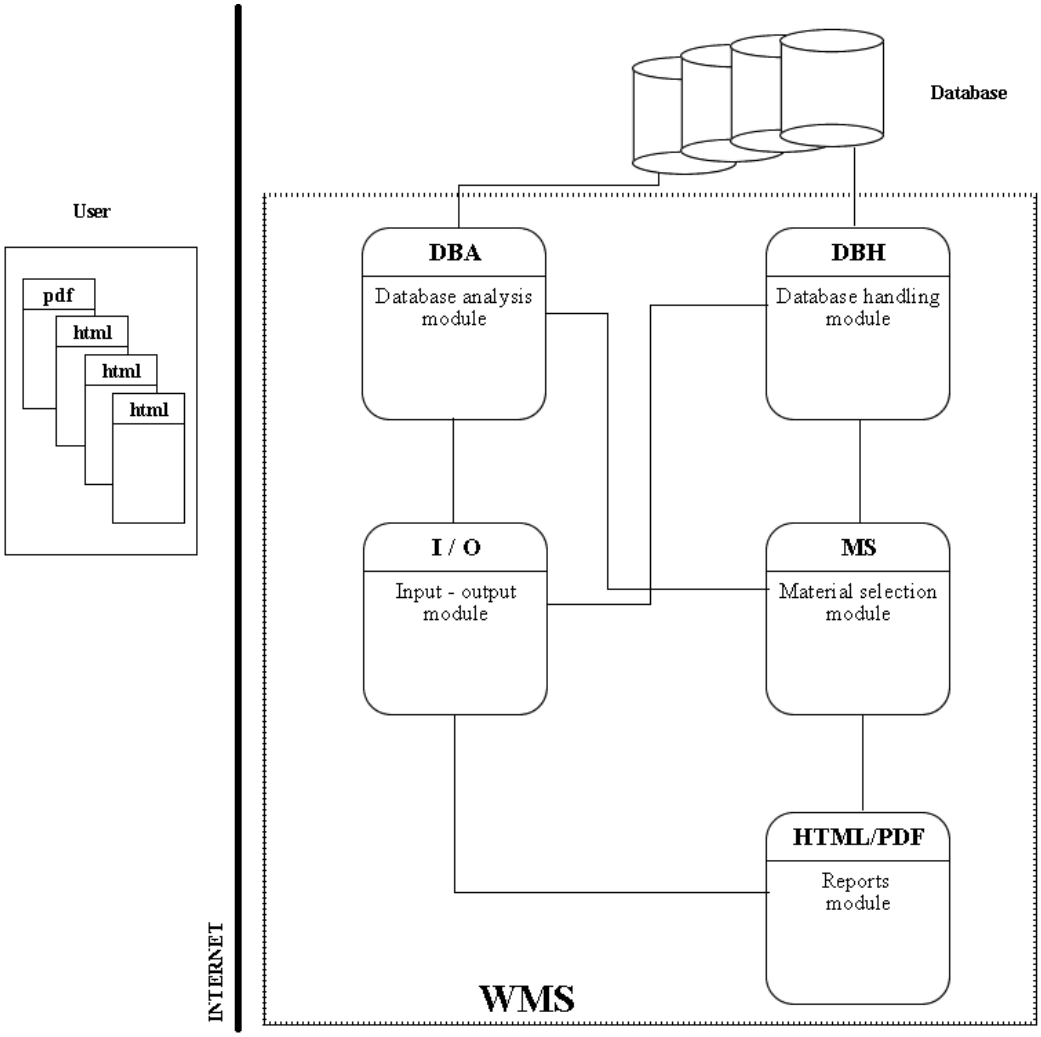


Fig. 2 Multi-layer system architecture of WMS

Originally the WMS program was supposed to eliminate these restrictions. It should support accurate material data for the largest possible number of materials, enable multi-criteria search of database and use optimization algorithms supporting the choice of the best materials. Additionally, in line with current trends in computing, it was decided to create a multi-user network applications with the ability of Internet service and the use of professional database management system of the company IBM. To be able to create a complete documentation of the process of material selection a print option of reports should be introduced.

WMS application designed for computer-aided materials selection is placed on the servers of Rzeszów University of Technology. It was established in accordance with the standard J2EE(Java Enterprise Edition) that defines the rules for creating applications in the Java programming language based on multi-layered architecture. Users gain access to the application using a web browser.

The system has a structure typical of an expert system where the machine is separated from the applicant database (Fig. 2). Entering the data related to the material sought takes place through dynamically generated web pages. Selection of appropriate criteria is determined by a database connected to the system, from which a web page is generated through analysis module and the input-output module.

For proper operation of the program it is required to establish a connection to a database of materials, for demonstration purposes there are two sample databases available. The first database "Plastics" contains information on over 500 plastics and polymer matrix composites. These materials can be found in the commercial offer of five leading manufacturers of plastic products. The second database shows the properties of the materials belonging to the group of stainless steels. It contains information on 300 materials (about 100 kinds in various forms of delivery and subjected to different heat treatment processes) and their properties. The program also offers the possibility to create their own databases, so that users can create their own database of materials.

The application service is very intuitive and is done using a web browser. It provides three basic modes of operation:

- Browsing a database;
- Editing and creating a database;
- Computer-aided materials selection.

The action of this last option is illustrated in Fig. 3. After choosing the appropriate database and the option "Selection of materials" in the menu we can determine information about the executed project. This option allows to enter information concerning the name, description, and additional design elements (such as graphics, calculation, charts); it enables to create a full documentation of the process of material selection.

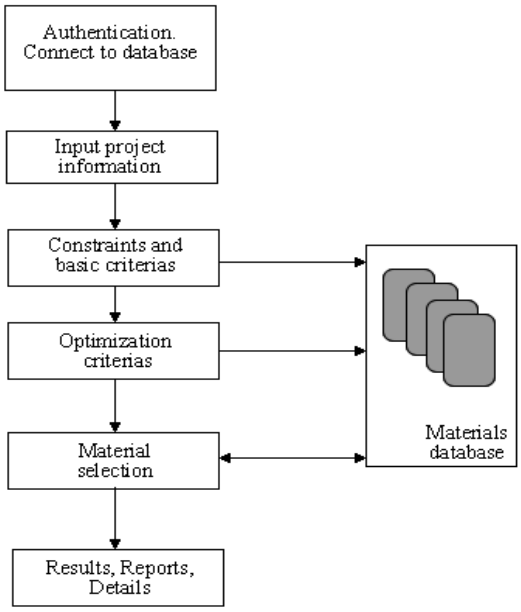


Fig. 3 Selection of materials in the WMS program

The process of selection of materials is done by defining the basic criteria, and thus choosing the right properties, relationships, and entering the boundary value. You can enter up to 10 criteria, and determine the relationship between them. From the declared database the program searches for materials meeting the specified criteria that will provide a solutions space. Then we introduce the optimization criteria determining the property, the purpose (min, max), and the priority of the given criterion (in the range of 1 - 10, where 10 is the highest priority). Using the defined data and information contained in the database the program prioritizes the space of solutions. The ranked list of materials is displayed on the screen (Fig. 4).

At this point, it is possible to view detailed information about a particular material and generate a report of the process of material selection as a PDF file.

WMS is a flexible tool for the selection of materials, the criteria are defined on the basis of the property in the database so that they can be closely matched to the needs of users.The program works with any database created using an appropriate WMS program options or compatible with the documentation supplied with the tool.

IV. Optimization Algorithm

Simple searching the database based on the basic criteria in the presence of a large number of materials and not restrictive criteria brings about choosing a variety of materials.

In recent years there has been many attempts to develop effective an way to find the optimal material from materials that meet the basic criteria. Most of these solutions are MCDM methods[3-6] - multi-criteria decision-making methods, but a statistical method was also developed [7] and a solution based on fuzzy logic [8,9].

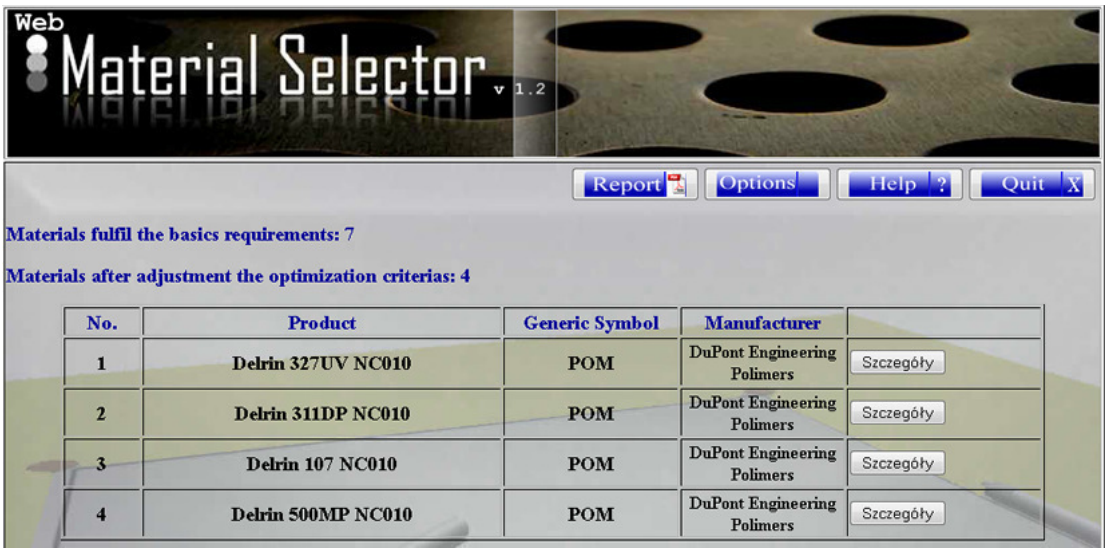


Fig. 4 The screen of WMS program - the results of searching the best material

To facilitate the selection of materials, an own MCDM method was implemented in the program - a simplified method for optimization. This method gives results similar to methods developed by other authors but computationally it is much easier what - in the presence of the large data sets - significantly reduces system's uptime

The starting point of the method is a matrix of size m x n (Fig. 5), where n is the number of materials that meet the basic criteria, and m is the number of optimization criteria. In the first step we calculate the arithmetic mean (xj) and the standard deviation (sigma j) for each criterion. Then we determine the value of the assessment factor Cji for each material.The final grade is the sum of the products of factor Cji and weights Pj corresponding to the priorities set by the user at the introduction of optimization criteria.

$$CK_i = \sum_{i=1,\,j=1}^{n,\,m} C_{ij} * P_j$$

If in the given criterion materials with smaller properties match design assumptions better (that is the objective function aims to minimize), for example, the material sought should be as light as possible so the smaller the density of the material the better it matches the assumption, then Pj of such a criterion is negative.

The last step involves sorting ascending the table of materials on the basis of final assessments and the possible rejection of the materials that meet the specified criteria the worst.

V. Summary

To improve the process of selection of materials we need solutions that fully use possibilities of information technology. One of such solutions is the Web Material Selector. This is a free, web-based application supporting the selection of materials, enabling multi-criteria search of the database of material. An additionally implemented optimization algorithm arrange according to preset criteria and limits the number of alternative solutions that should be analyzed in detail in the final phase of the materials selection. The program also makes it possible to create complete documentation of the process of material selection.

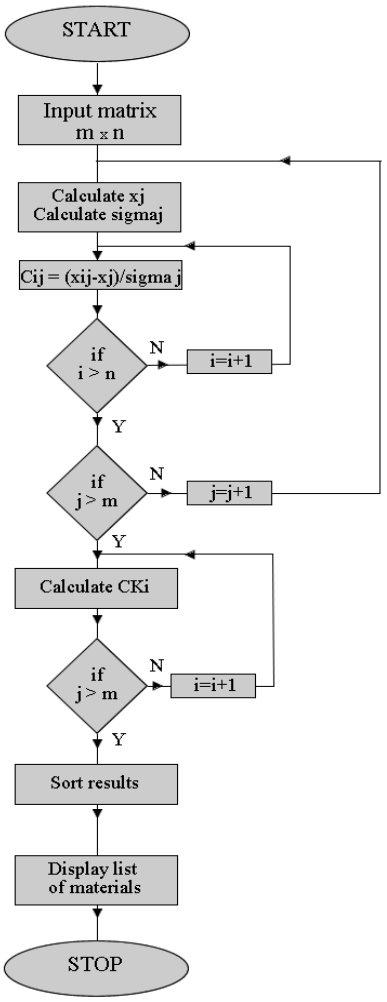


Fig. 4 Algorithm of the simplified optimization method

REFERENCES

[1] L. Dobrzański, "Zasady doboru materiałów inżynierskich z kartami charakterystyk", Wydawnictwo Politechniki Śląskiej, Gliwice 2001.
[2] M.F. Ashby, "Materials selection in mechanical design", Butterworth-Heinemann, Oxford 1999.
[3] K.M. Rajan, K. Narasimhan, "An approch to selection of material and manufacturing processes for rocket motor cases using weighted performance index", J Mat Eng Perform, vol. 11, pp. 444-449, 2002.
[4] R.V. Rao, "A material selection model using graph theory and matrix approach,, Materials Science and Enginiering, vol. 431, pp. 248-255, 2006. MCDM
[5] P. Chatterjee, V.M. Athawale, S. Chakraborty, "Selection of materials using compromise ranking and outranking methods, Materials and Design, vol. 30, pp. 4043-4053, 2009.
[6] A. Jahan, M.Y. Ismail, F. Sapuan, "Material selection based on ordinal data, Materials and Design, vol. 31, pp. 3180-3187, 2010.
[7] K. Fayazbakhsh, A. Abedian, B.D. Manshadi, R.S. Khabbaz, "Introducing a novel method for materials selection in mechanical design: Combination of non-linear normalization and mosified digital logic method", Mater Des, vol. 28, pp.8-15, 2007.
[8] T.W. Liao, A fuzzy multicriteria decision-making method for material selection, J Manuf Syst, vol. 15, pp. 1-12, 1996.
[9] S-M. Chen, "A new method for tool steel materials under fuzzy environment", Fuzzy Sets Syst, vol. 92, pp. 265-274, 1997.

3rd International Workshop on Advances in Semantic Information Retrieval

Recent advances in semantic technologies form a solid basis for a variety of methods and instruments that support information retrieval, knowledge representation, and text analysis. They influence the way and form of representing documents in the memory of computers, approaches to analyze documents, techniques to mine and retrieve knowledge. The abundance of video, voice and speech data also raises new challenging problems to information retrieval systems.

We believe that our workshop will facilitate discussion of new research results in this area, and will serve as a meeting place for researchers from all over the world. Our aim is to create an atmosphere of friendship and cooperation for everyone, interested in computational linguistics and information retrieval. The second ASIR workshop will continue to maintain high standards of quality and organization, set in the previous year. We welcome all the researchers, interested in semantics and information retrieval, to join our event.

Topics

The workshop addresses semantic information retrieval theory and important matters, related to practical Web tools. The topics and areas include, but not limited to:

- Domain-specific semantic applications.
- Evaluation methodologies for semantic search and retrieval.
- Models for document representation.
- Natural language semantic processing.
- Ontology for semantic information retrieval.
- Ontology alignment, mapping and merging.
- Query interfaces.
- Searching and ranking.
- Semantic multimedia retrieval.
- Visualization of retrieved results..

Event Chairs

Klyuev, Vitaly, University of Aizu, Japan
Mozgovoy, Maxim, University of Aizu, Japan

Program Committee

Borgo, Stefano, Laboratory for Applied Ontology, Italy
Budzynska, Katarzyna, Institute of Philosophy and Sociology of the Polish Academy of Sciences, Poland
Carrara, Massimiliano, Universita di Padova, Italy
Cybulka, Jolanta, Poznan University of Technology, Poland
Dobrynin, Vladimir, Saint Petersburg State University, Russia
Goczyła, Krzysztof, Gdansk University of Technology, Poland
Haralambous, Yannis, Institut Telecom - Telecom Bretagne, France
Homenda, Wladyslaw, Warsaw University of Technology, Poland
Jin, Qun, Waseda University, Japan
Kaczmarek, Janusz, Lódz University, Poland
Kakkonen, Tuomo, University of Eastern Finland, Finland
Kulicki, Piotr, John Paul II Catholic University of Lublin, Poland
Lai, Cristian, CRS4, Italy
Leonelli, Sabina, University of Exeter, United Kingdom
Ludwig, Simone, North Dakota State University, United States
Martinek, Jacek, Poznan University of Technology, Poland
Mirenkov, Nikolay, University of Aizu, Japan
Morshed, Ahsan, CSIRO ICT Centre, Commonwealth Scientific and Industrial Research Organisation, Australia
Mozgovoy, Maxim, University of Aizu, Japan
Nalepa, Grzegorz J., AGH University of Science and Technology, Poland
Palma, Raúl, Poznan Supercomputing and Networking Center, Poland
Piasecki, Maciej, Wroclaw University of Technology, Poland
Pyshkin, Evgeny, Saint Petersburg State Polytechnical University, Russia
Reformat, Marek, University of Alberta, Canada
Shtykh, Roman, CyberAgent Inc., Japan
Soldatova, Larisa, Brunel University, United Kingdom
Suárez de Figueroa Baonza, Mari Carmen, Ontology Engineering Group, Scool of Computer Science at Universidad Politécnica de Madrid, Spain
Tadeusiewicz, Ryszard, AGH University of Science and Technology, Poland
Trypuz, Robert, John Paul II Catholic University of Lublin, Poland
Vacura, Miroslav, University of Economics, Czech Republic
Vargiu, Eloisa, Barcelona Digital Technology Centre, Spain
Vazhenin, Alexander, University of Aizu, Japan
Wang, Haofen, Shanghai Jiao Tong University, China
Wu, Shih-Hung, Chaoyang University of Technology, Taiwan
Zadrozny, Slawomir, Systems Research Institute, Poland
Ławrynowicz, Agnieszka, Poznan University of Technology, Poland

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 11–18

Finer Investigation into Role

Riichiro Mizoguchi
Research Center for Service Science, Japan Advanced Institute of Science and Technology
mizo@jaist.ac.jp

Kouji Kozaki, Yoshinobu Kitamura
The Institute of Scientific and Industrial Research, Osaka University
{kozaki, kita}@ei.sanken.osaka-u.ac.jp

***Abstract*—This paper discusses three issues about role: (1) continuity of role playing in the case of role change, (2) role as specification and (3) vacant role. As to (1) we introduce decomposition of a role into two subroles: post subrole and constituent subrole. The latter is crucial for coping with the issue of continuity. The idea of role as a specification is proposed after detailed discussion about what a role should be. Finally, the third issue is addressed by introducing a new idea of virtual player to harmonize all the new ideas proposed in this paper. A preliminary result of formalization of the new theory is presented.**

I. Introduction

ROLE instantiation is problematic, in particular, the notion of instance of "vacant role" is unsettled. Intuitively, a vacant role is a role which is not played by any entity. What is vacancy of a role how and where does its instance exist? A relational model of roles has a difficulty to answer this question, since when a participant disappears, then the relation also disappears at the same time. We have to confess that we do not know what vacant role is very well in the first place.

In an organization, people are promoted to a higher position and then change their roles. Suppose we let John play a role, say, Deputy director, then John would lose his participation in the organization at the very time of his promotion up to Director, and hence John's participation in the organization loses continuity.

Social roles like President of a country which is clearly defined in its constitution need descriptions [11]. Then, do all kinds of roles need descriptions? Some say no, but what kinds of roles need a description is unclear.

We discuss three issues of role as fundamental concepts useful for better understanding about role.

(1) As to the first one we propose decomposition of a role into two subroles: *post subrole* and *constituent subrole* to address the issue of role change and that of the instance management of roles.
(2) The second one is about what a role is. We propose a new view of roles, that is, view of a role as a specification.
(3) Our current understanding about vacant role [15] is a bit too weak to explain the reality of a vacant role in the real world. We introduce *virtual role-holder* and *virtual player* to address this issue. It contributes to better understanding about vacant role.

This paper is organized as follows. The next section overviews our model of roles discussed in [5,14,15] to provide readers with background of the discussion of a new model of roles. Then, these three issues are discussed in Sections III, IV and V, respectively. Related work is presented in Section VI followed by concluding remarks.

II. Overview of our role theory

A. Basic model of a role

The fundamental scheme of our roles at the instance level is the following (see the lower diagram in Fig. 1):

> "*In Osaka high school, John plays teacher role-1 and thereby becomes teacher-1*"

This can be generalized to the class level (see the upper diagram in Fig. 1):

> *"In schools, there are **persons** who play **teacher roles** and thereby become **teachers**."*

By **play**, we mean "acts as", that is, it contingently acts as according to the role (role concept). By "**teacher**", we mean a class of dependent entities which roughly correspond to persons who are playing teacher roles and which are often called *qua individuals* [11]. Here, we introduce a couple of important concepts to enable finer distinctions among role-related concepts: ***role concept***, ***context***, ***role-holder***, ***potential player*** and ***role-playing thing***. In the above example, these terms are used as "*In a context, there are **players who** play **role concepts** and thereby become **role-holders***".

By **context**, we mean a class of things that should be considered as a whole. Unitary entities and relations can be a context. **Role concept** is defined as a concept which is played by some other entity within a context. So, it essentially depends on the context. By **potential player**, we mean a class of entities which are able to play an instance of a role concept. In many cases, basic concepts (natural types) can be used to denote classes of **potential players**. When an instance of potential player is playing an instance of a role concept, we call the instance a

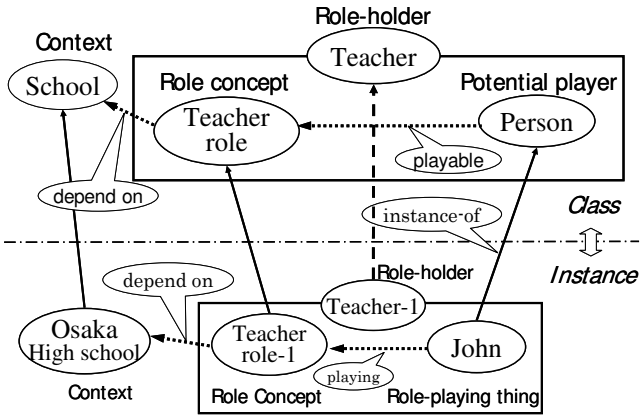


Fig. 1. Fundamental scheme of a role concept and a role-holder.

role-playing thing. In this example, we say a person can play an instance of a teacher role. In particular, John is actually playing a specific teacher role *teacher role-1*. By doing so, he is associated with the instance *teacher-1*, an individual teacher **role-holder.** A role-holder class is a super class of dependent entities like teacher-1. As such, it is neither a specialization of a potential player class (e.g., person) nor that of a role concept class (e.g., teacher role), but an abstraction of a composition of a role-playing thing and an instance of role concept, as is shown in Fig. 1, which is the heart of our Role model.

The link from Teacher-1 to Teacher is a broken arrow rather than a solid one like *instance-of* link to show the relation is not completely same as *instance-of* relation in Fig. 1. Neither our model nor Hozo tool allows people to directly generate an instance of role-holder classes because the individual role-holder as a dependent entity to be instantiated first requires an instance of a potential player class and that of a role concept class. Then, when the playing link is asserted, it virtually acquires the properties of the potential player and the role concept. This is why role-holders are dependent entities.

All the concepts introduced here are core of our role model and contain rich implications which are elaborated in [15]. The above shows that we divide the conventional notion of "Role" into two kinds: role concept and role-holder in our model. Therefore, our model of roles does not have the concept of "Role" explicitly. In particular, it is understood conventionally that a role existing at the instance level must be something being played by something, since people understand the role instantiation and the action of playing the role as happening at the same time. In contrast, in our model a role concept can exist at the instance level without being played, since it depends only on its context and not on its player. While the concept of role is the target of the ontological research on roles, at the same time, this term has been the source of confusion, since it hides the difference between role concept and role holder.

B. Hozo representation of our role model

Fig. 2 shows the correspondence between the role model and its Hozo representation. Because Hozo is based on frames, the representation is rather straightforward. Additionally, we discussed theoretical solid foundation and formal definitions of our role theory in previous work [15]. In the paper, we discussed the solid foundation of role model and presented its semantics using OWL to clarify its formal definitions. The details of role representation model using OWL and SWRL are discussed in [7]. Hozo also can export ontologies in OWL.

Let us explain Hozo's representation conventions by using the example shown in Fig. 2. In Hozo each concept defined as a class is represented in a rectangle like *School* and *Person*. Each class is defined by specifying its parts (denoted by "p/o") and/or attributes (denoted by "a/o") as slots. *School* is here defined as an entity composed of teachers and students where *teacher role* and *student role* are role concepts played by individuals specified by the rectangle at the far right, instances of *Person* in this case. As shown in Fig.2, the key idea of class definition in Hozo is that all concepts, which can theoretically be parts of something, are defined independently of the possible wholes which they can be part of, and each class as a whole is defined by specifying the roles whose parts play. In other words, all the class definitions in Hozo are reciprocal, in the sense that a whole (School) is defined in terms of its parts (Person) playing their own roles, and at the same time, the roles (teacher role) played by the parts (Person) are defined there under the context of the whole (School).

Is-a (super-sub) relations between basic concepts are represented by *is-a links* as shown Fig.3. In this example, *University* is defined as a sub concept of *School*. Sub concepts inherit all role concepts from their super concepts, and sometimes they specialize inherited role concepts to define the role concepts in their context. Fig.3 shows two role concepts *(full) professor role* and *associate professor role* which are defined in the context of *University*. They are sub concepts of *teacher role* in *School*. The relationships between these role concepts are represented by describing the super concepts on the right

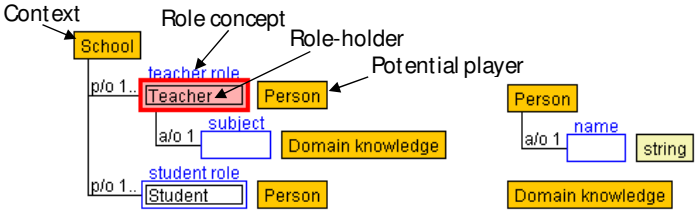


Fig. 2. Hozo's representation our role model.

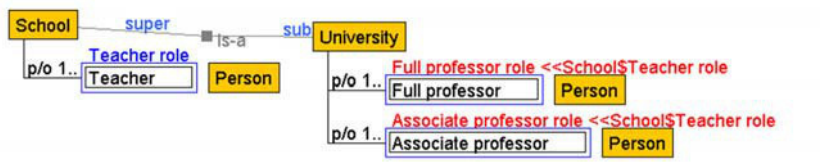


Fig. 3. Representaion of an *i-a* hierarchy of role concepts.

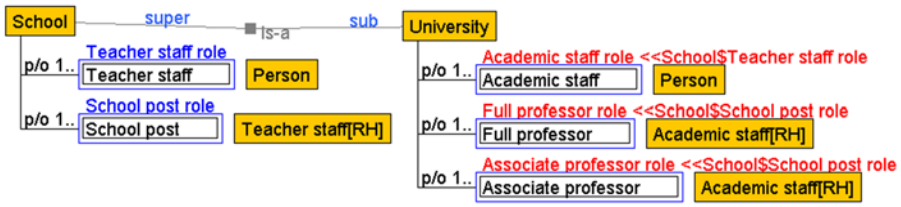


Fig. 4. Distinguishing between academic staff role and professor/associate professor role.

of role concepts with double angles "<<" as shown in Fig. 3. It means the hierarchy of role concepts is analogue to the hierarchy of basic concepts because all role concepts are defined within the basic concepts as their contexts.

III. Introduction of Post Role and Constituent Role [7]

A. Motivating example

Let us consider the behavior of an instance of roles in the case of promotion in a university. We take an *associate professor role* and a *full professor role* as examples which are defined in the context of a university. In this example, an associate professor John in Osaka University is represented as follows:

John plays an instance of the associate professor role and thereby becomes an associate professor (role-holder).

Now, we assume a case where John is promoted from associate professor to full professor. It means that John stops to play the instance of the *associate professor role*, then plays the instance of the *full professor role*. In this example, when John plays the instance of associate professor role or full professor role, he also plays *teacher role* in the university at the same time. Because <both full professor role and associate professor role *is-a* teacher role> as shown Fig.3, the semantics of *is-a* relation tells us that John stops to play the instance of teacher role at the very moment when he changes the role to play. This is because one has to stop to play the current role when he/she starts to play the new role. In other words, the continuity of playing the *teacher role* is lost. Obviously, this example model does not capture the behavior of the instances in the real world accurately. In the real world, John must have been the same teacher both when he is the associate professor and the full professor.

B. An informal solution

The problem discussed in the previous section is caused by the fact that the player of the role is John himself rather than an academic staff role-holder, so that he/she comes back to himself when he/she stops to play a role. If the player of associate professor role was an academic staff role-holder rather than John, then the player of the associated professor role can stay being an academic staff member when he/she stops to play the role. This can be explained from another perspective. That is, this difficulty is caused by the confusion between academic staff role and associate professor roles. The former means that the player of the role is a member of the university (as an academic staff member), and the latter means what the player is expected to perform in the university. The above motivating example and its model fails to capture the separation between an associate professor role which is a composite of the post subrole and a staff member subrole.

As it is already apparent, this problem can be solved by distinguishing between *academic staff role* and *professor/associate professor role*, and using *academic staff role-holder* rather than *person* as the player of the associate professor role as shown in Fig.4.

Firstly, when John is an associate professor of Osaka University, he plays an instance of an *academic staff role* in Osaka University and thereby becomes an *academic staff role-holder* (referred to as RHi). Then, RHi

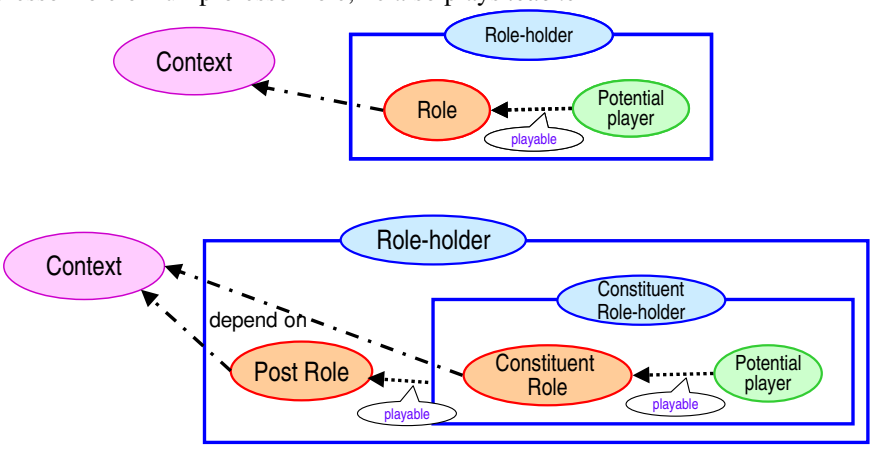


Fig. 5 Extension of the framework.

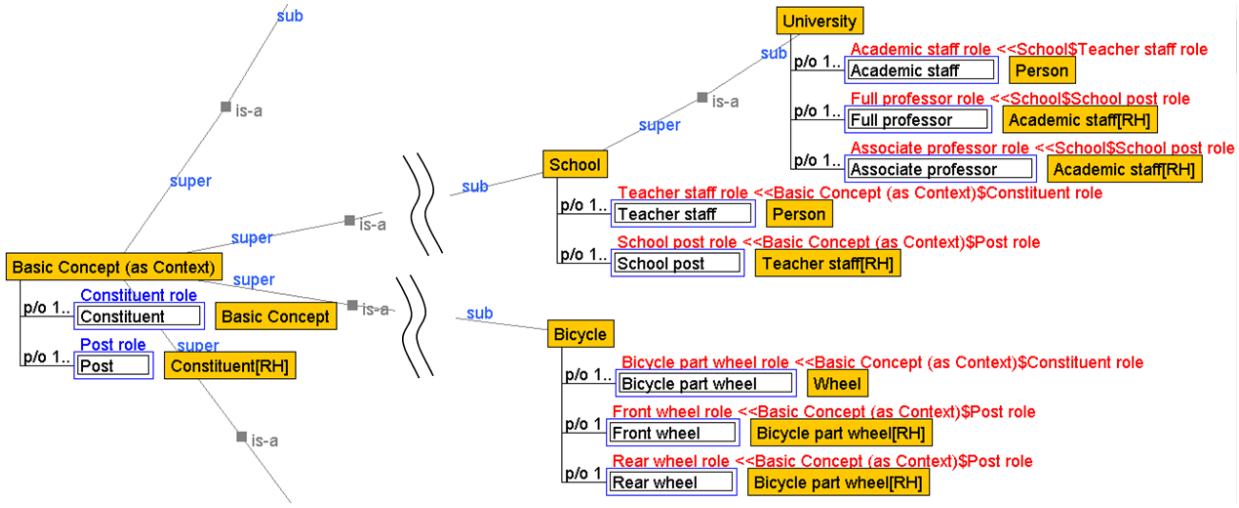


Fig. 6. Class definitions of Constituent role and Part role

plays an instance of *associate professor role* and thereby becomes the associate professor (role-holder). Next, when John is promoted from associate professor to full professor, he (RHi) stops to play the *associate professor role* and plays a *full professor role* while he does not stop to play the *academic staff role*, since he comes back to not an ordinary *person* but RHi which is a role-holder of *academic staff role* when he stops to play the *associate professor role*. This problem is thus resolved successfully.

C. Constituent subrole and post subrole [7]

We generalize the notion of subroles introduced in the above. Fig.5 shows the original framework of our role theory (the upper one) and the extended one (the lower one). In the original framework, a role is simply played by a player. In the new framework, on the other hand, the post subrole is played by constituent subrole-holder which is made of constituent subrole played by a usual player. The original structure is kept and the new one is composed of the original one in a nested structure. As to the structure, we can say that the potential player of the original framework is extended and replaced by constituent subrole-holder.

The ordinary role is decomposed into two subroles: post subrole and constituent subrole. The post subrole represents the main content of the original role and hence is responsible for specifying what job/task is expected to be performed by the player. On the other hand, the constituent subrole is introduced just to represent participation of the player in the context to create theoretically appropriate players of corresponding post subroles. It also contributes to the context by being its component/participant.

Note that this idea of separation of a role into two subroles applies to all kinds of roles defined in the context of an entity as a whole in theory. The utility is apparent in the case of a company in which people often move across departments and get promoted to an upper post. In the example discussed in *B teaching staff role* in a school and *academic staff role* in a university are *constituent subroles*[1] which means their players are participating in the organization (school and university). And, *school post subrole* in a school, *associate professor role* and *full professor role* are *post subroles* which mean roles (posts) are performed in the organization by role-holders of the *constituent subroles*.

We can find the same issue concerning artifacts. For example, suppose that a front wheel of a bike is replaced by a rear wheel in the same bike. The replaced wheels change its role while keeping participation in the context. To capture this case appropriately, we should define *bike part role* as a constituent subrole and some post subroles (e.g. *front wheel role*, *rear wheel role*) played by role-holders of the *bike part roles*. Fig.6 shows examples of them discussed in this section.

D. Instance of role concept

Let us take a teacher John's case in a high school to investigate the role instance problem as an example. Imagine a situation where, after John retired, a new teacher was hired. The question is whether or not the teacher role instance of John was playing is the same as what the new person plays. If there is only one role instance like prime minister role, the answer is trivial. So, the answer of the John's case would be the same role is played by the new comer. However, if two teachers retired at the same time, and two new teachers were hired, then there is no way to identify which role instance is

[1] We can introduce a superrole of them such as member (of the organization) role whose subroles include constituent roles for clerks or students if it is needed. That is, we can define several kinds of constituent roles according to their contexts [7].

played by which. If we think about the case where a student Mike comes back to the same university after he once graduated from it, it is not clear if he plays the same student role as the former one or not.

Our goal here is not to determine in which cases the role instances are reused. We intend to come up with a flexible role instantiation model which can cope with either case so that users can use it as they like. Interestingly, the two roles: *constituent subroles* and *post subroles* do a significant contribution to coping with this issue. In fact, post subrole is created according to each kind of role concepts independently of players. For example, when there are three English and two math teacher roles in a school, one English teacher post subrole and one math teacher post subrole instances are created. On the other hand, constituent subrole instances are created for each player instance. Each post subrole instance is shared by multiple constituent subrole instances and is reused by them. Such roles as prime minister, manager, president and teacher roles attract people's focus on their post subroles, and hence they seem to be reused by multiple players. Even in the case of restaurant guest roles, we create one guest post subrole for a restaurant and reuse it by multiple players (guests) because all the players should be equally treated by restaurant staff as guests.

We know the term "guest post subrole" sounds strange comparing with manager post and president post subroles. However, we believe it is just a terminological problem and does not degrade the idea of post subroles which is applicable to any role.

In the case of constituent subrole, on the other hand, its instance is created whenever a new playing event happens independently of the player is the same or not. So, even if the same player has played the same role multiple times, each instance of the constituent subrole should be different. When people concentrate their focus on constituent subrole, then the role seems to require new instances whenever a new player participates in the context.

When we investigate properties role instances should have, we notice important information is lacking, that is, the information of when the player plays the role. Candidates which can possess such information are the following three: player itself, post subrole and constituent subrole[2]. If we put it to players, then we lose the information when he/she left. Post subroles are shared by multiple constituent role-holders, so it is not appropriate to hold such player-dependent information. *Constituent subroles* are created for each playing event, so it would be the best to let each constituent subrole to possess such information.

In summary, when a context instance is created, all the instances of post subroles corresponding each role kind are created. All post subroles are played by the corresponding constituent subrole-holders which are created whenever a player participates in the context to play the role. All the players participated in the context must play the constituents subroles first and thereby become a constituent subrole-holder to play the post subrole corresponding to the original role. Constituent subroles thus represent participation events of players as well as generic components of the context. The problem of instance generation of roles has been resolved by the introduction of these two subroles. As already explained, post subroles are created according to the corresponding role independently of the existence of players, while each constituent subrole is created according to each participation of a player. Whether the constituent subrole instance should be kept as a vacant role after the player has gone is determined by the necessity of keeping the participation history.

IV. WHAT IS A ROLE?

A role is something to be played. When a person plays a teacher role, he/she must fulfill the requirement the school expects. If it is a math teacher, he/she must be qualified with math teaching. A role is thus a required specification expected to be fulfilled by a player. This view suggests that what a player has when he/she plays the role is capacity to fulfill the role specification. The players realize their capacity to fulfill the specification required to play the role. Therefore, the role is external to the player. What is internal to the player is capacity to realize/fulfill the role as specification. The decomposition of a role into *post subrole* and *constituent subrole* also contributes to the idea of "a role is a specification" because *post subrole* is what a player fulfills after removing *constituent subrole* from the original role and corresponds to what functionality/responsibility for the player to perform/have

Let us investigate the idea that a role is a specification further for each of the cases where roles defined in the context of an object and an occurrent.

(1) Object-dependent roles

Every part of an object contributes to making it as a whole. In the case of functional objects which have functional parts, all parts play a role to help the whole realize its function by performing respective functions. An engine of a car plays a *power source role* by generating torque and a blood vessel of a circulatory system of a human body plays a *conduit role* by allowing blood to flow, etc. Such roles that are associated with parts specify what function(s) must be performed. In particular, parts/components of an artifact are designed as specification which must be satisfied by real parts/components which are installed at the right place with right qualities to realize the real artifact.

(2) Occurrent-dependent roles

[2] The role-holder is the strongest candidates, of course. However, it is a composite of these three, so it is excluded.

Any occurrent necessarily needs participants. Objects can participate in an occurrent. Typical participants are *actor* and *operand* either one of which is a role. Examples of actor role include *runner, singer, speaker, etc.* Actor as a role subsumes these three roles each of which specifies what it must be, that is, *runner role* specifies that its player must participate in a running process as a doer.

The discussion thus far suggests us a temporary answer to the question: what is a role. A role individual exists as a specification as far as its context exists independently of if it is played or not and it exists not in the player but in its context. Players have capacity to fulfill the specification of the role to play. This idea fits the notion of *play* very well. Borrowing Guarino's claim[3], "to play a piece of music essentially means to play an instrument using a musical score which is a specification".

V. VACANT ROLE AND VIRTUAL ROLE-HOLDER

A. Vacant role

At first glance, a role seems to be a property rather than a relation. Although it is true that a role is not a relation, the problem is that a role can be a property only when it is played by an entity. Note here that it is a property of the player rather than so is the role itself intrinsically. This suggests that if a role is a property or not is not a simple problem which is caused by the fact that a role is something to play, and hence there is a situation where a role is not played by anything. Thus, there appears a serious issue: **what a role is when it is vacant**.

To answer this question, we need to investigate what is the instance of a role in the first place: If it is a property, then whose property is it, the player or another? There are two states in roles: being played and not played. If a vacant role, which is not played by any entity, were a property of the player, then the issue would have been much easier as that a role is a property of a player in any case. However, the vacant role cannot exist in the player, but exists in the context where it is defined, which is outside of the player. For example, a teacher role is defined in a school and does not exist in the player when it is not played. Teacher roles are key components of a school which is composed of other roles such as student roles and lecture room roles as well as teacher roles.

How much vacant is a vacant role? We claimed a role is a specification in IV. Then, what is an instance of role as a specification? If it is just a specification, then a vacant teacher role of a school does not make sense well and it does not explain the reality. How about the Japanese Prime Minister role? We do not think it is just a specification when it is vacant. To investigate this issue, here are two stories[4].

> ***Story 1:*** *A letter has been sent to the President of a country from Mr. A who does not know Mr. President resigned a few days ago. The letter is valid and received successfully.*
>
> ***Story 2:*** *A set of cords connected to the right headlight of a car is said to be connected to the right headlight even when the light bulb is taken off.*

Both stories say that a vacant role is not just vacant and is something more realistic than mere specification. In Story 1, even if President cannot answer the letter, it is validly received and the next real President will be able to answer it. In Story 2, engineers talk about the right headlight independently of if the bulb is there or not, though the vacant right headlight cannot light. In short, these vacant roles are almost real entities. So, we call such a vacant role *virtual entity*. Then, the next issue is how to concretize such an observation.

B. Virtual player

This idea suggests a possibility to introduce *virtual role-holder* of constituent subrole to let it play the post subrole. It is anyway effective to let anything play a post role for making a mere specification more realistic. But, we have no real player. Then, the only possible way is to create a *virtual player* who is empty but to let it play a constituent subrole to produce a *virtual constituent subrole-holder*.

Then, what is a virtual player? It is made by removing all the values of specified properties derived from its material part, so that if such an immaterial player already there and play the *constituent subrole* to become a *virtual constituent subrole-holder,* it would resolve our problem. Note, in our case, that we have two virtual role-holders: *virtual constituent subrole-holder* held by a *virtual player* and *virtual post subrole-holder* held by *virtual constituent subrole-holder*.

Accordingly, we modify the idea of role-holder to adapt to the notion of *virtual player*. In the current theory, role-holder instance cannot exist without a player, while, in the new theory, it can exist as a *virtual role-holder* by considering the role is played by a virtual player. When a role is played by a real entity, the corresponding *role-holder* comes to exist as a real entity. Thus, the change of our theory has been made on the treatment of vacant cases. Although the theory needs more adaptation, it is omitted due to space limitation.

We define a ***virtual role-holder*** as a role-holder in the case of the player is a ***virtual player***. See Fig. 7 which shows *virtual teacher-1* is made when a *teacher post subrole* is played by a *virtual constituent-1* which is made when a *constituent subrole* is played by a virtual player,

[3] This captures only meaning of his claim in personal discussion made in the context of EuJoint project: Project n. 247503.

[4] These two stories are given by Nicola Guarino while discussion with him and his colleagues in the context of collaborative research supported by EuJoint: Project n. 247503.

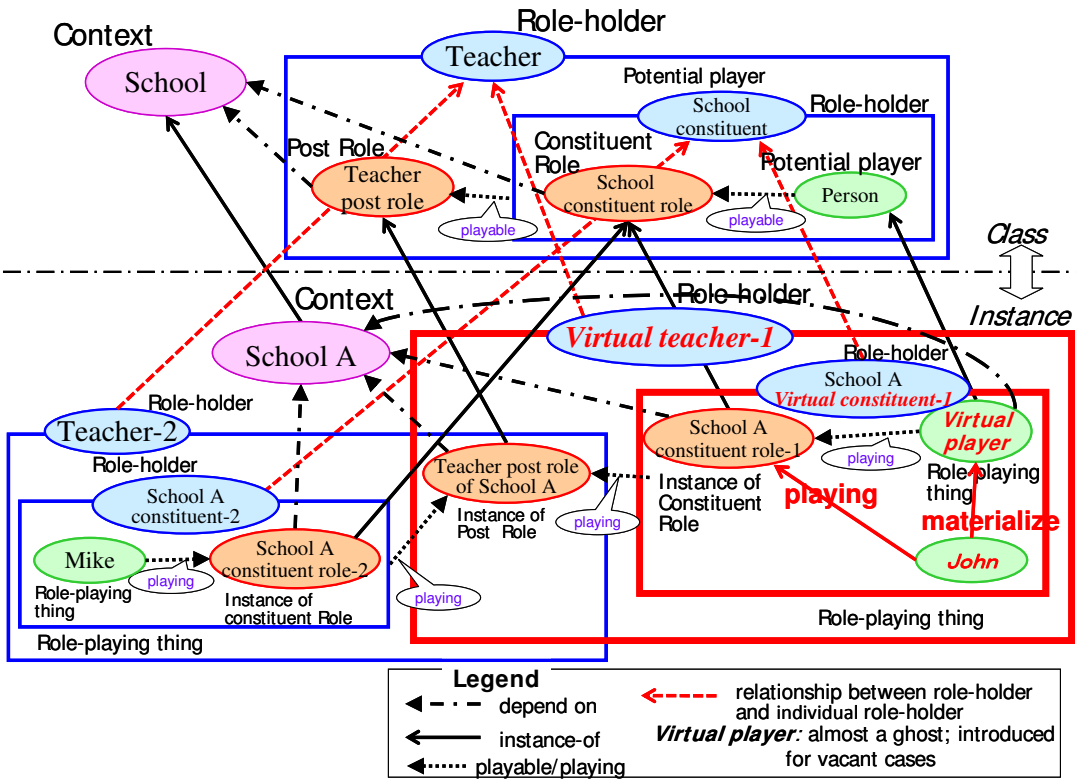


Fig.7 Virtual teacher and virtual constituent (see the two red rectangles).
All circles represent a set of properties.

that is, when no real player exists. When John participates in the context (school A) and plays the teacher role, he **materializes** the virtual player by giving all his property values to it. Then, the next issue is what is the virtual player materialized by John? Are there two different individuals at the same time: John and a virtual player materialized by John? Because the relation between the two is *constituted-of*, the answer is yes, and both of them exist by sharing the same region in the spatiotemporal space like a vase. It is analogous to the case of a vase constituted of an amount of clay. After making a vase using the amount of clay, both keep their identities unless the vase has been destroyed. They co-exist in the same region of the spatiotemporal space. Thus, when a real player participates in the context, the *virtual role-holder* continues to exist even after a real player has participated in, that is, it exists independently of it is filled or not.

In this way, a letter sent to the President who recently resigned is received by the *virtual President role-holder*, and the cords to the right headlight are successfully connected to the *virtual right headlight role-holder*.

The introduction of virtual player does not have to be done to all kinds of roles. It should be dependent on the kinds of roles. It should be introduced to such roles that have a unique individual or pre-fixed number of role individuals. On the other hand, roles whose number of role instances is indefinite like guest role of a restaurant do not need any virtual player.

C. Toward formalization

Although it is very preliminary, the following is a first step towards formalization of role playing and vacancy of roles mainly for cases where virtual players are necessary.

Primitives: *context(x), playing(x, r, rh){x is playing a role r and becomes a role-holder rh}, post-sub-role(x), real(x), virtual(x), constituent-sub-role(x), inherit(x, y) {x inherits all properties of y}, part(x, y) {x is a part of y}, constituted-of(x, y, vo) {x is constituted of y using vo as a template}*

The original model:

role-concept(r, x) => context(x).

role-holder(rh, r) => ∃x,y; context(y), role-concept(r, y), playing(x, r, rh), inherit(rh, x), inherit(rh, r).

role-holder(rh, r) <=> ∃x; playing(x, r, rh).

vacant(r) => Not (∃x; playing(x, r, rh)).

The new (nested) model:

role-concept(r, w) => ∃x,y,z; context(w), part(x, r), part(y, r), post-sub-role(x), constituent-sub-role(y). {x and y are non-overlapping and exhaustive}

c-role-holder(rh, y) =>. x, csr; context(y), role-concept(r, y), part(csr, r), constituent-sub-role(csr), real-playing(x, csr, rh).
p-role-holder(rh, r) =>. x,y, psr; c-role-holder(rh1, r), part(psr, r), post-sub-role(psr), playing(rh1, r, rh). {a psr can be shared by multiple csr's}
vacant(r) <=> Not (x; real(x), playing(x, r, rh)).
v-c-role-holder(rh, y) =>. x, csr; context(y), role-concept(r, y), part(csr, r), constituent-sub-role(csr), virtual-playing(x, csr, rh).
v-p-role-holder(rh, r) =>. x,y, psr; v-c-role-holder(rh1, r), part(psr, r), post-sub-role(psr), playing(rh1, r, rh).
real-playing(x, r, rh) =>. y, vo; real(y), virtual(vo), constituted-of(x, y, vo), playing(x, r, rh), virtual-playing(vo, r, rh).
virtual-playing(x, r, rh) => virtual(x), playing(x, r, rh).
context(x) =>. rh, rh1; v-c-role-holder(rh1, x), v-p-role-holder(rh, r), playing(rh1, r, rh).

VI. Related Work

As far as the authors know, there are few papers which tackle the main topics discussed in this paper except [4]. Existing literature on roles scatters in several areas of ontology engineering [3][11][9,10], database model[16], software engineering [1], and agent systems [2].

Guarino discussed vacant roles and claims that they are more real than the current understanding about them in the role community in his recent paper [4]. We share the same issue with him. He nicely analyzes ontological status of related entities and concepts such as system component, replaceability, functional roles in the context of vacant role (missing player). While he does not present a solution to this issue, we have proposed a solution to the problem in this paper.

VII. Concluding Remarks

We have proposed (1) post and constituent subroles to cope with continuity of role-playing in the case of change of roles, (2) a new view of role as a specification and (3) virtual player. These contribute to refinement of role theory. Future work should be completion of the formalization of the new theory.

Acknowledgment

The authors are grateful for Nicola Guarino, Laure Vieu and Claudio Masolo for their comments in the context of collaborative research supported by EuJoint: Project n. 247503. This material is partially based on research sponsored by the Air Force Research Laboratory, under agreement number FA2386-13-1-4047**.** The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

Disclaimer:

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

References

[1] Chernuchin, D. & Dittrich, G. (2005): Role Types and their Dependencies as Components of Natural Types, Proceedings of the 2005 AAAI Fall Symposium 'Roles, an Interdisciplinary Perspective: Ontologies, Languages, and Multiagent Systems' Technical Report FS-05-08, Arlington, Virginia, pp.39-46.

[2] Colman,A., & Han, J. (2007). Roles, players and adaptable organizations, Applied Ontology **2** , 105-126.

[3] Guarino, N. (1992): Concepts, attributes and arbitrary relations: Some linguistic and ontological criteria for structuring knowledge bases. Data & Knowledge Engineering **8**(2), 249–261.

[4] Guarino, N. (2013): Artefactual systems, missing components, and replaceability, Chapter 11, Artefact Kinds: Ontological and the Human-Made World. Syntese Library, Springer (to appear).

[5] Kozaki, K. and et al. (2002). An Environment for Building/Using Ontologies Based on a Fundamental Consideration of "Role" and "Relationship", In Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW2002), 213-218.

[6] K. Kozaki, et. al. (2007) Role Representation Model Using OWL and SWRL, In Proc. of 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies, Berlin, July 30-31, pp.39-46.

[7] K. Kozaki, et. al. (2008) Instance Management Problems in the Role Model of Hozo, Proc. of the 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI-08), LNAI 5351, pp. 614-625, Dec. 15-19, Hanoi, Vietnam,

[8] Kozaki, K., Kitamura, Y. and Mizoguchi, R. (2012). A Model of Derived Roles, In Proc. of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW2012), pp.227-236.

[9] Loebe, F. (2005). Abstract vs. Social Roles - A Refined Top-Level Ontological Analysis. Papers from the AAAI Fall Symposium Technical Report FS-05-08, pp. 93-100, USA.

[10] Loebe, F. (2007). Abstract vs. social roles - towards theoretical account of roles, Applied Ontology **2**, 127–258.

[11] Masolo, C., Vieu, L. and et al. (2004). Social roles and their descriptions, In Proc. of the 9th International Conference on the Principles of Knowledge Representation and Reasoning (KR), , 267–277.

[12] Masolo, C, Guizzardi, G., and et al. (2005). Relational roles and qua-individuals, AAAI Fall Symposium on Roles, an interdisciplinary perspective, November 3-6, 2005, Hyatt Crystal City, Arlington, Virginia.

[13] Masolo, C., Vieu, L., Kitamura, Y., Kozaki, K. and Mizoguchi, R. (2011). The Counting Problem in the Light of Role Kinds, InProc. Of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, 76-82.

[14] Mizoguchi, R., Kozaki, K., Sano, T. & Kitamura, Y. (2000). Construction and Deployment of a Plant Ontology. In R. Dieng and O. Corby (Eds.), Proc. of the 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW2000), Juan-les-Pins, France, LNCS 1937, 113-128.

[15] Mizoguchi, R., Sunagawa, E., Kozaki, K. and Y. Kitamura. (2007). The model of roles within an ontology development tool: Hozo. J. of Applied Ontology, **2**, 159–179.

[16] Steimann, F. (2007). The role data model revisited, J. of Applied Ontology, **2** (2007), 89-103.

Computer Science & Network Systems

CSNS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to more technical aspects of computer science and related disciplines. The CSNS area spans themes ranging from hardware issues close to the discipline of computer engineering via software issues tackled by the theory and applications of computer science and to communications issues of interest to distributed and network systems. Events that constitute CSNS are:

- CANA'2013 - Computer Aspects of Numerical Algorithms
- MMAP'2013 - International Symposium on Multimedia Applications and Processing

Computer Aspects of Numerical Algorithms

Numerical algorithms are widely used by scientists engaged in various areas. There is a special need of highly efficient and easy-to-use scalable tools for solving large scale problems. The workshop is devoted to numerical algorithms with the particular attention to the latest scientific trends in this area and to problems related to implementation of libraries of efficient numerical algorithms. The goal of the workshop is meeting of researchers from various institutes and exchanging of their experience, and integrations of scientific centers.

Topics

- Parallel numerical algorithms
- Novel data formats for dense and sparse matrices
- Libraries for numerical computations
- Numerical algorithms testing and benchmarking
- Analysis of rounding errors of numerical algorithms
- Languages, tools and environments for programming numerical algorithms
- Numerical algorithms on GPUs
- Paradigms of programming numerical algorithms
- Contemporary computer architectures
- Heterogeneous numerical algorithms
- Applications of numerical algorithms in science and technology

Event Chairs

Bylina, Beata, Maria Curie-Sklodowska University, Poland

Bylina, Jarosław, Maria Curie-Sklodowska University, Poland

Stpiczyński, Przemysław, Maria Curie-Sklodowska University, Poland

Program Committee

Amodio, Pierluigi, Università di Bari, Italy

Anastassi, Zacharias, Qatar University, Qatar

Brugnano, Luigi, Universita' di Firenze, Italy

Czachorski, Tadeusz, Poland

Filippone, Salvatore, University Rome Tor Vergata, Italy

Gansterer, Wilfried, University of Vienna, Austria

Georgiev, Krassimir, IICT - BAS, Bulgaria

Gimenez, Domingo, University of Murcia, Spain

Gravvanis, George, Democritus University of Thrace, Greece

Kierzenka, Jacek, United States

Knottenbelt, William, Imperial College London, United Kingdom

Lirkov, Ivan, Institute of Information and Communication Technologies , Bulgarian Academy of Sciences, Bulgaria

Maksimov, Vyacheslav, Institute of Mathematics and Mechanics, Russia

Marowka, Ami, Bar-Ilan University, Israel

Mauro, Francaviglia, University of Torino, Italy

Petcu, Dana, West University of Timisoara, Romania

Pultarova, Ivana, Czech Technical University in Prague, Czech Republic

Rybka, Piotr, The University of Warsaw, Poland

Satco, Bianca-Renata, Stefan cel Mare University of Suceava, Romania

Sedukhin, Stanislav, The University of Aizu, Japan

Sergeichuk, Vladimir, Institute of Mathematics of NAS of Ukraine, Ukraine

Srinivasan, Natesan, Indian Institute of Technology, India

Szajowski, Krzysztof, Institute of Mathematics and Computer Science, Poland

Szyld, Daniel, Temple University, United States

Tokarzewski, Jerzy, Warsaw University of Technology, Poland

Tudruj, Marek, Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland

Ustimenko, Vasyl, Marie Curie-Sklodowska University, Poland

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 23–26

Parallel algorithms for adaptive time-stepping in radiofrequency liver ablation simulation: implementation on an IBM Blue Gene/P computer

K. Georgiev, N. Kosturski, S. Margenov, Y. Vutov
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences

***Abstract*—This work concerns the mathematical modeling and computer simulations of the heat transfer process. The core is solving the time-dependent partial differential equation of parabolic type. Instead of a uniform discretization of the considered time interval, an adaptive time-stepping procedure is applied in an effort to decrease the simulation time. The procedure is based on the local comparison of the Crank Nicholson and backward Euler approximations. Computer simulation on geometry obtained from a magnetic resonance imaging (MRI) scan of the patient is performed. Results of some preliminary numerical experiments performed on a selected test problems are presented and discussed**

I. Introduction

THE minimally invasive treatment called radio-frequency ablation (RFA) guided by imaging techniques, the doctor inserts a thin needle through the skin and into the tumor, (see Fig. 1, [11]). High-frequency electrical energy delivered through this needle heats and destroys the tumor. The circuit is closed with a ground pad applied to the patient's skin.

The right procedure parameters are very important for the successful killing of all of the tumor cells with minimal damage on the non-tumor cells.

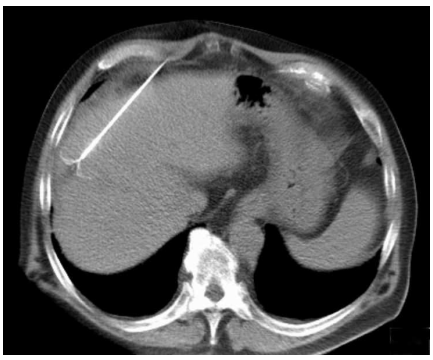


Fig. 1. CT Scan, Showing Radio-Frequency Ablation of a Liver Lesion

Computer simulation on geometry obtained from a magnetic resonance imaging (MRI) scan of the patient is performed.

In this work, an adaptive time stepping algorithm is applied to the simulation in order to reduce the computational time.

The rest of the paper is organized as follows. The mathematical and numerical models are shortly presented in Section 2. The adaptive time-stepping algorithm is described in Section 3. Section 4 is devoted to the computer simulations and analysis of the results obtained on an IBM Blue Gene/P supercomputer. Finally, some concluding remarks can be found in Section 5.

II. The mathematical and numerical models

The bio-heat time-dependent partial differential equation [4], [5] is the governing equation describing the RFA process. It can be presented as follows:

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot k \nabla T + J \cdot E - \alpha\, h_{\mathrm{B}}\,(T - T_{\mathrm{B}}), \qquad (1)$$

where the thermal energy arising from the current flow is described by $J \cdot E$ in (1) and $\alpha\, h_{\mathrm{B}}\,(T - T_{\mathrm{B}})$ accounts for the heat loss due to blood perfusion in the capillaries. The heat produced from metabolic functions of the liver is neglected. The initial and boundary conditions which are used in this approach are as follows:

$$T = 37^{\circ}\mathrm{C} \quad \text{when } t = 0 \text{ at } \Omega, \qquad (2a)$$

$$T = 37^{\circ}\mathrm{C} \quad \text{when } t \geq 0 \text{ at } \partial\Omega, \qquad (2b)$$

$$-k\frac{\partial T}{\partial n} = \alpha(T - T_{\mathrm{B}}) \quad \text{when } t \geq 0 \text{ at } \Gamma_{\mathrm{R}} \qquad (2c)$$

The notations which are used in (1) and (2) are given bellow:

- Ω – the entire domain of the model;
- $\partial\Omega$ – the boundary of the domain;
- Γ_r – the boundary of the blood vessel;
- ρ – density [kg/m^3];
- c – specific heat [J/kg K];
- k – thermal conductivity [W/m K];
- J – current density [A/m];
- E – electric field intensity [V/m];
- t – time [s];
- T – temperature [K];
- T_{B} – blood temperature (37°C);
- w_{B} – blood perfusion coefficient[s^{-1}];
- $h_{\mathrm{B}} = \rho_{\mathrm{B}} c_{\mathrm{B}} w_{\mathrm{B}}$ – convective heat transfer coefficient accounting for the blood perfusion in the model;
- α – tissue state coefficient;
- n – the outward-pointing normal vector of the boundary.

The cumulative damage integral $\Psi(t)$ is used as a measure of ablated region [1], [9]:

$$\Psi(t) = \ln\left(\frac{c(0)}{c(t)}\right) = A\int e^{-\frac{\Delta E}{RT(t)}}dt, \tag{3}$$

where $c(t)$ is the concentration of living cells, R is the universal gas constant, A is the "frequency" factor for the kinetic expression [s^{-1}], and ΔE is the activation energy for the irreversible damage reaction [J mol^{-1}]. The values used $A = 7.39 \times 10^{39} s^{-1}$ and $\Delta E = 2.577 \times 10^5$J mol^{-1} are taken from [1]. Tissue damage $\Psi(t) = 4.6$ corresponds to 99% probability of cell death. The value of $\Psi(t) = 1$, corresponding to 63% probability of cell death is significant, because at this point the tissue coagulation first occurs and blood perfusion stops.

The tissue state coefficient α is expressed as

$$\alpha(t) = \begin{cases} e^{-\Psi(t)} \text{ if } \Psi(t) < 1, \\ 0 \text{ if } \Psi(t) \geq 1. \end{cases}$$

In the presented algorithm the bio-heat problem (1) is solved in two steps (see [9] for more details):

1) Finding the heat source $J \cdot E$ using that: (a) $E = -\nabla V$ (V is the electric potential in the computational domain Ω), and (b) $J = \sigma E$, where σ is the electric conductivity [S/m];
2) Finding the temperature T by solving the heat transfer equation (1) using the heat source $J \cdot E$ obtained in the first step.

For the numerical solution of (1) the finite element method in space is used ([7]). *Linear conforming tetrahedral elements* are used in this study. They are directly defined on the elements of the used unstructured mesh (see Fig. 2). An *algebraic multigrid* (AMG) preconditioner is used [3]. The time derivative is discretized via finite differences and the both the *backward Euler* and the *Crank-Nicholson* schemes are used ([8]).

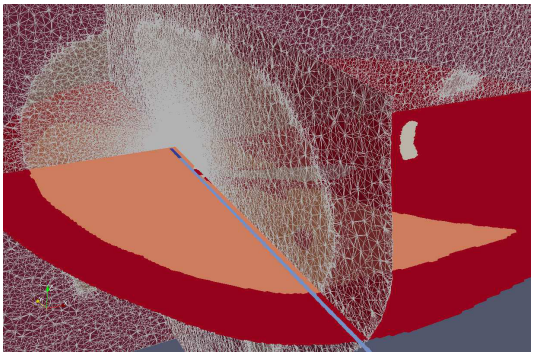


Fig. 2. Inserted RF Probe and the Finite Element Mesh

Let the matrices K and M be the stiffness and mass matrices from the finite element discretization of (1):

$$K = \left[\int_\Omega k\nabla\Phi_i \cdot \nabla\Phi_j d\mathbf{x}\right]_{i,j=1}^{N},$$

$$M = \left[\int_\Omega \rho c\Phi_i\Phi_j d\mathbf{x}\right]_{i,j=1}^{N}.$$

Let us also denote with Ω_B the subdomain of Ω where we account for the blood perfusion (the liver tissue) and with M_B the matrix

$$M_\mathrm{B} = \left[\int_\Omega \delta_\mathrm{B} h_\mathrm{B}\Phi_i\Phi_j d\mathbf{x}\right]_{i,j=1}^{N},$$

where

$$\delta_\mathrm{B}(x) = \begin{cases} \alpha \text{ for } x \in \Omega_\mathrm{B}, \\ 0 \text{ for } x \in \Omega \setminus \Omega_\mathrm{B}. \end{cases}$$

The influence of the Robin boundary conditions given in (2c) and the electric field intensity is presented by:

$$M_\mathrm{R} = \left[\int_{\Gamma_\mathrm{R}} \alpha\Phi_i\Phi_j d\mathbf{x}\right]_{i,j=1}^{N}, \tag{4}$$

and

$$F = \left[\int_\Omega JE\Phi_i\Phi_j d\mathbf{x}\right]_{i,j=1}^{N}, \tag{5}$$

Than, the spatially discretized parabolic equation (1) can be written in matrix form as:

$$M\frac{\partial T}{\partial t} + (K + M_\mathrm{B} + M_\mathrm{R})T = F + M_\mathrm{B}T_\mathrm{B} + M_\mathrm{R}T_\mathrm{B}. \tag{6}$$

III. Adaptive Time-Stepping Algorithm

To ensure accuracy and not waste computational effort, it is important to adapt the time steps to the behavior of the solution.

The time discretization for both backward Euler method and the Crank-Nicolson one can be written in the form

$$\begin{aligned}(M+&\tau^n\theta(K + M_\mathrm{B} + M_\mathrm{R}))T^{n+1} \\ =&(M - \tau^n(1-\theta)(K + M_\mathrm{B} + M_\mathrm{R}))T^n \\ &+ (\tau^n\theta + \tau^n(1-\theta))(F + M_\mathrm{B}T_\mathrm{B} + M_\mathrm{R}T_\mathrm{B}),\end{aligned} \tag{7}$$

where the current (n-th) time-step is denoted with τ^n, the unknown solution at the next time step – with T^{n+1}, and the solution at the current time step – with T^n. If we set the parameter $\theta = 1$, (7) gives a system for the backward Euler discretization. When $\theta = 0.5$ (7) becomes Crank-Nicolson one. The solution of the linear system (7) with $\theta = 1$ and $\theta = 0.5$ gives us T_BE and T_CN respectively.

A suitable adaptive time-stepping procedure is based on a local comparison of the backward Euler (T_BE) and Crank-Nicolson (T_CN) approximations for the current timestep, and is controlled by the ratio

$$\eta = \frac{\|T_\mathrm{CN} - T_\mathrm{BE}\|}{\|T_\mathrm{BE}\|}. \tag{8}$$

This approach has a down side, that solving two linear systems is required to obtain T_BE and T_CN. This is, from the computational point of view, expensive. Nevertheless overall decrease in computational time is expected.

The algorithm below, describing our adaptive time-stepping procedure, is based on the one for adaptive time stepping for processes in spent nuclear fuel repositories [2]. It has several parameters:

1) τ^1 – initial timestep;
2) N_{Adapt} – a parameter showing how often the adaptive time stepping strategy is applied, e.g. $N_{\text{Adapt}} = 1$ shows that the adaptive time stepping is used on each step while $N_{\text{Adapt}} = 3$ – that the adaptive time stepping is performed at every third time step, $N_{\text{Adapt}} = 0$ indicates that all time steps are non-adaptive.
3) $\lambda_{\text{NonAdapt}}$ – a parameter showing whether and by how much the time step is multiplied, in non-adaptive time steps, e.g. $\lambda_{\text{NonAdapt}} = 1$ means that the time step is not changed, while $\lambda_{\text{NonAdapt}} = 1.2$ means that the time step on the current level is multiplied by 1.2 for the next time level.
4) $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are minimal and maximal thresholds for the error estimate η.

Algorithm 1 (Adaptive Time-Stepping Procedure):

```
1.  for k = 1, 2, ... until the end of time do
2.      if CurrentStepIsAdaptive(N_Adapt, k)
2.      then
3.          do
4.              compute T_BE, T_CN with τ^k
5.              compute η
6.              if η < ε_min then τ^{k+1} = 2τ^k
7.              if η > ε_max then τ^k = 0.5τ^k
8.          while η > ε_max   // if too big error
                              // stay on the same timestep
9.          T^{k+1} = T_BE
10.     else
11.         compute T_BE with τ^k
12.         T^{k+1} = T_BE
13.         τ^{k+1} = τ^k λ_NonAdapt
14.     end if
15. end for
```

The last timestep is always truncated to the time of simulation.

Inner PCG iteration with the BoomerAMG [3] preconditioner, part of the software package HYPRE [10], is used for the solution of (7). The preconditioner is reconstructed if the number of inner iterations goes above 12. The reconstruction takes place before the solution of the next timestep.

IV. Computer Simulations and Analysis of the Output Results

The IBM Blue Gene/P computer, located at the Bulgarian Supercomputing Center, is used for the simulations and numerical experiments with the new adaptive time stepping algorithm. This machine consists of two racks, 2048 Power PC 450 based compute nodes, 8192 processor cores and a total of 4 TB random access memory. Each processor core has a double-precision, dual pipe floating-point core accelerator. Sixteen I/O nodes are connected via fiber optics to a 10 Gbps Ethernet switch.

The material properties which are used in the simulations are taken from [4]. The applied electrical power is 15 W, and the simulation is done for 7 minutes.

We run several test to choose a suitable set of values for the threshold parameters $\varepsilon_{\min}$ and $\varepsilon_{\max}$. As a quantitative criterion of quality of the solution we used two volumes – the volume Vol_1, which is the volume of the tissue, where the cumulative damage integral Ψ is greater than 1, and $Vol_{4.6}$ – the volume of the tissue, where $\Psi > 4.6$. The results of the nonadaptive algorithm with step $\tau = 1$ s were compared with the ones from adaptive runs. As a result of these tests we found that an acceptable variation in the two important volumes less than 3 % occurs when the threshold interval is $[2.5\times10^{-4}, 1.25\times10^{-3}]$ and this interval is used in the further computer simulations.

Based on these preliminary tests, a number of runs were done both using 128 and 1024 processors. Uniformly refined mesh was used for the runs on 1024 processors. Some of the output results obtained during the simulations are presented in Table I and Table II. An excellent scalability is observed – we solve eight times bigger problems on eight times more processors for almost the same time. One can see in both tables that the best results with regards to CPU time and number of the inner iterations are obtained when the adaptive strategy is applied at each second time step and meanwhile, at the intermediate time steps τ is multiplied by 1.2. In this case, comparing the total CPU times of the algorithm without the adaptive time-stepping and using this strategy, it is seen that the time of the new algorithm is almost three times shorter.

V. Conclusions

The first experimental results show that the new algorithm is scalable. The tests allowed us to find some suitable parameters and showed the practical usefulness of the developed solver for such kind of computer simulations. One can observe that the computing time is decreased more than three times, the number of outer iterations is decreased from 420 to 71, and the number of inner iteration decreases form 2233 to 535. This preliminary results are a good motivation for further improving the algorithm and doing more simulations.

Acknowledgments

This research is supported in part by Grants DFNI I01/5 and DCVP-02/1 from the Bulgarian NSF and the Bulgarian National Center for Supercomputing Applications (NCSA) giving access to the IBM Blue Gene/P computer.

References

[1] Isaac A. Chang and Uyen D. Nguyen *Thermal modeling of lesion growth with radiofrequency ablation devices* BioMedical Engineering OnLine 2004, 3:27

[2] R. Blaheta, P. Byczanski, R. Kohut, J. Stary, *Algorithms for parallel FEM modelling of thermo-mechanical phenomena arising from the disposal of spent nuclear fuel*, in: O. Stephansson, J.B. Hudson, L. Jing (Eds.), Coupled Thermo-Hydro-Mechanical-Chemical Processes in Geosystems, Elsevier, 2004.

[3] V.E. Henson, U.M. Yang, *BoomerAMG: A parallel algebraic multigrid solver and preconditioner*, Applied Numerical Mathematics 41 (1), Elsevier, 2002, 155–177.

TABLE I
NUMBER OF ITERATIONS AND THE CPU TIME IN THE ADAPTIVE TIME-STEPPING ALGORITHM IN THE CASE OF 128 PROCESSORS.

N_{Adapt}	$\lambda_{\text{NonAdapt}}$	No. of inner iterations	No. of outer iterations	CPU time [s]	Vol_1 cm^3	$Vol_{4.6}$ cm^3
0	1.0	2233	420	7608	22.14	15.60
1	1.0	917	102	3968	22.72	15.93
2	1.0	731	104	3137	22.63	15.87
	1.2	535	71	2321	22.87	16.00
	1.3	587	77	2624	22.87	16.02
3	1.0	700	113	3053	22.58	15.83
	1.2	539	76	2329	22.88	16.03
	1.3	592	77	2559	22.81	15.97

TABLE II
NUMBER OF ITERATIONS AND THE CPU TIME IN THE ADAPTIVE TIME-STEPPING ALGORITHM IN THE CASE OF 1024 PROCESSORS.

N_{Adapt}	$\lambda_{\text{NonAdapt}}$	No. of inner iterations	No. of outer iterations	CPU time [s]	Vol_1 $[\text{cm}^3]$	$Vol_{4.6}$ $[\text{cm}^3]$
0	1.0	604	420	7259	22.21	15.65
1	1.0	777	101	4234	22.70	15.92
2	1.0	594	101	3488	22.70	15.92
	1.2	478	71	2619	23.01	16.10
	1.3	539	77	2982	22.94	16.07
3	1.0	549	104	3121	22.70	15.93
	1.2	455	76	2530	22.85	16.01
	1.3	514	75	2740	22.94	16.06

[4] S. Tungjitkusolmun, S.T. Staelin, D. Haemmerich, J.Z. Tsai, H. Cao, J.G. Webster, F.T. Lee, D.M. Mahvi, V.R. Vorperian, *Three-dimensional finite-element analyses for radio-frequency hepatic tumor ablation*, IEEE transactions on biomedical engineering 49 (1), 2002, 3–9.

[5] S. Tungjitkusolmun, E.J. Woo, H. Cao, J.Z. Tsai, V.R. Vorperian, J.G. Webster, *Thermal-electrical finite element modelling for radio frequency cardiac ablation: Effects of changes in myocardial properties*, Medical and Biological Engineering and Computing 38 (5), 2000. 562–568.

[6] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1996.

[7] S. Brenner, L. Scott, *The mathematical theory of finite element methods*, Texts in applied mathematics, 15, Springer-Verlag, 1994.

[8] E. Hairer, S.P. Norsett, G. Wanner *Solving ordinary differential equations I, II*, Springer Series in Comp. Math., 2000, 2002

[9] N. Kosturski, S. Margenov, Y. Vutov, *Supercomputer Simulation of Radio-Frequency Hepatic Tumor Ablation, AMiTaNS'12 Proceedings*, AIP CP vol. 1487, pp. 120–126.

[10] Lawrence Livermore National Laboratory, *Scalable Linear Solvers Project*, http://www.llnl.gov/CASC/linear_solvers/.

[11] http://en.wikipedia.org/wiki/File:RFA_CT_Leber_001.jpg

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 27–32

Performance Evaluation of MPI/OpenMP Algorithm for 3D Time Dependent Problems

Ivan Lirkov
Institute of Information
and Communication Technologies
Bulgarian Academy of Sciences
Acad G. Bonchev, bl. 25A
1113 Sofia, Bulgaria
ivan@parallel.bas.bg
http://parallel.bas.bg/~ivan/

Marcin Paprzycki Maria Ganzha
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw, Poland
paprzyck@ibspan.waw.pl
maria.ganzha@ibspan.waw.pl
http://www.ibspan.waw.pl/~paprzyck/
http://inf.ug.edu.pl/~mganzha/

Paweł Gepner
Intel Corporation
Pipers Way
Swindon Wiltshire SN3 1RJ
United Kingdom
pawel.gepner@intel.com

***Abstract*—We consider the 3D time dependent Stokes equation on a finite time interval and on a uniform rectangular mesh, approached in terms of velocity and pressure. In a parallel algorithm, based on a novel direction splitting approach, the pressure equation is derived from a perturbed form of the continuity equation, in which the incompressibility constraint is penalized in a negative norm induced by the direction splitting. In order to achieve good parallel performance, the solution of the Poison problem for the pressure correction is replaced by solving a sequence of one-dimensional second order elliptic boundary value problems in each spatial direction. The parallel code was developed using MPI and OpenMP and tested on modern computer systems. The performed tests illustrate the parallel efficiency, and the scalability, of the direction-splitting based algorithm.**

I. Introduction

THE objective of this paper is to analyze the parallel performance of a novel fractional time stepping technique, based on a direction splitting strategy, developed to solve the incompressible Navier-Stokes equations.

Projection schemes were introduced in [4], [12] and they have been used in Computational Fluid Dynamics (CFD) for the last forty years. During these years, such techniques went through some evolution, but the main paradigm, consisting of decomposing vector fields into a divergence-free part and a gradient, has been preserved; see [7] for an overview. In terms of computational efficiency, projection algorithms are far superior than the methods that solve the coupled velocity-pressure system, making them the most popular techniques for solving unsteady Navier-Stokes equations.

The alternating directions algorithm proposed in [5], [6] reduces the computational complexity of the enforcement of the incompressibility constraint. The key idea consists of abandoning the projection paradigm, in which vector fields are decomposed into a divergence-free part and a gradient part. Departure from the projection paradigm has been proved to be very efficient for solving variable density flows [8], [9]. In the new method, the pressure equation is derived from a perturbed form of the continuity equation, in which the incompressibility constraint is penalized in a negative norm induced by the direction splitting. The standard Poisson problem for the pressure correction is replaced by the series of one-dimensional second-order boundary value problems. This technique has been proved to be stable and convergent (see [5], [6]). The parallel performance of the direction splitting algorithm was evaluated (in [10]) on three different parallel architectures when solving 2D problems . The aim of this paper is to study the parallel performance of the algorithm for solving of three dimensional problems.

II. Stokes Equation

Let us start from the mathematical formulation of the problem. We consider the time-dependent Navier-Stokes equations on a finite time interval $[0,T]$, and in a rectangular domain Ω. Since the nonlinear term in the Navier-Stokes equations does not interfere with the incompressibility constraint, we henceforth mainly focus our attention on the time-dependent Stokes equations written in terms of velocity $\mathbf{u}$ and pressure p:

$$\begin{cases} \mathbf{u}_t - \nu\Delta\mathbf{u} + \nabla p = \mathbf{f} & \text{in} \quad \Omega\times(0,T) \\ \nabla\cdot\mathbf{u} = 0 & \text{in} \quad \Omega\times(0,T) \\ \mathbf{u}|_{\partial\Omega} = 0, \quad \partial_n p|_{\partial\Omega} = 0 & \text{in} \quad (0,T) \\ \mathbf{u}|_{t=0} = \mathbf{u}_0, \quad p|_{t=0} = p_0 & \text{in} \quad \Omega \end{cases}, \tag{1}$$

where $\mathbf{f}$ is a smooth source term, ν is the kinematic viscosity, and $\mathbf{u}_0$ is a solenoidal initial velocity field with a zero normal trace. Let us also assume that the time interval $[0,T]$ was discretized on a uniform mesh and τ was the time step.

III. Parallel Alternating Directions Algorithm

As what concerns the solution method, Guermond and Minev (in [5], [6]) introduced a novel fractional time stepping technique for solving the incompressible Navier-Stokes equations. This technique is based on a direction splitting strategy. They used a singular perturbation of the Stokes equation with a perturbation parameter τ. In this way, the standard Poisson problem was replaced by series of one-dimensional second-order boundary value problems.

A. Formulation of the Scheme

The scheme used in the algorithm is composed of the following parts: pressure prediction, velocity update, penalty step, and pressure correction. We now describe an algorithm that uses the direction splitting operator

$$A := \left(1 - \frac{\partial^2}{\partial x^2}\right)\left(1 - \frac{\partial^2}{\partial y^2}\right)\left(1 - \frac{\partial^2}{\partial z^2}\right).$$

- *Pressure predictor*. Denoting by p_0 the pressure field at $t = 0$, the algorithm is initialized by setting $p^{-\frac{1}{2}} = p^{-\frac{3}{2}} = p_0$. Then, for all $n \geq 0$, a pressure predictor is computed:

$$p^{*,n+\frac{1}{2}} = 2p^{n-\frac{1}{2}} - p^{n-\frac{3}{2}}. \tag{2}$$

- *Velocity update*.
The velocity field is initialized by setting $\mathbf{u}^0 = \mathbf{u}_0$, and for all $n \geq 0$ the velocity update is computed by solving the following series of one-dimensional problems

$$\frac{\boldsymbol{\xi}^{n+1} - \mathbf{u}^n}{\tau} - \nu\Delta\mathbf{u}^n + \nabla p^{*,n+\frac{1}{2}} = \mathbf{f}|_{t=(n+\frac{1}{2})\tau},$$

$$\frac{\boldsymbol{\eta}^{n+1} - \boldsymbol{\xi}^{n+1}}{\tau} - \frac{\nu}{2}\frac{\partial^2(\boldsymbol{\eta}^{n+1} - \mathbf{u}^n)}{\partial x^2} = 0, \tag{3}$$

$$\frac{\boldsymbol{\zeta}^{n+1} - \boldsymbol{\eta}^{n+1}}{\tau} - \frac{\nu}{2}\frac{\partial^2(\boldsymbol{\zeta}^{n+1} - \mathbf{u}^n)}{\partial y^2} = 0, \tag{4}$$

$$\frac{\mathbf{u}^{n+1} - \boldsymbol{\zeta}^{n+1}}{\tau} - \frac{\nu}{2}\frac{\partial^2(\mathbf{u}^{n+1} - \mathbf{u}^n)}{\partial z^2} = 0, \tag{5}$$

where $\boldsymbol{\xi}^{n+1}|_{\partial\Omega} = \boldsymbol{\eta}^{n+1}|_{\partial\Omega} = \boldsymbol{\zeta}^{n+1}|_{\partial\Omega} = \mathbf{u}^{n+1}|_{\partial\Omega} = 0$.

- *Penalty step* The intermediate parameter ϕ is approximated by solving $A\phi = -\frac{1}{\tau}\nabla \cdot \mathbf{u}^{n+1}$. Owing to the definition of the direction splitting operator A, this is done by solving the following series of one-dimensional problems:

$$\begin{array}{lll} \theta - \theta_{xx} = & -\frac{1}{\tau}\nabla \cdot \mathbf{u}^{n+1}, & \theta_x|_{\partial\Omega} = 0, \\ \psi - \psi_{yy} = & \theta, & \psi_y|_{\partial\Omega} = 0, \\ \phi - \phi_{zz} = & \psi, & \phi_z|_{\partial\Omega} = 0, \end{array} \tag{6}$$

- *Pressure update*
The last sub-step of the algorithm consists of updating the pressure:

$$p^{n+\frac{1}{2}} = p^{n-\frac{1}{2}} + \phi - \chi\nu\nabla \cdot \frac{\mathbf{u}^{n+1} + \mathbf{u}^n}{2} \tag{7}$$

The algorithm is in a standard incremental form when the parameter $\chi = 0$; while the algorithm is in a rotational incremental form when $\chi \in (0, \frac{1}{2}]$. The convergence tests reported in [5], [6] confirm that the rotational form of the incremental version of the method is second-order in time for the L^2-norm of the velocity field.

IV. Experimental Results

Lest us recall, that our aim is to evaluate the performance of just described method for 3D problems. Therefore, we have solved the problem (1) in the domain $\Omega = (0, 1)^3$, for $t \in [0, 2]$, with Dirichlet boundary conditions. The discretization in time was done with time step 10^{-2}. The parameter in the pressure update sub-step was $\chi = \frac{1}{2}$, and the kinematic viscosity was $\nu = 10^{-3}$. The second order central differences were used for the discretization in space on a rectangular mesh with mesh sizes $h_x = \frac{1}{n_x - 1}$, $h_y = \frac{1}{n_y - 1}$, and $h_z = \frac{1}{n_z - 1}$. Thus, the equation (3) resulted in linear systems of size n_x, the equation (4) resulted in linear systems of size n_y, and the equation (5) – in linear systems of size n_z. The total number of unknowns in the discrete problem was $800\, n_x\, n_y\, n_z$.

To solve the problem, a portable parallel code was designed and implemented in C. The parallelization was based on the MPI and OpenMP standards [2], [3], [11], [13]. In the code, we used the LAPACK subroutines DPTTRF and DPTTS2 (see [1]) for solving tridiagonal systems of equations resulting from equations (3), (4), (5), and (6), for the unknowns corresponding to the internal nodes of each sub-domain. The same subroutines were used to solve the tridiagonal systems with the Schur complement.

The parallel code has been tested on the following three systems: (1) a cluster computer *Galera*, located in the Polish Centrum Informatyczne TASK, (2) on a cluster computer system *HPCG*, located in the Institute of Information and Communication Technologies, and (3) on the IBM Blue Gene/P machine, at the Bulgarian Supercomputing Center. Table I summarizes the information about compilers and libraries used on the three computers. In our experiments, times have been collected using the MPI provided timer, and we report the best results from multiple runs. However, in all our experiments, recorded times did not differ by more than 5%. In the following tables, we report the elapsed (wall-clock) time T_p in seconds using m MPI processes and k OpenMP processes, where $p = m \times k$, and the parallel speed-up $S_p = T_1/T_p$.

Tables II and III show the results collected on the Galera. It is a Linux cluster with 336 nodes, and two Intel Xeon quad core processors per node. Each processor runs at 2.33 GHz. Processors within each node share 8, 16, or 32 GB of memory, while nodes are interconnected with a high-speed InfiniBand network (see also http://www.task.gda.pl/kdm/sprzet/Galera). Here, we used an Intel C compiler, and compiled the code with the option "-O3 -openmp". For solving the tridiagonal systems of equations using LAPACK subroutines we linked our code to multi-threaded layer Intel Math Kernel Library (see http://software.intel.com/en-us/articles/intel-mkl/).

Note that, the discrete problem with $n_x = n_y = 400$, $n_z = 800$ requires 22 GB of memory. That is why, for larger problems, we could not run the code for small number of nodes. As a matter of fact, the largest considered problem required a minimum of 64 nodes.

Table IV contains the speed-up obtained on Galera. For the reasons described above, it was impossible to calculate the standard speed-up (time on a single core vs. time on P cores). That is why we report the speed-up on Galera only for problems with $n_x, n_y = 100, 200, 400$, $n_z = 100, 200, 400, 800$. However it is easy to calculate the "normalized" speed-up even for the largest problem. For instance, when comparing execution time on 128 and 256 cores with that on 64 cores

TABLE I
COMPILERS AND LIBRARIES ON THE THREE COMPUTER SYSTEMS

	Galera	HPCG	IBM Blue Gene/P
Compiler	Intel C Compiler 12.1.0	Intel C Compiler 12.1.0	IBM XL C Compiler 9.0
MPI	OpenMPI 1.4.3	Intel MPI 4.0.3.008	MPICH2
LAPACK	Intel Math Kernel Library 10.0	Intel Math Kernel Library 10.0	Engineering and Scientific Subroutine Library 5.1

TABLE II
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON ONE NODE OF GALERA.

n_x	n_y	n_z	k			
			1	2	4	8
100	100	100	179.13	108.92	82.91	78.31
100	100	200	386.53	233.35	176.05	160.64
100	200	200	848.59	506.58	377.94	339.81
200	200	200	1802.57	1057.07	788.10	685.73
200	200	400	3695.46	2157.89	1593.42	1374.58
200	400	400	7618.41	4440.56	3272.34	2765.14
400	400	400	15839.00	9120.79	6545.47	5473.50
400	400	800	32763.20	18849.10	13582.80	11524.80

TABLE III
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON MANY NODES OF GALERA.

n_x	n_y	n_z	nodes							
			2	4	8	16	32	64	128	256
			k=8							
100	100	100	42.3	21.1	10.8	5.8	3.2	2.6	2.2	1.8
100	100	200	83.2	42.2	21.0	11.1	6.1	3.3	2.4	1.7
100	200	200	174.5	86.0	43.3	22.8	11.6	6.2	4.6	2.6
200	200	200	343.2	166.3	81.0	44.3	22.3	11.5	7.1	4.1
200	200	400	692.3	337.8	165.3	89.1	45.5	22.8	13.5	7.8
200	400	400	1410.2	706.4	347.7	187.0	90.7	46.3	29.2	14.2
400	400	400	2795.8	1429.7	703.8	357.8	174.2	85.5	51.6	25.9
400	400	800	5470.0	2758.6	1415.7	723.1	359.2	177.1	108.0	56.5
400	800	800	11613.7	5594.8	2797.6	1466.9	741.7	372.7	225.5	113.4
800	800	800		11926.1	5669.0	2802.2	1485.4	753.8	404.5	207.9
800	800	1600			12028.3	5660.6	2874.8	1515.1	817.3	428.5
800	1600	1600				12006.2	5719.0	2868.4	1653.2	853.2
1600	1600	1600					12151.7	5757.7	3022.9	1646.3
1600	1600	3200						12231.2	6076.1	3306.1

one can say that such speed-up is 2.015 and 3.699 respectively. Here, the slight superlinear speed-up can be attributed to the well-known effect caused by halving the problem size and improving memory management. These results indicate also that the communication in our parallel algorithm is mainly local. Specifically, if halving the problem leads to superlinear speed-up, it means that communication between nodes is not as important as memory contention within a node.

In our previous work [10] we observed slower performance of the MPI code for solving 2D problem on a single node, while using all available cores. Therefore, we have used OpenMP for solving of 3D problem on a single node. However, the slower performance using 8 OpenMP processes is clearly visible. It can be stipulated that this effect is a result of limitations of the memory subsystems, and their hierarchical organization. One of them might be the limited bandwidth of the main memory bus. This causes processors to "starve" for data, thus, decreasing the overall performance. This stipulation is consistent with the above observation based on superlinearity of speed-up for largest problems.

Tables V and VI show the results collected on the HPCG cluster. HP Cluster Platform Express 7000 consists of 36 blades BL 280c, dual Intel Xeon X5560 processors (total of 576 cores). Each processor runs at 2.8 GHz. Processors within each blade share 24 GB RAM, while nodes are interconnected with non-blocking DDR Interconnection via a Voltaire Grid director 2004 with latency 2.5 μs and bandwidth 20 Gbps (see also http://www.grid.bas.bg/hpcg/). Again, we used an Intel C compiler, and compiled the code with the option "-O3 -openmp". For solving the tridiagonal systems of equations using LAPACK subroutines, we linked our code to the multi-threaded layer Intel Math Kernel Library. The execution time presented in Table VI shows that for solving small systems using the multi-threaded library the best results are obtained when using 8 OpenMP processes per blade. However, hyper-threading (using 2 OpenMP processes per physical core) helps only when solving discrete problems with more than 16×10^6 grid points.

To provide an analytical view on performance, the speed-up obtained on the HPCG is reported in Table VII. Here, let us

TABLE IV
SPEED-UP ON GALERA.

n_x	n_y	n_z	$m \times k$										
			2	4	8	16	32	64	128	256	512	1024	2048
100	100	100	1.64	2.16	2.29	4.23	8.50	16.64	30.66	55.35	69.74	83.10	99.59
100	100	200	1.66	2.20	2.41	4.64	9.17	18.43	34.65	63.66	116.14	158.43	233.33
100	200	200	1.68	2.25	2.50	4.86	9.87	19.61	37.22	72.81	136.44	185.16	327.56
200	200	200	1.71	2.29	2.63	5.25	10.84	22.24	40.72	80.80	156.55	254.74	439.52
200	200	400	1.71	2.32	2.69	5.34	10.94	22.35	41.45	81.29	162.28	274.49	473.96
200	400	400	1.72	2.33	2.76	5.40	10.78	21.91	40.73	84.03	164.69	260.76	537.08
400	400	400	1.74	2.42	2.89	5.67	11.08	22.51	44.27	90.91	185.23	306.84	611.41
400	400	800	1.74	2.41	2.84	5.99	11.88	23.14	45.31	91.21	184.96	303.23	579.91

TABLE V
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON ONE NODE OF HPCG.

n_x	n_y	n_z	k				
			1	2	4	8	16
100	100	100	86.14	48.47	30.23	26.26	27.42
100	100	200	190.45	98.81	63.44	52.07	54.51
100	200	200	397.90	224.73	137.70	110.45	109.30
200	200	200	899.24	456.19	279.53	207.31	219.29
200	200	400	1891.51	945.52	598.95	447.60	449.36
200	400	400	3806.96	1987.19	1225.93	889.76	843.52
400	400	400	6519.28	4162.41	2507.75	1900.73	1722.60
400	400	800	14343.40	8174.26	5048.87	3840.93	3648.71

recall that (similarly to the case of Galera) we were not able to solve largest problems for small number of nodes. Specifically, the largest discrete problem that we could solve on a single blade had 128×10^6 grid points. Furthermore, problems larger than these reported above did not fit into the available memory at all. However, we can see that the normalized speed-up for the largest solvable problem was 1.94.

Tables VIII and IX present times collected on the IBM Blue Gene/P, at the Bulgarian Supercomputing Center. It consists of 2048 compute nodes with quad core PowerPC 450 processors (running at 850 MHz). Each node has 2 GB of RAM. For the point-to-point communications a 3.4 Gb 3D mesh network is used. Reduction operations are performed on a 6.8 Gb tree network (for more details, see http://www.scc.acad.bg/). We have used the IBM XL C compiler and compiled the code with the following options: "-O5 -qstrict -qarch=450d -qtune=450 -qsmp=omp". For solving the tridiagonal systems using LAPACK subroutines, we linked our code to the multi-threaded Engineering and Scientific Subroutine Library (ESSL) (see http://www-03.ibm.com/systems/software/essl/index.html). Note that, the memory of a node of the IBM supercomputer is substantially smaller than that on the clusters and was not sufficient for solving 3D problem larger than $n_x = n_y = n_z = 200$. We solved these problems on two and more nodes. Table X shows the speed-up obtained on the Blue Gene. Because of smaller memory per node we calculated the actual speed-up only for $n_x, n_y, n_z = 100, 200$. Interestingly, a super-linear speed-up was observed using up to 128 processes. Here, note that individual processors on supercomputer are slower than these on clusters while the communication is faster (due to special networking used in the Blue Gene). Furthermore, due to the memory scarcity the above mentioned problem halving effect is particularly pronounced. Separately, note that the normalized speed-up for the largest problem we were able to solve was 1.98.

Finally, we represent (in Figure 1) performance of the three computer systems. Here, we represent results for discrete problem sizes that were obtainable across all machines. Here, it can be seen that, for up to 128 processes, the HPCG cluster is the most efficient for our problem. For 256 processes (using 16 OpenMP processes), for the smallest problems, we observe a performance drop of the HPCG. Here, the Galera becomes the performance leader. We observed a performance drop on both clusters when we use 8 OpenMP processes on one node. This happens regardless of the fact that we have used a hybrid OpenMP+MPI approach, which should have resulted in optimal performance of computers with mixed shared-distributed memory.

V. CONCLUDING REMARKS

In this paper we have investigated performance of a novel approach to the solution of 3D Navier-Stokes equations on three parallel computers. We have found out that the mixed OpenMP+MPI approach, used to implement the proposed algorithm, works well on shared-distributed memory computers. Furthermore, the Blue Gene/P machine is not well balanced, with memory that is relatively too small, networking that is relatively too fast, while processors being relatively too slow. The HPCG and Galera machines exhibits unusual behavior when all of its processors and cores on one node are used. The Galera seems to be best balanced (even though not the fastest of the three). It was also the Galera that allowed us to solve the largest problems.

ACKNOWLEDGMENTS

Computer time grants from the Bulgarian Supercomputing Center and the TASK computing center in Gdansk, Poland are

TABLE VI
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON MANY NODES OF HPCG.

			nodes			
			2	4	8	16
n_x	n_y	n_z	k=8			
100	100	100	15.10	7.40	4.38	2.56
100	100	200	28.61	17.06	8.14	5.05
100	200	200	61.72	35.38	17.58	9.45
200	200	200	118.89	68.54	35.96	18.63
200	200	400	246.84	120.25	70.53	37.90
200	400	400	500.00	259.89	152.23	83.25
400	400	400	914.44	505.90	248.07	146.35
400	400	800	2109.01	985.69	529.95	321.91
400	800	800	4271.35	2074.35	1042.33	626.71
800	800	800		4463.87	2204.67	1167.19
800	800	1600			4495.82	2378.40
n_x	n_y	n_z	k=16			
100	100	100	15.45	9.08	5.73	3.92
100	100	200	28.30	18.56	10.47	7.09
100	200	200	57.77	34.18	19.10	12.49
200	200	200	115.16	61.55	35.88	21.00
200	200	400	229.05	122.69	72.39	39.07
200	400	400	439.35	237.16	132.74	71.45
400	400	400	894.70	462.40	242.39	137.17
400	400	800	1744.18	886.36	485.19	261.39
400	800	800	3733.36	1928.69	1014.47	524.18
800	800	800		4138.85	2209.53	1000.43
800	800	1600			4169.04	2138.83

TABLE VII
SPEED-UP ON HPCG.

n_x	n_y	n_z	$m \times k$							
			2	4	8	16	32	64	128	256
100	100	100	1.78	2.85	3.28	5.71	11.64	19.69	33.63	21.96
100	100	200	1.93	3.00	3.66	6.66	11.16	23.40	37.71	26.85
100	200	200	1.77	2.89	3.60	6.45	11.25	22.64	42.10	31.86
200	200	200	1.97	3.22	4.34	7.56	13.12	25.01	48.27	42.83
200	200	400	2.00	3.16	4.23	7.66	15.73	26.82	49.90	48.41
200	400	400	1.92	3.11	4.28	7.61	14.65	25.01	45.73	53.28
400	400	400	1.57	2.60	3.43	7.13	12.89	26.28	44.54	47.53
400	400	800	1.75	2.84	3.73	6.80	14.55	27.07	44.56	54.87

TABLE VIII
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON ONE NODE OF IBM BLUE GENE/P.

n_x	n_y	n_z	k		
			1	2	4
100	100	100	882.72	516.00	306.88
100	100	200	1815.72	1040.66	632.95
100	200	200	3782.91	2153.78	1291.85
200	200	200	7685.65	4343.88	2710.08

TABLE IX
EXECUTION TIME FOR SOLVING OF 3D PROBLEM ON MANY NODES OF IBM BLUE GENE/P.

n_x	n_y	n_z	nodes								
			2	4	8	16	32	64	128	256	512
			k=4								
100	100	100	146.8	74.9	38.3	20.9	10.9	6.2	4.5	3.0	1.9
100	100	200	309.0	157.3	77.0	41.0	21.1	11.2	7.1	4.9	3.1
100	200	200	626.8	317.8	159.9	79.3	39.8	21.3	13.2	7.9	5.1
200	200	200	1315.4	645.1	323.4	164.2	79.1	39.5	24.9	14.7	8.5
200	200	400	2645.4	1325.9	647.5	337.9	165.2	80.2	42.8	26.0	15.0
200	400	400		2656.2	1341.7	657.7	330.0	159.4	87.5	46.2	27.4
400	400	400			2719.6	1383.3	677.9	334.4	176.2	96.3	48.7
400	400	800				2807.4	1387.8	681.4	348.8	184.2	95.7
400	800	800					2753.6	1379.3	692.2	362.7	184.6
800	800	800						2853.6	1457.6	748.5	374.5
800	800	1600							2914.0	1524.5	749.6
800	1600	1600								2980.3	1498.6

TABLE X
SPEED-UP ON IBM BLUE GENE/P.

n_x	n_y	n_z	$m \times k$										
			2	4	8	16	32	64	128	256	512	1024	2048
100	100	100	2.11	4.43	8.78	17.02	33.73	64.93	111.54	191.55	351.96	421.32	461.68
100	100	200	2.04	4.45	8.96	17.61	34.78	68.73	125.50	218.75	407.52	517.90	593.41
100	200	200	2.08	4.54	9.16	18.38	36.92	72.02	134.22	246.66	452.54	616.02	746.84
200	200	200	2.00	4.18	8.41	17.23	35.15	75.19	138.63	259.63	479.53	770.25	1006.01

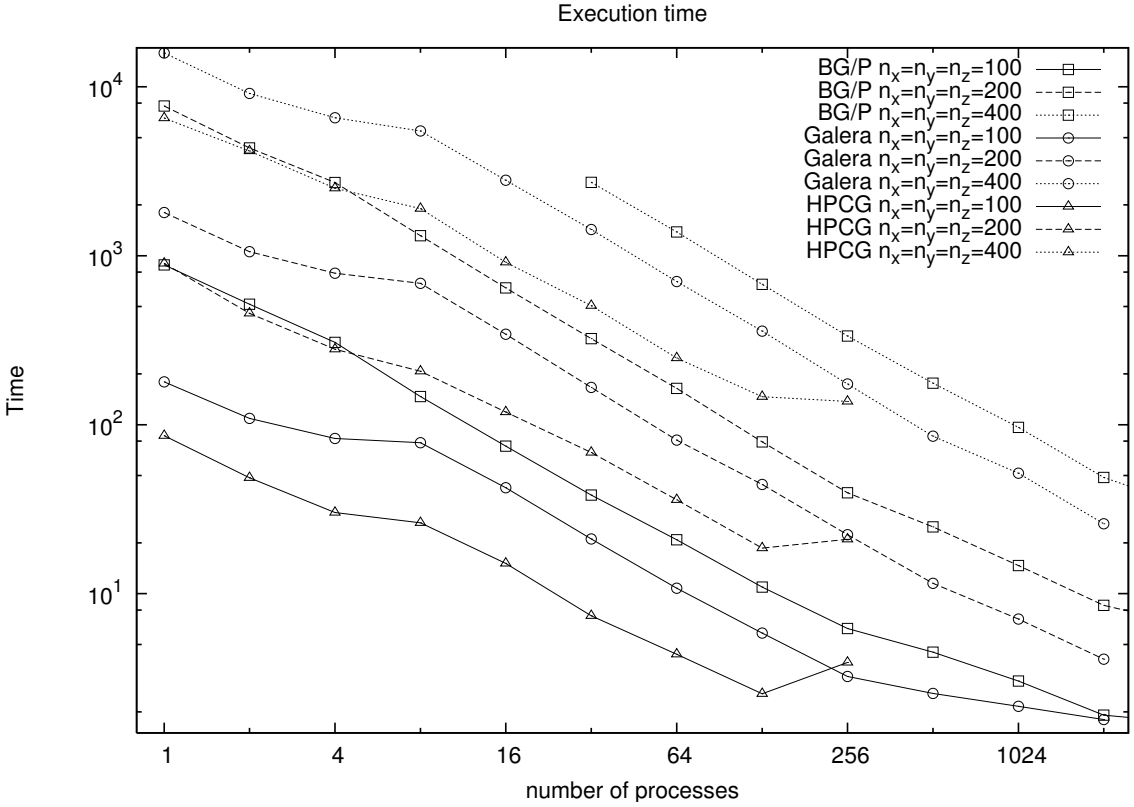


Fig. 1. Execution time for $n_x = n_y = n_z = 100, 200, 400$.

kindly acknowledged. This research was partially supported by grants DCVP 02/1 and I01/5 from the Bulgarian NSF. Work presented here is a part of the Poland-Bulgaria collaborative grant "Parallel and distributed computing practices".

REFERENCES

[1] Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Croz, J.D., Greenbaum, A., Hammarling, S., McKenney, A., Sorensen, D.: LAPACK Users' Guide. SIAM, Philadelphia, third edn. (1999)

[2] Chandra, R., Menon, R., Dagum, L., Kohr, D., Maydan, D., McDonald, J.: Parallel programming in OpenMP. Morgan Kaufmann (2000)

[3] Chapman, B., Jost, G., Van Der Pas, R.: Using OpenMP: portable shared memory parallel programming, vol. 10. MIT press (2008)

[4] Chorin, A.J.: Numerical solution of the Navier-Stokes equations. Math. Comp. 22, 745–762 (1968)

[5] Guermond, J.L., Minev, P.: A new class of fractional step techniques for the incompressible Navier-Stokes equations using direction splitting. Comptes Rendus Mathematique 348(9–10), 581–585 (2010)

[6] Guermond, J.L., Minev, P.: A new class of massively parallel direction splitting for the incompressible navier–stokes equations. Computer Methods in Applied Mechanics and Engineering 200(23), 2083–2093 (2011)

[7] Guermond, J.L., Minev, P., Shen, J.: An overview of projection methods for incompressible flows. Comput. Methods Appl. Mech. Engrg. 195, 6011–6054 (2006)

[8] Guermond, J.L., Salgado, A.: A fractional step method based on a pressure poisson equation for incompressible flows with variable density. Comptes Rendus Mathematique 346(15–16), 913–918 (2008)

[9] Guermond, J.L., Salgado, A.: A splitting method for incompressible flows with variable density based on a pressure Poisson equation. Journal of Computational Physics 228(8), 2834–2846 (2009)

[10] Lirkov, I., Paprzycki, M., Ganzha, M.: Performance analysis of parallel alternating directions algorithm for time dependent problems. In: Wyrzykowski, R., Dongarra, J., Karczewski, K., Waśniewski, J. (eds.) 9th international conference on Parallel Processing and Applied Mathematics, PPAM 2011, Part I. Lecture notes in computer science, vol. 7203, pp. 173–182. Springer (2012)

[11] Snir, M., Otto, S., Huss-Lederman, S., Walker, D., Dongarra, J.: MPI: The Complete Reference. Scientific and engineering computation series, The MIT Press, Cambridge, Massachusetts (1997), second printing

[12] Temam, R.: Sur l'approximation de la solution des équations de Navier-Stokes par la méthode des pas fractionnaires. Arch. Rat. Mech. Anal. 33, 377–385 (1969)

[13] Walker, D., Dongarra, J.: MPI: a standard Message Passing Interface. Supercomputer 63, 56–68 (1996)

Education, Curricula & Research Methods

ECRM is a FedCSIS conference area aiming at interchange of information, ideas, new viewpoints and research undertakings related to university education and curricula as well as recommended methods of doing research in all computing disciplines, i.e. computer science, computer engineering, software engineering, information technology, and information systems. This area spans typical FedCSIS events (conferences, workshops, etc.) with rigorous paper submissions and review processes as well as panels, PhD and research consortia, summer schools, etc. Events that constitute ECRM are:

- DS-RAIT'13 - Doctoral Symposium on Recent Advances in Information Technology
- ISEC'2012 - Information Systems Education & Curricula Workshop

Doctoral Symposium on Recent Advances in Information Technology

THE first international Doctoral Symposium on Recent Advances in Information Technology (DS-RAIT 2013) will be held in Cracow (Poland) on September 8-11, 2013 as a satellite event of the Federated Conference on Computer Science and Information Systems (FedCSIS 2013) and Education, Curricula & Research Methods (ECRM 2013) conference.

The aim of this meeting is to provide a platform for exchange of ideas between early-stage researchers, in Computer Science, PhD students in particular. Furthermore, the symposium will provide all participants an opportunity to get feedback on their studies from experienced members of the IT research community invited to chair all DS-RAIT thematic sessions. Therefore, submission of research proposals with limited preliminary results is strongly encouraged.

Besides receiving specific advice for their contributions all participants will be invited to attend plenary lectures on conducting high-quality research studies, excellence in scientific writing and issues related to intellectual property in IT research. Authors of the two most outstanding submissions will have a possibility to present their papers in a form of short plenary lecture.

Topics

DS-RAIT 2013 invites the submission of papers on all aspects of Information Technology including, but not limited to:

- Automatic Control and Robotics
- Bioinformatics
- Cloud, GPU and Parallel Computing
- Cognitive Science
- Computer Networks
- Computational Intelligence
- Cryptography
- Data Mining and Data Visualization
- Database Management Systems
- Expert Systems
- Image Processing and Computer Animation
- Information Theory
- Machine Learning
- Natural Language Processing
- Numerical Analysis
- Operating Systems
- Pattern Recognition
- Scientific Computing
- Software Engineering

Honorary Conference Chair

Kacprzyk, Janusz, Systems Research Institute of the Polish Academy of Sciences, Poland

Event Chairs

Golunska, Dominika, Cracow University of Technology, Poland

Kowalski, Piotr Andrzej, Systems Research Institute of the Polish Academy of Sciences, Poland

Lukasik, Szymon, Systems Research Institute, Polish Academy of Sciences, Poland

Program Committee

Arabas, Jaroslaw, Warsaw University of Technology, Poland

Atanassov, Krassimir T., Bulgarian Academy of Sciences, Bulgaria

Balazs, Krisztian, Budapest University of Technology and Economics, Hungary

Castrillon-Santana, Modesto, University of Las Palmas de Gran Canaria, Spain

Charytanowicz, Malgorzata, Catholic University of Lublin, Poland

Corpetti, Thomas, University of Rennes, France

Courty, Nicolas, University of Bretagne Sud, France

Fournier-Viger, Philippe, University of Moncton, Canada

Gil, David, University of Alicante, Spain

Hu, Bao-Gang, Chinese Academy of Sciences, China

Koczy, Laszlo, Szechenyi Istvan University, Hungary

Kulczycki, Piotr, Systems Research Institute, Polish Academy of Sciences, Poland

Lilik, Ferenc, Szechenyi Istvan University, Hungary

Mesiar, Radko, Slovak University of Technology, Slovakia

Noguera i Clofent, Carles, Academy of Sciences of the Czech Republic, Czech Republic

Petrik, Milan, Masaryk University, Czech Republic

Tormasi, Alex, Szechenyi Istvan University, Hungary

Yang, Yujiu, Tsinghua University, China

Zadrozny, Slawomir, Polish Academy of Sciences, Poland

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 37–42

Full 3D Robotic Arm Control with Stereo Cameras Made in LabVIEW

Roland Szabó, Aurel Gontean
Applied Electronics Department
Faculty of Electronics and
Telecommunications,
"Politehnica" Univ. of Timişoara
Timişoara, România
roland.szabo@etc.upt.ro

***Abstract*—In this paper we shall present a full application made in LabVIEW, which can control a robotic arm. The control is made upon image recognition system. The image recognition library is the NI Vision Development Module, which has all the tools necessary to create industrial computer vision applications. LabVIEW was also a good choice, because it has integrated a lot of interface drivers. We used the VISA (Virtual Instrument Software Architecture) driver for controlling the robotic arm on the RS-232 interface [3], [4]. We also have a very friendly GUI (Graphical User Interface) and a very friendly graphical programming environment. Our algorithm is to detect certain key parts of the robotic arm, then unite tem with straight lines, create the robotic arm's skeleton and introduce the data in the PC. After this, with some mathematical calculations we can compute exactly the angles needed to be moved by each motor to move the robotic arm in the desired point in the space. The joints of the robotic arms are detected with color recognition algorithm. We placed colored bottle stopper over each joint that interested us and recognized it with color filtering, this way we had extracted from the capered image that key points which we needed.**

I. Introduction

THIS paper presents an application made in LabVIEW which can control the robotic arm with stereo cameras.

In the industrial environment most of the robots have no "eyes", have no cameras, they just move following the previously learned paths and they make no decisions during execution of tasks. Their control software is simple, with no complications, most of the robots have clear paths with no obstacles and the engineers are always close to correct if a problem occurs [1].

As we maybe see in today's factories the robotic arm are even separated in cages, so no obstacles can get in their way and no person can get close to the even accidentally. These rules of caging robotic arm were are made for user safety, because these robotic arms can really injure a human if he gets in their way. If you take a robotic arm programming class of one semester, you shall see that almost one semester you just learn the work protection and safety task and only after than you can get to the fun part, to really program the robotic arm.

This is all good, but maybe this can't go forever, there has to be a time when a robot and human can work together, without to always be aware that a robotic arm can hurt the operator.

Here comes the computer vision in the robotic industry, which can change the way robots were controlled.

Of course we know that computer vision was integrated in the robotic industry even a decade ago, but wasn't so spread, and in the industrial robotic arms the cameras for detecting the robotic arm's movement weren't really used. Cameras were used only for object detection, but not to control the robotic arms itself.

This way we can get rid of the cages in the industry for the robotic arms and if an operator or an obstacle gets in the way of the robotic arm, this can avoid it, or it can stop or recalculate it's trajectory on the move.

Another this is we can reduce service and maintenance time of a robotic arm, because we shall need even less calibration, because using cameras the arm can detect it's position and can auto calibrate.

Maybe this robotic arm detection with camera method is not very mature, but is a good start. Our idea is to combine all the robotic arm control algorithms together and not to replace an existing one. We can use the PTP (Point To Point) method to teach the trajectory to the robot, we can use the coordinate system conversion algorithms like the inverse and forward kinematics; we can use one camera for object detection and two or more cameras for the robotic arm position detections in space.

II. Problem Formulation

We had the task quite clear. We had a Lynxmotion AL5 type robotic arm, two webcams and a PC. We could use any programming environment; the idea was to make the job done. To create an application which can recognize the robotic arm in space using computer vision and control it according to its position using the RS-232 interface.

We could use even other auxiliary object too, which could help us, but the important fact was to finish the job successfully.

An idea was to try to find some sort o programming language which can incorporate all the necessary tasks, to be simple to use and to program, because in very complex

systems it helps no to do it in a very low level programming language, because this way we never finish the job. To be also user friendly, to have a string image recognition library built in, to have great execution speeds and to also have many communication interfaces built in to be able to control almost any robotic arm.

III. Problem Solution

We found the suitable programming language which can fulfill the requirements from the previous paragraph, this programming language is called: LabVIEW.

LabVIEW made everything possible, like user friendly interface and user friendly graphical programming environment, NI vision Development Module which is a strong computer vision library, and a library for almost all the computer interfaces. We could accomplish everything in only one programming language and create a single application which can do the entire job.

A. Theoretical Background

Our idea was to recognize the robotic arm with the two webcams, but of course to recognize all the arm would be a more complicated job, so the best way is to recognize only key part of the object. Today every object recognition system recognizes only key characteristics of the object and upon these a decision is made to recognize the whole object.

We placed some colored bottle stoppers on the joints of the robotic arm and we recognized these with color recognition algorithm. These joints were than united with straight lines and this way we had the skeleton of the whole robotic arm, which could be introduced in the PC. This method is used when athletes are tested by their performance; they have some luminous spots or colored spots placed on their joints and these spots are united and their movement is tested. This method also is used in computer games when animating characters. Sometimes the animation of the characters is made with the data introduced from real actors which have these spots on them to recognize their movements. We used the same principle to our robotic arm. Of course we can enhance the method by recognizing some key elements on the robotic arm, like some screws or bolts, this way we don't have to place colored objects on them.

After this we drew circles around each recognized joint, because the joints of the robot are exactly in the place where the motors are, so these circles are the paths where the robotic arm part can move around the center point, which is the specific motor.

We drew tangents to the circles and parallels to the tangents, this way we created a 2D coordinate system. This drawing looked like a parallelogram, so our coordinate system is not orthogonal, but it's still good for us. In this coordinate system can calculate the movement on 0Y and 0Z axes. The main information for us was the lengths of the sides of the parallelogram. To know them we used the following algorithm. First we drew the tangent to the circle as long as its radius, it is logic too that this tangent is orthogonal to the radius. We drew the two tangents of the two circles which had the same starting point, because on circles had the center in the base joint and the other in the elbow joint. We knew the target point too. We needed to compute two parallel lines with the tangents which pass through the target point. After this with the equation of the straight line we compute the intersections of one line which is parallel to the tangent and passes though the target point and the other line which has the length of the radius and it's orthogonal to it. This we knew the exact size of the sides of the parallelogram. This is the first information; the other is the size of the radius of the circles, which is also known, so this way we have an orthogonal triangle, from where we could compute the angle which needs to move each motor with arctangent. From this angle we computed the robotic value, constructed the SCPI (Standard Commands for Programmable Instruments) command and sent it on the serial interface [3], [4].

For the movement in the 0X axis we used stereo distance calculation, where we used both of the cameras. We computed the distance to the object and the distance two the gripper, so we used the triangle geometric stereo distance algorithm twice, we computed their difference and that, from that the angle and from that the robotic value which was included in an SCPI command and sent to the robotic arm. So as we seen with these ideas the robotic arm can be moved in 3D, and 3 of its joints manipulated according to its recognized position by the camera.

We needed to compute the robotic values from the angles, for this we came up with equation (1) [1], [2]. After this we concatenated these values into the SCPI (Standard Commands for Programmable Instruments) commands sent to the robotic arm on the RS-232 interface [5].

$$\alpha = \frac{\Delta\omega}{180° - 0°} = \frac{2500 - 500}{180° - 0°} = 11,(1)\ robotic\ values \quad (1)$$

Where 2500 is the maximum and 500 is the minim robotic value and we know that the motors from the robot can make a 180º maximum movement. This means the following shown in equation (2) [1], [2].

$$1°{\sim}11,(1)\ robotic\ values \quad (2)$$

On Fig. 1 we can see the block diagram of the experiment. The image recognition algorithm is presented next.

We first make a color threshold which can be done to and RGB (Red, Green and Blue) image or a HSV (Hue, Saturation and Value) image. This is done for each color, so four times for the blue, yellow red and green bottle stoppers. After this we remove the small particles to exclude false positives and noise. Finally we want to know the exact coordinate of the searched color, so we make a particle

analysis and we search the center of the mass, which is the center of the searched color blob. After this we computer our lines and circles, we create our coordinate system, we overlay everything on the image, we construct the SCPI commands and send them thought the VISA driver to the robotic arm [3], [4].

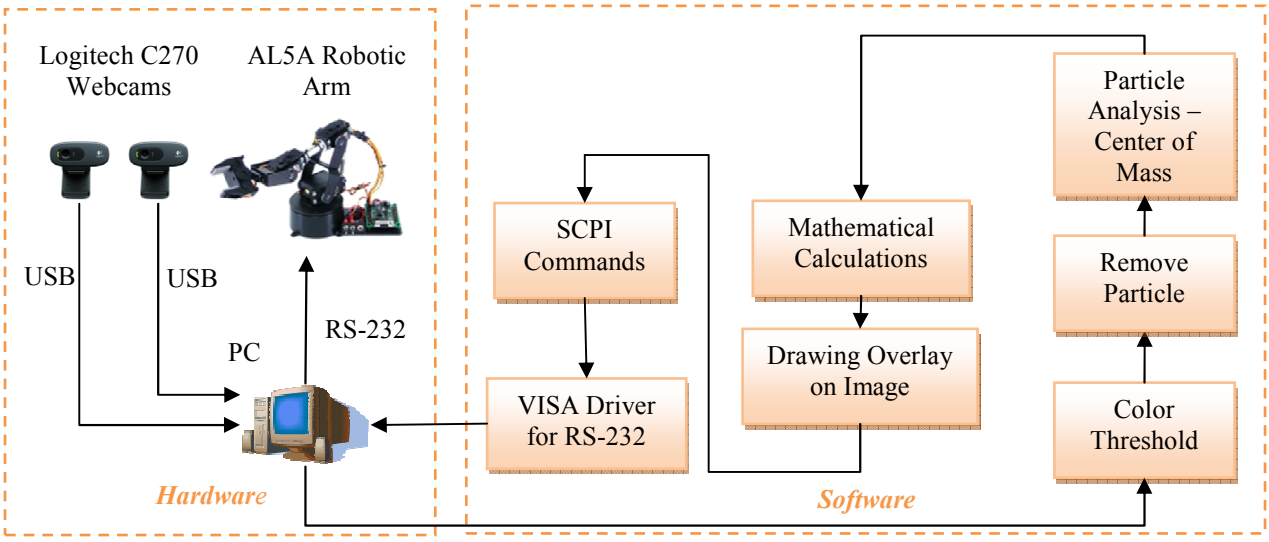


Fig. 1 Block diagram of the experiment

B. Mathematical Calculations

On Fig. 2 we can see the overlay drawings on the acquisitioned image with webcam.

The difference between vectors we calculate the following way.

$$vx = x_2 - x_0 \tag{3}$$

$$vy = y_2 - y_1 \tag{4}$$

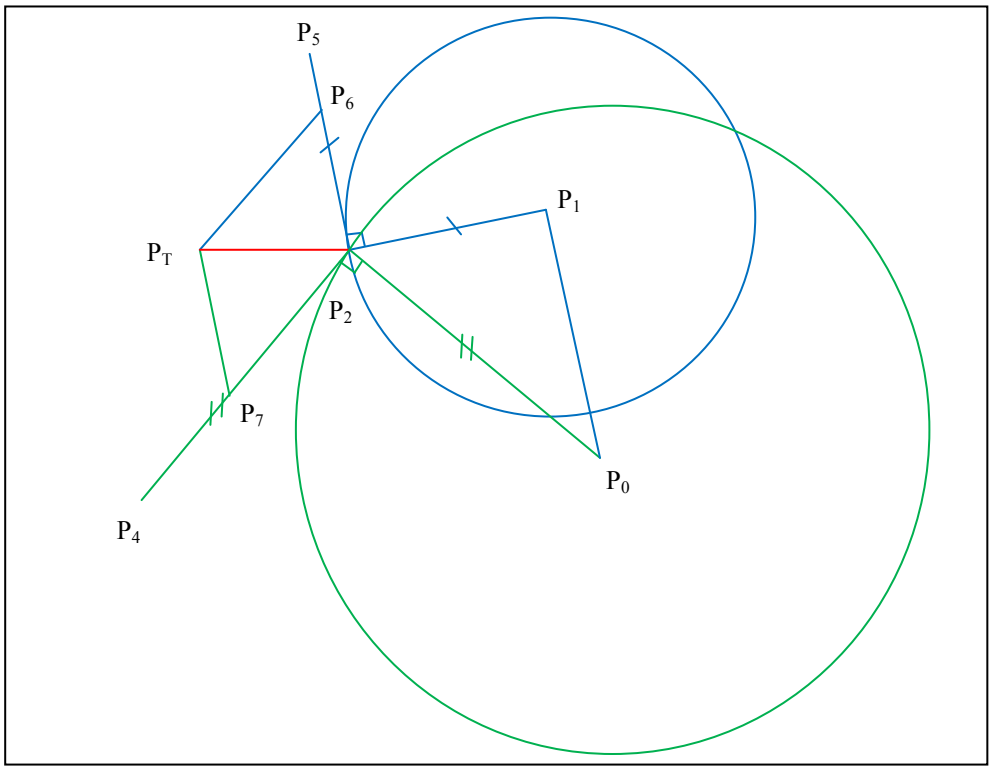


Fig. 2 The overlay drawings on the robotic arm for the mathematical calculations

The vector length we calculate with the following formula.

$$\|l\| = \sqrt{x^2 + y^2} \tag{5}$$

We can compute the orthogonal vector in the following way.

$$temporary = x \tag{6}$$

$$x_O = y \tag{7}$$

$$y_O = -temporary \tag{8}$$

$$x_4 = vx_{Ox} - 2 \cdot vx_{Ox} + x_2 \tag{9}$$

$$y_4 = vx_{Oy} - 2 \cdot vx_{Oy} + y_2 \tag{10}$$

$$x_5 = vy_{Ox} - 2 \cdot vy_{Ox} + x_2 \tag{11}$$

$$y_5 = vy_{Oy} - 2 \cdot vy_{Oy} + y_2 \tag{12}$$

The parallelogram is calculated the following way.

$$m_1 = \frac{y_5 - y_2}{x_5 - x_2} \tag{13}$$

$$m_2 = \frac{y_4 - y_2}{x_4 - x_2} \tag{14}$$

$$y = mx + b \tag{15}$$

$$y_2 = m_1 x_2 + b \tag{16}$$

$$y = m_1 x + y_2 - m_1 x_2 \tag{17}$$

$$y_6 = m_1 x_6 + y_2 - m_1 x_2 \tag{18}$$

$$y_7 = m_2 x_7 + y_2 - m_2 x_2 \tag{19}$$

$$y_6 = m_2 x_6 + y_T - m_2 x_T \tag{20}$$

$$y_7 = m_1 x_7 + y_T - m_1 x_T \tag{21}$$

$$m_1 x_6 - m_2 x_6 = y_T - m_2 x_T - y_2 + m_1 x_2 \tag{22}$$

$$x_6 = \frac{(m_1 x_2 - m_2 x_T + y_T - y_2)}{m_1 - m_2} \tag{23}$$

$$y_6 = m_1 (x_6 - x_2) + y_2 \tag{24}$$

$$m_2 x_7 - m_1 x_7 = y_T - m_1 x_T - y_2 + m_2 x_2 \tag{25}$$

$$x_7 = \frac{(m_2 x_2 - m_1 x_T + y_T - y_2)}{m_2 - m_1} \tag{26}$$

$$y_7 = m_2 (x_7 - x_2) + y_2 \tag{27}$$

The stereo distance calculation is done as shown on Fig. 3.

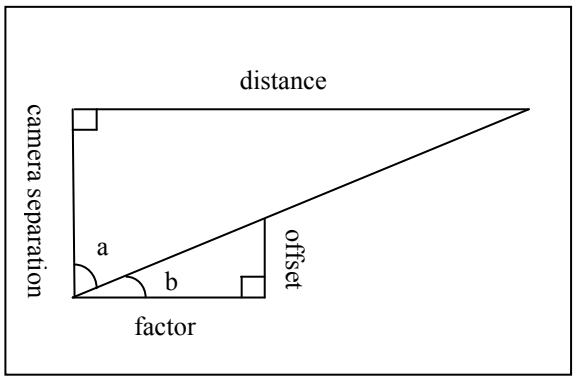


Fig. 3 Stereo distance calculation algorithm

$$tg(a) = \frac{distance}{camera\ separation} \tag{28}$$

$$tg(b) = \frac{offset}{factor} \tag{29}$$

$$a = \frac{\pi}{2} - b \tag{30}$$

$$offset = |x_{0R} - x_{0L}| \tag{31}$$

$$factor = \frac{offset}{tg\left(\frac{\pi}{2} - arctg\left(\frac{inital\ distance}{camera\ separation}\right)\right)} \tag{32}$$

$$offset = |x_{2R} - x_{2L}| \tag{33}$$

$$final\ dinstance = tg\left(\frac{\pi}{2} - arctg\left(\frac{offset}{factor}\right)\right) \cdot camera\ separation \tag{34}$$

The angle is calculated in the following way.

$$\alpha = \frac{180°}{\pi} \cdot arctg\left(\frac{tangent\ length}{radius\ length}\right) \tag{35}$$

The robotic values are calculated the following way.

$$robotic\ value = \alpha \cdot \frac{2000\ robotic\ values}{180°} \tag{36}$$

The real world coordinates were converted to pixels with the following formulas.

$$horizontal\ resolution = 320; \quad vertial\ resolution = 240; \quad diagonal = 19" \tag{37}$$

$$display\ size = 15.2" \times 11.4" = 21.05\ PPI \tag{38}$$

$$distance\ on\ display\ [pixels] = \\ = real\ world\ distance[cm] \times \frac{21.05\ [PPI]}{2.54\ [cm]} \qquad (39)$$

C. Software Implementation

The software was implemented in LabVIEW.

We chose LabVIEW, because as we can see on Fig. 4 this software environment has very friendly and rich UI (User Interface). It has also the possibility to program almost any interface of the PC, like serial port (RS-232), USB, parallel port, Ethernet, IEEE-488 or GPIB (General Purpose Interface Bus), Bluetooth and IrDA (Infrared Data Association). It has the libraries in the base package, so no additional libraries are needed to control the previously enumerated computer interfaces [3], [4].

It has also a very strong computer vision library which can control almost any camera on FireWire (IEEE-1394), USB 2.0, Gigabit Ethernet (GigE Vision) or Camera Link [1].

It has many computer vision functions which smartly combined can create a string computer vision application.

As we can see everything to control the robotic arm is integrated in one single application.

We have first the serial port configuration for commanding the robotic arm on RS-232 interface. We have 115200 baud rate, 8 data bits, no parity, 1 stop bit and no flow control. [1], [2]

We have 4 buttons to make to communication possible on the serial interface between the PC and the robotic arm. The first button will place the SSC-32 servo control board from the robotic arm in remote mode and will read its version. This buttons after it's pressed it will display the response in the indicator near it, we can see clearly that we have the SSC-32 board version 2.01. [1], [2]

The Initialize Motor button will test all the digital port from the servo control board and the motors it's self too. We have 32 digital port, but only 5 motors, from where in our application we use only 3 motors, but with this button we test all ports and all motors [1], [2].

With the All Servos 1500 we "wake" the robotic arm , we put it in the initial position which will look like the Greek capital gamma letter (Γ) [1], [2].

The move button will start the image acquisition and will overlay the lines and circles to the original images acquisitioned by the two webcams. It will also move the robotic arm in the desired position.

We can see that we have also two drop down boxes to choose the webcams, this way we can change between webcams if we have more than two connected to the PC or we can change their order or switch the right webcam with the left one.

In the middle we have the left and right images with the acquisitioned image of the robotic arm ant he overlays after the mathematical calculation. We have the color filtering for all for colors for left and for right.

We have also 24 sliders for each color filtering to set the color thresholds.

Finally we have the STOP button, because no LabVIEW application with loops should be left without it, this way we can get rid of problems like force quit and end task when we want to exit the application.

IV. Conclusion

As we could see we created a whole robotic arm control application all in one program.

The system work surprisingly well, but we know that nothing is perfect. We think about enhancements of it, maybe to create it on text based programming language too or to port it to another operating system, because our system works now in Windows operating system and in LabVIEW programming language.

We would also like to port it on an embedded hardware, this way the only solution could be and FPGA, maybe a Spartan-6 board like the Digilent ATLYS.

After this we could create even a layout of the chip and create our own ASIC (Application-Specific Integrated Circuit), which does the robotic arm control.

Maybe a best enhancement would be to extend to control of the number of joints, the best goal is to extend the number of controlled joints from 3 to 7, this way we can please even NASA's 7 joint robotic arm from the ISS (International Space Station).

The implementation solutions are endless, but the basic idea will stay the same, to use one or more cameras to detect the robotic arm's position, not just to know it by "blind" calculations.

References

[1] R. Szabó, I. Lie, "Automated Colored Object Sorting Application for Robotic Arms," *International Symposium on Electronics and Telecommunications (ISETC)*. Tenth Edition, 2012, pp. 95–98.

[2] R. Szabó, A. Gontean, I. Lie, "Smart Commanding of a Robotic Arm with Mouse and a Hexapod with Microcontroller," *18th International Conference on Soft Computing (MENDEL)*, 2012, pp. 476–481.

[3] R. Szabó, A. Gontean, I. Lie, M. Băbăiță, "Comparison between Agilent and National Instruments Functional Test Systems," *8th International Symposium on Intelligent Systems and Informatics (SISY)*, 2010, pp. 87–92.

[4] A. Gontean, R. Szabó, I. Lie, "LabVIEW Powered Remote Lab," *15th International Symposium for Design and Technology of Electronics Packages (SIITME)*, 2009, pp 335–340.

[5] *SSC-32 User Manual*. Available: http://www.lynxmotion.com/images/data/ssc-32.pdf, last visited April 20, 2013.

[6] Wong Guan Hao, Yap Yee Leck, Lim Chot Hun, "6-DOF PC-Based Robotic Arm (PC-ROBOARM) with efficient trajectory planning and speed control," *4th International Conference On Mechatronics (ICOM)*, 2011, pp. 1–7.

[7] Woosung Yang, Ji-Hun Bae, Yonghwan Oh, Nak Young Chong, Bum-Jae You, Sang-Rok Oh, "CPG based self-adapting multi-DOF robotic arm control," *International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 4236–4243.

[8] E. Oyama, T. Maeda, J.Q. Gan, E.M. Rosales, K.F. MacDorman, S. Tachi, A. Agah, "Inverse kinematics learning for robotic arms with fewer degrees of freedom by modular neural network systems,"

International Conference on Intelligent Robots and Systems (IROS), 2005, pp. 1791–1798.

[9] N. Ahuja, U.S. Banerjee, V.A. Darbhe, T.N. Mapara, A.D. Matkar, R.K. Nirmal, S. Balagopalan, "Computer controlled robotic arm," *16th IEEE Symposium on Computer-Based Medical Systems*, 2003, pp. 361–366.

[10] M.H. Liyanage, N. Krouglicof, R. Gosine, "Design and control of a high performance SCARA type robotic arm with rotary hydraulic actuators," *Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2009, pp. 827–832.

[11] M. C. Mulder, S. R. Malladi, "A minimum effort control algorithm for a cooperating sensor driven intelligent multi-jointed robotic arm," *Proceedings of the 30th IEEE Conference on Decision and Control*, vol. 2, pp. 1573–1578, 1991.

[12] M. H. Liyanage, N. Krouglicof, R. Gosine, "High speed electro-hydraulic actuator for a scara type robotic arm," *International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 470–476.

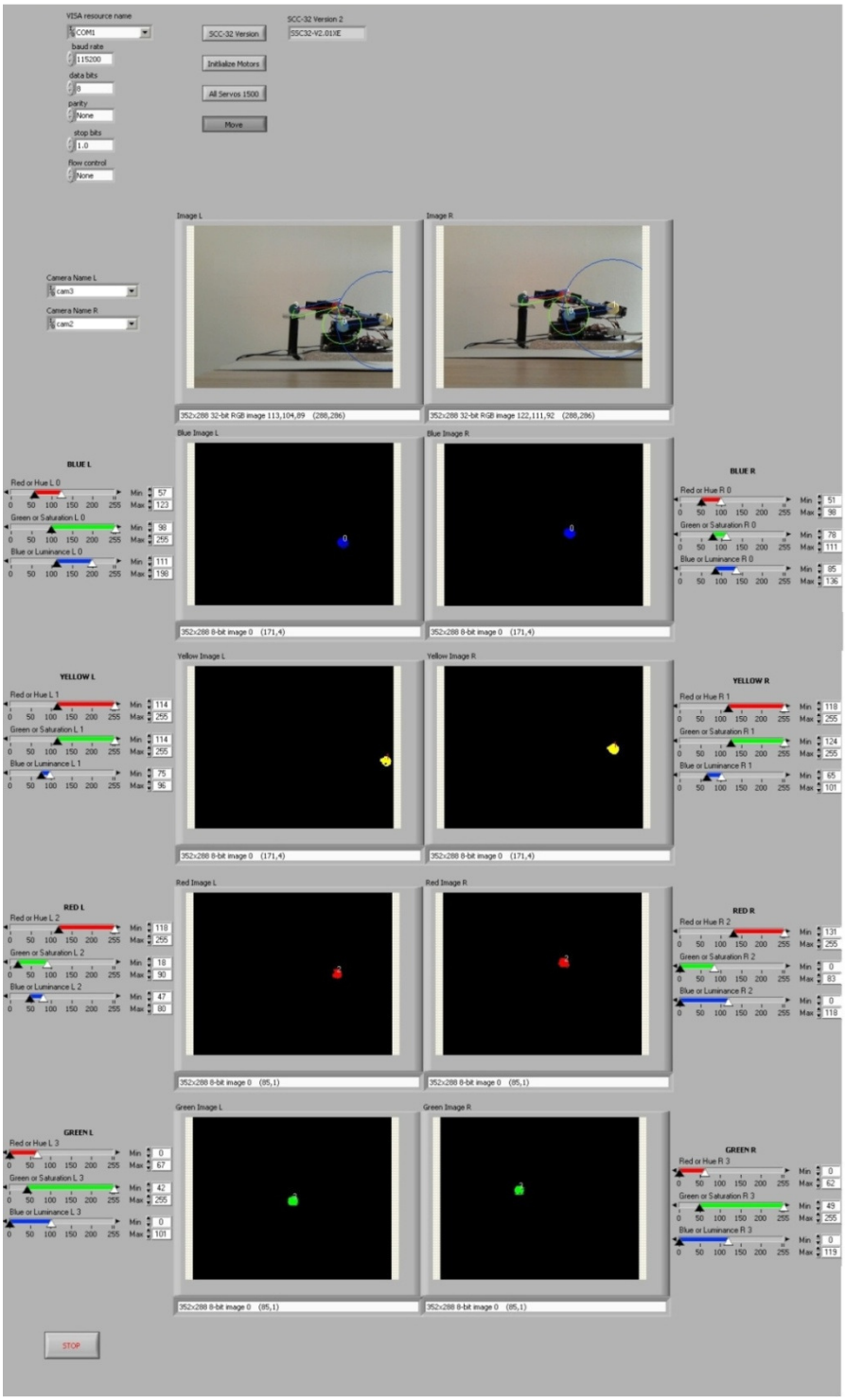


Fig. 4 The robotic arm control application made in LabVIEW

Frontiers in Network Applications, Network Systems and Web Services

SYMPOSIUM SoFAST-WS focuses on modern challenges and solutions in network systems, applications and service computing. The Symposium builds upon the success of Frontiers in Network Applications and Network Systems (FINANS'2012) and 4th International Symposium on Web Services (WSS' 2012) held in 2012 in Wroclaw, Poland. These two events are now integrated into one event to fully exploit the synergy of topics and cooperation of research groups.

The topics discussed during the symposium include different aspects of network systems, applications and service computing. The primary objective of the symposium is to bring together researchers and practitioners analyzing, developing and administering network systems, with particular emphasis on Internet systems. Authors are invited to submit their papers in English, presenting the results of original research or innovative practical applications in the field.

Topics

Topics include (but are not limited to):

- Architecture, scalability and security of Open API solutions,
- Technical and social aspects of Open API and open data,
- Service delivery platforms - architecture and applications,
- Telecommunication operators API exposition in Telco 2.0 model,
- The applications of intelligent techniques in network systems,
- Mobile applications,
- Network-based computing systems,
- Network and mobile GIS platforms and applications,
- Computer forensic,
- Network security,
- Anomaly and intrusion detection,
- Traffic classification algorithms and techniques,
- Network traffic engineering,
- High-speed network traffic processing,
- Heterogeneous cellular networks,
- Wireless communications,
- Security issues in Cloud Computing,
- Network aspects of Cloud Computing,
- Control of networks,
- Standards for Web services,
- Semantic Web services,
- Context-aware Web services,
- Composition approaches for Web services,
- Security of Web services,
- Software agents for Web services composition,
- Supporting SWS Deployment,
- Architectures for SWS Deployment,
- Applications of SWS to E-business and E-government,
- Supporting Enterprise Application Integration with SWS,
- SWS Conversational Protocols and Choreography,
- Ontologies and Languages for Service Description,
- Ontologies and Languages for Process Modeling,
- Foundations of Reasoning about Services and/or Processes,
- Composition of Semantic Web Services,
- Innovative network applications, systems and services.

Event Chairs

Furtak, Janusz, Military University of Technology, Poland

Grzenda, Maciej, Orange Labs Poland and Warsaw University of Technology, Poland

Legierski, Jarosław, Orange Labs Poland, Poland

Luckner, Marcin, Warsaw University of Technology, Poland

Szmit, Maciej, Orange Labs Poland, Poland

Program Committee

Afonso, Joao, Foundation for National Scientific Computing, Portugal

Baghdadi, Youcef, Sultan Qaboos University, Oman

Benslimane, Sidi Mohammed, University of Sidi Bel-Abbès, Algeria

Chainbi, Walid, ENISO, Tunisia

Chojnacki, Andrzej, Military University of Technology, Poland

Dabrowski, Andrzej, Warsaw University of Technology, Poland

Davies, John, Glyndwr University, United Kingdom

Fernández, Alberto, Universidad Rey Juan Carlos, Spain

Frankowski, Jacek, Orange Labs, Poland

Fuchs, Lothar, Institute for technical and scientific hydrology, Germany

Furtak, Janusz, Military University of Technology, Poland

Gaaloul, Walid, Institut Mines Télécom, France

García-Domínguez, Antonio, University of Cádiz, Spain

García-Osorio, César, University of Burgos, Spain

Grabowski, Sebastian, Orange Labs, Poland

Kiedrowicz, Maciej, Military University of Technology, Poland

Korbel, Piotr, Technical University of Lodz, Poland

Kowalczyk, Emil, Orange Labs, Poland

Kowalski, Andrzej, Orange Labs, Poland

López Nores, Martín, University of Vigo, Spain

Macukow, Bohdan, Warsaw University of Technology, Poland

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 45–50

USSD communication channel as alternative to XML SOAP in mobile Unified Communication applications

Dariusz Bogusz
Siemens Enterprise Communications*
ul. Mińska 63A
03-828 Warsaw, Poland
Email:
dariusz.bogusz@siemens-enterprise.com

Grzegorz Siewruk
Warsaw University of Technology
Faculty of Electronics and Information Technology
ul. Nowowiejska 15/19
00-665 Warsaw, Poland
Email:gsiewruk@gmail.com

Jarosław Legierski
Orange Labs
ul. Obrzeżna 7,
02-691 Warsaw, Poland
Email: jaroslaw.legierski@orange.com

Jan S. Kunicki
University of Warmia and Mazury
Faculty of Mathematics
and Computer Science
Słoneczna 54 Street
10-710 Olsztyn
Email:
jan.kunicki@matman.uwm.edu.pl

Abstract—**Traditional mobile Unified Communications (UC) clients use data channel over GSM/UMTS network for communication with UC services broker and providers. UC signaling traffic characteristics require a large number of the short data connection sessions and can cause overload of the mobile network. The reason is a ratio of the overhead data to the useful signaling information.**

This paper presents an analysis of the mobile UC architectures and a possible use of Unstructured Supplementary Service Data (USSD) channel for UC application in the mobile domain. The usage of USSD channel can help to provide more effective applications for the mobile UC user.

I. Introduction

UNIFIED Communications is a new trend on the enterprise market. The basic idea is to handle all communications of a business user (voice call, video, instant messaging, teleconferencing, e-mails, voice mails, web collaboration, mobility and customer contact center etc.) over a single scalable, software-based, unifying communications platform working with any IT, voice, and application environment. Some of the key features are:

- one universal telephone number independent on a physical device (desk phone, soft phone, home office, mobile),
- presence management - including personal and media status,
- mobility - including virtualization of the communication services irrespective of a client (softphone on PC, browser based client, personal voice based client, mobile smartphone client),
- communication enabled business processes CEBP – multimedia communication built into the enterprise business applications via API or the available plugins.
- social media integration – possibility to communicate via Web 2.0 portals in the Cloud via API or federation

Architecture of a typical UC system is presented in Fig. 1.

Proliferation of the smart mobile terminals allowed to offer UC users a dedicated mobile application that significantly increases functionality of the office communication systems for the business end users. Integration of the UC applications with other systems [1], [2] allows to offer a complete end extensible environment. In the next sections we discuss evolution of the architectural approach to the mobile UC, we also suggest a possible optimization in this area and present a prototype implementation.

II. Mobile UC Applications Evolution

Mobility of the end users is a very important aspect for many enterprises. This section presents evolution of the enterprise mobile applications architectures based on the solutions available on the market during the last 20 years.

A. Hardware DECT/GSM system

The first mobile application was announced in 1995 as a joint trial development project between TELIASONERA AB; ERICSSON Inc. [6]. Ericsson was to deliver

5,000 dual-mode GSM/DECT handsets, which should operate over both wireless networks. This concept was beneficial to the telecommunications providers due to better utilization of RF, and to the customers because of better radio coverage and maintaining in-campus mobility.

However such a system requires a specialized terminal (hardware) capable of using DECT and GSM transmission (GSM means here GSM900 as well as DCS1800) [7]. Such terminal must be able to maintain two concurrent communication sessions; one over GSM and the other over DECT. Additionally it should be available over one personal number irrespective of the terminal location; via DECT in campus and via GSM outside. Intelligence is located in the network infrastructure, no intelligence in the terminal.

With all these advantages this model requires a specialized dual-mode GSM/DECT handset. This in turn increases price of the whole solution and diminishes battery life. Some vendors tried to use another approach.

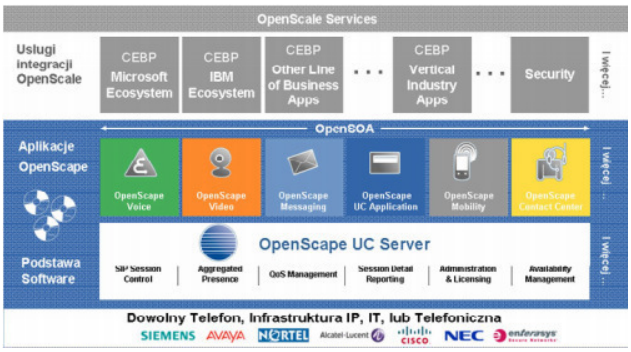


Fig. 1 OpenScape UC system architecture

B. Corporate GSM

The idea of corporate GSM has been presented by many vendors. The first fully developed architecture named DWOS, was presented by Ericsson[8]. Implementation of the critical functionality is split between the enterprise and telecommunications provider core infrastructure [14]. This unfortunately requires many integrations between the provider and enterprise telecommunication in both organisations and in the infrastructure. The market has not developed any appropriate business model and such solutions have not been widely adopted.

C. OpenScape UC Mobile Client

By the year 2002 the first Unified Communications systems were announced to the market [12]. They promised many advanced features irrespective of the user location, administrative domain, or access network technology. One of the first UC systems in the market was Openscape UC issued by 2004 [13]. In this paper we cover the mobility aspect of this solution. Openscape UC mobile Client is mainly CTI (Computer Telephony Integration) client extensively using WiFi or GSM Data channel for UC signaling traffic. The payload (voice only) to the mobile client is transmitted over a standard GSM channel. GUI of this client is a simple implementation of UI for managing Web Services responsible for the advanced UC features. Openscape UC mobile Client has no built in intelligence, which could allow us to name it an agent system.

D. Mobile Connect

At CEBIT 2006 Siemens presented a prototype, and at CEBIT 2007 the commercial product of a different Architecture, commercially named HiPath Mobile Connect. Mobile Connect is a solution based on a dual connectivity (WiFi and GSM) of the commercially available mobile phones [9].

The main architectural difference is the application server called FMC Appliance, located within the enterprise's IP network. It is fully managed by the Enterprise. As the contractual relationships do not go beyond a standard commercial contract between the enterprise and telecommunications provider, such system is commercially much easier to deploy. Thus much wider adoption of this model. With a lack of information from the mobile provider, a network implementation of the advanced services like seamless handover between enterprise WiFi and public GSM is implemented partially as FMC client function. It means that more intelligence is moved to the endpoint. This architecture implements for the first time FMC client as an intelligent agent system interworking with FMC Controller through implementation of its advanced functions.

E. OpenScape MObile

Openscape OSMO (OpenScape MObile) is a client combining rich UC functionality of OpenScape UC Mobile Client with SIP client implementation on a smartphone [15]. Though such functionality is beneficial for the users it poses many challenges to the network.

OSMO client maintains two parallel connections:

- CTI connection to a UC proxy named OpenScape Facade
- Payload and SIP Signalling connection to an enterprise Session Border Controller (SBC) named OpenScape SBC

This architecture is necessary (despite its complications), because SIP is a standard protocol. Standard voice and video functions are implemented within SIP signalling accessing the enterprise net via SBC. All non standard UC protocols are transmitted via data channel to a UC proxy. Such architecture with split of functions requires the client on a smartphone to perform as an agent. User interface of UC mobile client OSMO is presented in Fig. 2.

Fig. 2 OpenScape OSMO GUI on iPhone

III. Mobile UC Signaling Traffic

A. Data traffic from UC mobile application

Signalling traffic characteristics of such an agent in mobile domain creates an interesting problem with many practical implications for the mobile provider in the signalling and data path. They are also important from the enterprise security and availability perspective. Issues resulting from integration of those two worlds are planned for further investigation. Here we cover only some aspects of it.

Typical Unified Communication mobile user uses the application installed in his mobile phone to:

- Change personal status (available, free, out of the office on holiday, etc.)
- Change preferred device (office phone, mobile phone , home phone, etc.)
- Read phone log (answered/ not answered calls, connections redirected to another device, etc.)
- Read telephone book records (global and personal address books)

The most frequently used function focuses on the change of personal status of a user. The end users can change their status using OpenScape Mobile application. Request of the status change is transmitted to Unified Communication server as Web Service call using XML data and SOAP [3] protocol. The connection is established as data channel via IP network. Using this method every status change results in transmission of 4890 bytes in 4 TCP packets (without TCP/IP packet headers).

Information of Unified Communication subscriber status change can be transmitted using 3-5 digits USSD code. Fig. 4 presents USSD communication captured using Wireshark in SIGTRAN (SS7 over IP protocol) MAP messages. The amount of data sent using USSD (without packets headers) is about 632 bytes in SCTP protocol in both directions (between USSD gateway – server exposed API).

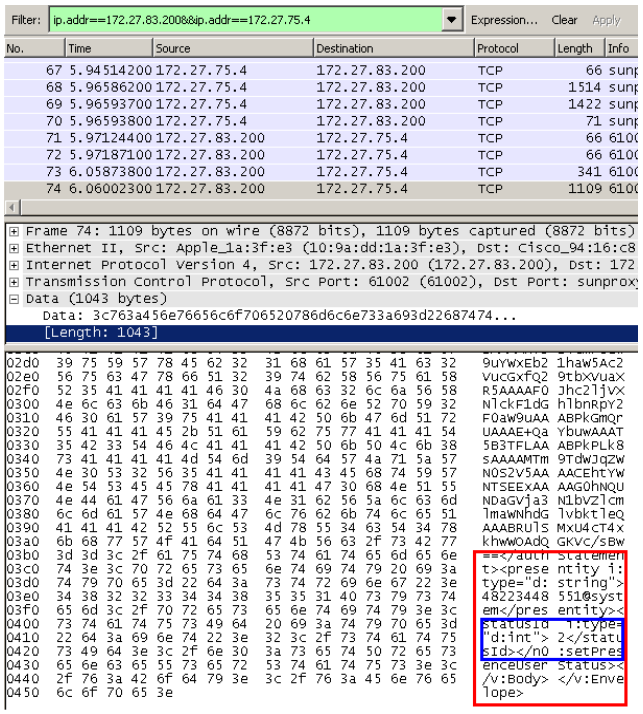


Fig. 3 TCP packet with user status change

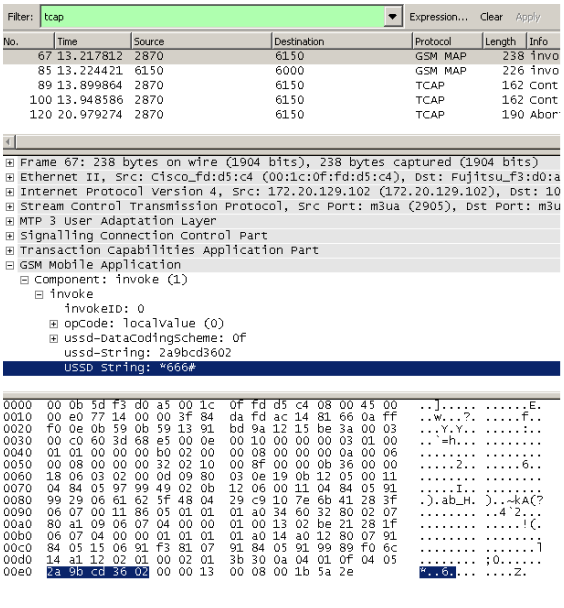


Fig. 4 TCP packet with USSD code

Fig 5 presents typical usage of OSMO application based on the application logs recorded at a mobile phone.

From the data presented at Fig 5 we notice that the number of communication requests per hour per user ranges between 1 and 21. Average value is more than 6 (6,42). Technically UC system can grow up to 100 000[13] users. By a conservative assumption of 1000 users/enterprise we come to 6,42*1000=6420 requests per hour. The observed application traffic characteristics re-

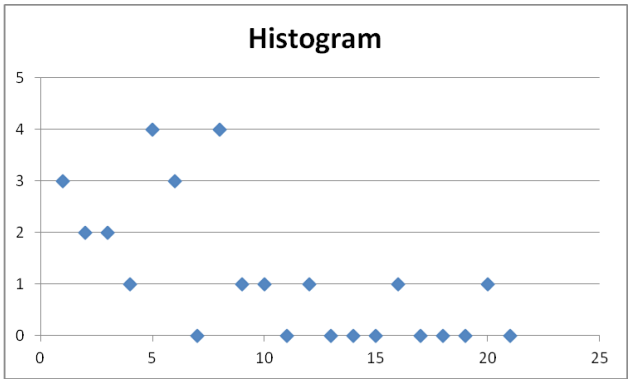


Fig. 5 Histogram of personal status change requests in time for OSMO UC application. x-req/h, y-frequency over 48h

sults in a large number of short data connections over the mobile network (GSM or UMTS). The number of valid data (one Integer value) transferred in TCP/IP packets (Fig. 3) is very small in comparison to the amount of other data (TCP/IP packet headers, XML headers etc.).

We have also expected this data to exhibit more distinctively other structures like periodic and non-periodic components, etc., important for better understanding of this traffic nature. These however did not show with the data presented at Fig. 5. Such intriguing issues are to be subject to our further investigations.

B. USSD communication channel

Based on the information presented in section II, we can conclude that using the data channel for communication Mobile Application – UC system is not an optimal solution. The reason is the amount of the signalling data in relation to the payload data and the large number of a short time data connections (Fig. 5). Our concept focuses on using the USSD channel (rather than data-over-Internet) for communication between the mobile application and the UC server.

Unstructured Supplementary Service Data (USSD) is a communication protocol implemented in the popular mobile networks (GSM, UMTS). The main role of USSD is to provide the two sided communication between a mobile phone and various applications and the system on the telecommunications provider side.

The first USSD specification (Phase 1), described in GSM 02.90 documentation [10] supports only the communication initiated from a mobile phone – pull mode (using dedicated USSD codes e.g. *665*11#, *100*# etc.) The Second Phase specified in GSM 03.90 [11] supports both communication methods originated by the network (push) and terminal (pull). The network initiated communications allow sending information to the mobile terminal.

From a functionality point of view the USSD message contains up to 182 characters. Compared to Short Message Service (SMS), USSD is a real-time transmission using the signalling channel of the mobile network.

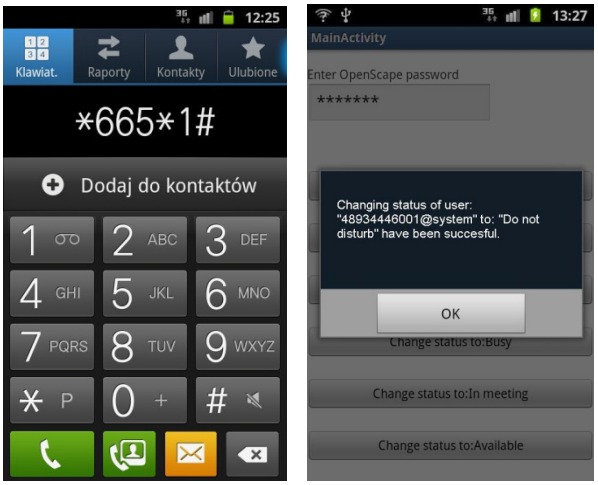


Fig. 6 Mobile phone originated USSD (left) and network originated USSD message (right)

In the mobile provider core network, USSD communication is implemented using MAP (Mobile Application Part) protocol. MAP is an application layer protocol dedicated for the mobile networks based on SS7 signalling stack. The USSD communication is supported using the following MAP messages:

- MAP_PROCESS_UNSTRUCTURED_SS_REQUEST – pull operation between terminal (originator) and application server,
- MAP_UNSTRUCTURED_SS_REQUEST- push USSD messages,
- MAP_UNSTRUCTURED_SS_NOTIFY – USSD - push operation.

USSD offers session-oriented connections and this functionality allows to develop applications supporting dialogue with the end user (e.g. using interactive menu displayed on a mobile terminal).

IV. SYSTEM ARCHITECTURE

The solution presented in this paper was implemented as a low cost application for UC mobile end user. Therefore, we expect it to be highly effective. The high level system architecture is presented at Fig. 7. In order to create a scalable service in line with the latest trend in the field of mobile services, our architecture consists of:

- OpenScape UC server – actual UC services provider
- Web server, which processes USSD requests to UC API
- Client - on user smartphone for operations like: login or changing users presence status

This section presents the implementation and configuration of the client endpoint and software architecture of USSDforUC environment.

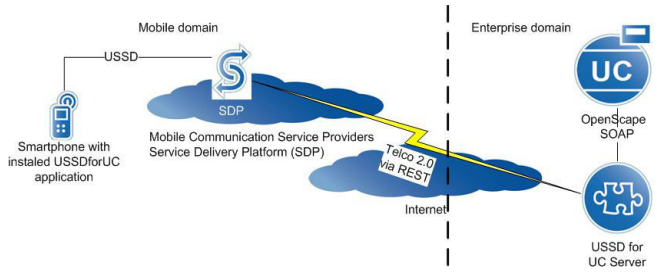


Fig. 7 USSDforUC High level system architecture

As presented at Fig. 8, *USSD* message sent by the user is redirected via provider's API to the USSDforUC application server which communicates using SOAP Web Services API with Open Scape UC server. Both communications: USSD and communication with UC server are created within the request-response model using Web Services. Because of the session timeout implemented in USSD, the USSDforUC server must send response to the particular request during the same communication session and in a limited time. The USSDforUC was developed as an application without its own database and the information about UC users and their settings (passwords, telephone numbers) is stored in Openscape UC repository.

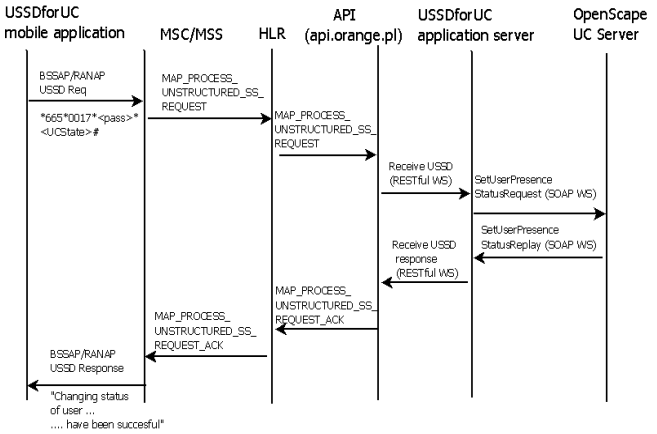


Fig. 8 APIs and telecommunications protocols request/response flow. MSC – Mobile Switching Center, MSS – Mobile Softswitch Solution, HLR – Home Location Register

A. USSDforUC server

This section describes configuration and implementation of USSDforUC server which translates USSD message from expressed in Internet provider's APIs (based on Restful Web Services [4]) to SOAP messages. The main objectives executed by this module are:

- login to OpenScape UC server using UC API,
- creation object represents OpenScape user profile in the memory,
- changing OpenScape user presence status to the one of the list: Do not disturb, Be right back, Unavailable, Busy, In meeting, Available.

User profiles are stored in the memory in a way which does not overload the system. Only the necessary data like login name and mobile phone number are saved in the profile. In order to maintain user base correctness the array of users is refreshed every 2 hours.

Changing the user presence status is more complex. The architecture of OpenScape UC server does not allow to use the technical account dedicated for API calls to change the particular user presence status. The only solution is to login via API as the user who wants to change his or her status and to complete the task from that level. Information about the user login and password is coded in USSD request as a special code. UC user account is recognized using MSISDN number from USSD request. Server looks for MSISDN in the array of profiles stored in its memory in order to login to the user account and change status.

B. USSD mobile application

This application has been developed as a mobile application for Android Operating System.

Application itself is created as a single activity (android screen) which contains a single text field and six buttons. To complete the first request it is mandatory to enter password before clicking any button. Otherwise the user will be provided with an error message. The USSDforUC application is actually using only one Android SDK function: 'MakeCall', with one parameter i.e. telephone number. API call requires the phone to set up connection with the number it gets. In our case the number is the USSD request in form of *665*17*code*, where:

- 665 is the USSD server id,
- 17 is application id (USSD server is aware of the URL it should redirect USSD messages to, if they come with particular id),
- code which contain the mash up of password and status code.

C. Security aspects

As is mentioned in section *B*, the message which contains the user password is sent through an unsecured USSD channel. Because of a simple USSD coding in SIGTRAN messages (Fig. 8) use of any network sniffer enables to read the passwords. To increase the security, we have used a simple encryption method. Every special character sign of the code page is replaced by a number. This method is not strong one, but this addition results in a random generated sequence of alphabet letters and is hard to decode for a person not knowing the key. To harden against a dictionary attack (trying to login on account with each combination of letters) we have implemented a module which counts the unsuccessful login attempts, adds them to server log and disables a user account for 1h hour.

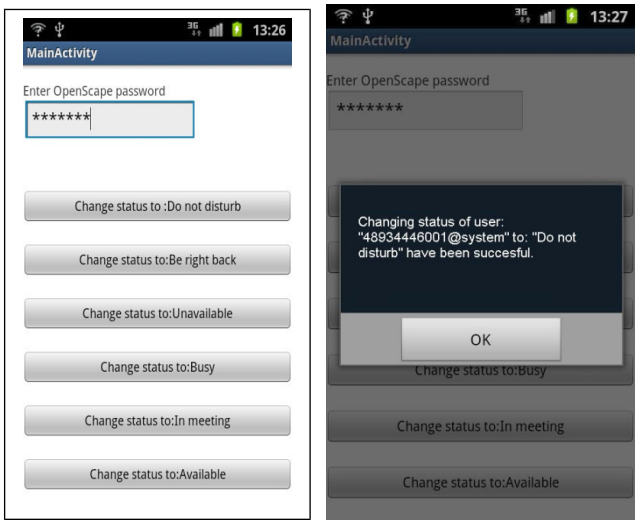


Fig. 9 USSDforUC mobile application GUI on Android

V. SUMMARY

This paper presents the traffic characteristics of a modern mobile UC user. It discusses the possible drawbacks of transmitting UC signaling traffic over IP in the mobile domain as well as the possible use of USSD channel in UC mobile application including a prototype implementation.

USSD communication allows bidirectional, half-duplex connections between the mobile terminal and application. This communication is completely transparent for the handset and mobile network. USSD can be used by any type of a mobile phone and does not depend on the vendor and operating system.

Using USSD channel for signaling, mobile UC application can avoid the large number of short time data connections. Because in the core provider's network USSD messages are coded using binary protocol the single UC status change request is transmitted using about 632 bytes as compared to 4890 bytes transmitted in the current solution - SOAP and XML based Web Services call in data traffic.

An interesting feature of the presented USSDforUC prototype is that the implemented system doesn't use any database what simplifies its overall system architecture.

Using USSD for UC signaling traffic shows that even the leading commercially available UC solutions have areas where they can be vastly improved.

Characteristics of UC signaling traffic in the mobile domain as well as on the interface to other ICT systems in the enterprise, telecommunications provider as well as in the public cloud will be subject of further investigation. The research area is to model and manage this traffic to avoid communication bottlenecks and to manage user experience.

ACKNOWLEDGMENT

Prototype of USSDforUC service was developed under the program Orange Labs Open Middleware 2.0 Community [5].

*) Siemens Enterprise Communications Sp. z o.o. is a Trademark Licensee of Siemens AG

REFERENCES

[1] D.Bogusz, J.Legierski, A. Podziewski, K. Litwiniuk, „Telco 2.0 for UC – an example of integration telecommunications service provider's SDP with enterprise UC system", Wrocław 2012. Results of operation in 2010, Warsaw 2011

[2] D. Bogusz, P. Korbel, J. Legierski, Integracja systemów Unified Communications z platformami usługowymi operatorów, KSTiT 2011 conferencing materials, Przegląd Telekomunikacyjny 8-9/2011

[3] SOAP specification W3C, http://www.w3.org/TR/soap/ [5.03.2013]

[4] L. Richardson, Sam Ruby, David Heinemeier Hansson, RESTful Web Services, O'Reilly, 2007

[5] Open Middleware 2.0 Community portal – http://www.openmiddleware.pl

[6] "Telia, Ericsson plan dual-mode, GSM/DECT", Electronic News (10616624);4/17/95, Vol. 41 Issue 2061, p28, April 1995

[7] Patric Lind, "Terminal aspects of DECT/GSM as an evolution towards UMTS", http://cordis.europa.eu/infowin/acts/analysys/products/thematic/mobile/ev_cellular_sys/dect_gsm.html

[8] Rune Johanson, Mikael Nilsson and Torbjorn Ward,"Mobile Advantage Wireless Office—A digital wireless office system for TDMA/136 networks", Ericsson Review No. 1,1999 19

[9] "Demystifying Enterprise Fixed Mobile Convergence", A white paper from Siemens Enterprise Communications, February 2007

[10] Unstructured Supplementary Service Data (USSD); Stage 1 3GPP specification: 02.90 www.3gpp.org/ftp/Specs/html-info/0290.htm

[11] Unstructured Supplementary Service Data (USSD); Stage 2 3GPP specification: 03.90 http://www.3gpp.org/ftp/Specs/html-info/0390.htm

[12] "Siemens ICN Announces LifeWorks Concept for Integrating Personal, Enterprise and Carrier Technologies", http://www.prnewswire.-com/news-releases-test/siemens-icn-announces-lifeworks-concept-for-integrating-personal-enterprise-and-carrier-technologies-71170057.html

[13] "Siemens Redefines Communication Over IP Market With Announcement of HiPath(TM) 8000 Real-Time IP System for Very Large Enterprises", http://www.prnewswire.com/news-releases/siemens-redefines-communication-over-ip-market-with-announcement-of-hipathtm-8000-real-time-ip-system-for-very-large-enterprises-72780072.html

[14] "HiPath CorporateConnect V2.0, CorporateConnect Server", Service Guide A31003-G3620-S300-4-7620, by Siemens AG 04/2006

[15] "OpenScape Mobile V7R1, Administrator Documentation", A31003-M4070-M100-8-76A9, by Siemens Enterprise Communications, 12/2012

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 51–55

Metering devices remote monitoring system Web-interface developing

Alexey Lagunov
Northen (Arctic) Federal University
named after M.V. Lomonosov
Severnaya Dvina Emb. 17,
Arkhangelsk, Russia, 163002
Email: a.lagunov@narfu.ru
http://www.narfu.ru

Dmitry Fedin
Northen (Arctic) Federal University
named after M.V. Lomonosov
Severnaya Dvina Emb. 17,
Arkhangelsk, Russia, 163002
Email: d.fedin@narfu.ru
http://www.narfu.ru

Nikita Gerasimov
Northen (Arctic) Federal University
named after M.V. Lomonosov
Severnaya Dvina Emb. 17,
Arkhangelsk, Russia, 163002
Email: n.gerasimov@narfu.ru
http://www.narfu.ru

***Abstract*— Recently appeared necessity of enterprises, homes and other objects energy consumption information automatic collection. For this purposes various metering devices by different vendors have been created. While using different metering devices, problem of central information collecting appears. This paper raises the problem of creating a web-interface for metering devices remote monitoring system for data collection from the various sensors and meters.**

I. Introduction

QUANTITY of sensors by different vendors, collecting different information leads to different meters managing software packages quantity. For this problem solution was required universal hardware-software system, which would can connect to any metering device, get every data, write it in to a database and show in user-friendly format. Common metering devices remote monitoring system scheme presented at “Fig. 1”.

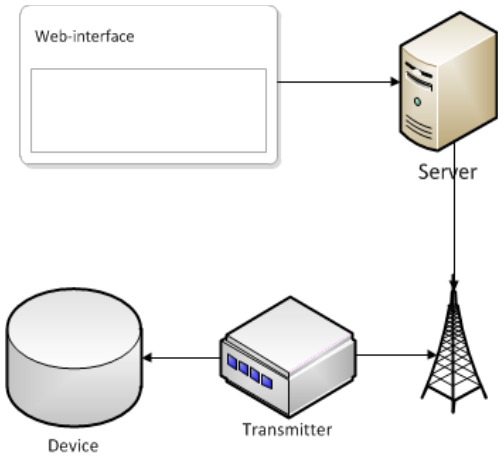


Figure 1. Metering devices remote monitoring system common scheme

Read out by metering device, data is transmitted to the server for following saving to DB. After these actions it can be shown to the customer in user-friendly view. In our case user-friendly format is provided by web-interface. Web-interface is adjusted such way to have ability to show every data types, collected from every sensor. Common web-interface structure is at “Fig.2”.

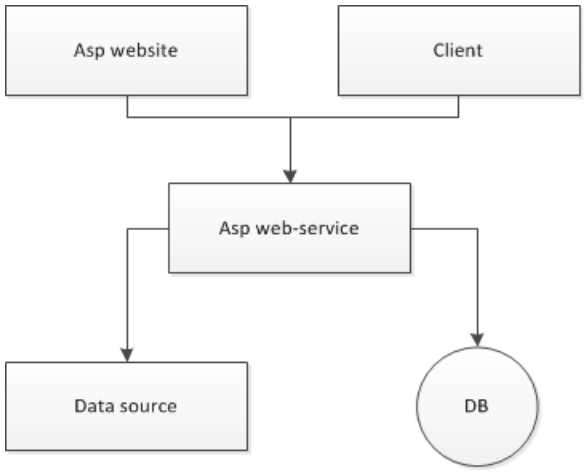


Figure 2. Web interface common model scheme

We decided to use ASP.NET framework because developing simplicity, reliability. ASP.NET framework provides all required tools for creating web-portals: implementation of SOAP REST protocol for web-service and web-site communication, database drivers. Also SAP provides simple reports generation integration. Other variants: PHP or Java requires manual implementation of above-mentioned functionality [1].

II. Device

Modern equipment installed in the pipe rooms, allows remotely collect data from the meters using wired and wireless communication channels. The majority of the existing systems is focused on a specific manufacturer of the energy counter. Many organizations have counters of different manufacturers and different modifications. Each manufacturer has its own data collection system, often not compatible with the systems of other manufacturer. The device allows to collect the data from the meters. Transfer of the received data is carried out through GPRS. The device includes the printing cost, which contains a power unit, a sim card, slot antennas, signal RS-232 v converter, signal CAN interface converter, on the circuit Board is additionally available programmable GSM module, independent RS-232 and CAN interfaces.

Performance indicators:

- ambient air temperature from -20 to +50'c;

- relative humidity 95% at 35 C;
- power supply 12 Volts +- 30%, 50 Hz;
- work mode - continuous;
- validity period – 3 years.

III. Web-interface Database Developing

A. Object structure scheme

For simplicity of object geographical managing and structuration ease we decided to allocate all objects, including geographical zones and meters, to different levels with private proprieties and inheritance ability. Common indication storing database diagram is presented at "Fig. 3".

Such scheme allows creating a root level, representing some geographical object with successor: another geographical object or metering device. First level always represent meter. Every object has its own unique proprieties, user privileges. Such scheme allows describing almost every structure, allows managing it conveniently, allows simply define every device.

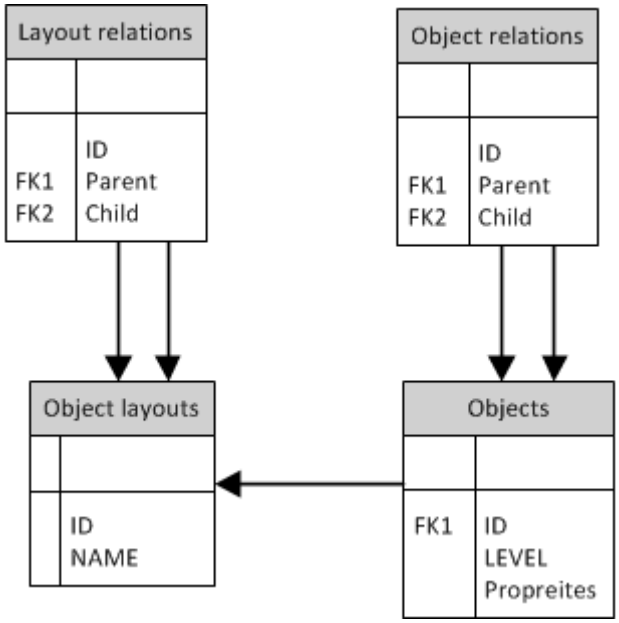


Figure 3. Object information storing database scheme

B. Data storing scheme

Collected data is storing in such database structure, presented at "Fig. 4":

- ID
- Data type ID
- Object ID
- Date and time
- Data relevancy duration in hours
- Applied function
- Value

Thus in the database data with data and relevancy period and applied function are storing. The "Applied function" field describes is the data average or total for some time period. Such structure allows storing collected information and saving user reports calculation results, e.g. average daily results. After collection data is being processed to get average daily results, is sending to a customer and saving in to the database with such structure:

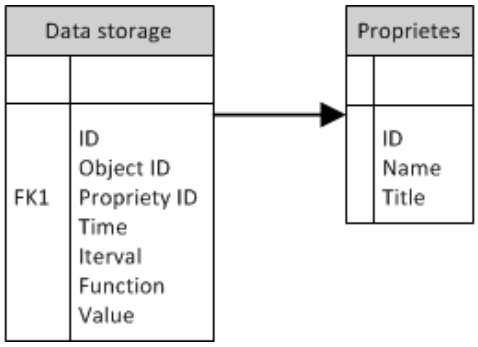


Figure 4. Collected data storing database scheme

- Data type ID
- Object Id
- Date with time equal 00:00
- Relevancy duration equal 24 hours
- Applied function - average
- Value

At first time system doesn't calculate any information in such way, database stores only information, transferred from metering device. In future will be realized algorithm, calculating required data

First variant was based on data division on several tables by period, but such way turned out to be inflexible.

Firstly our system contained every object info entry in specialized table field, but after level-oriented structure introduction, all information stores in united table, which containing object identifier, field name and field value: "Fig. 5".

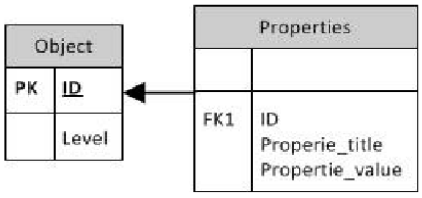


Figure 5. Properties storing.

As alternative to such table, arrays can be used, but .NET PostgreSQL drivers doesn't allow to use this feature.

Other database scheme part is rather typical for web-application: tables for user information and settings storing.

For data storing we choose DBMS PostgreSQL 9.x. Such decision was made for PostgreSQL high efficiency, high reliability and functionality. PostgreSQL allowed us to devide different databases to different schemes, also PostgreSQL allows to easily DB ranging with tables partitioning [2].

IV. ASP.NET WEB-Service Developing

Web-site is developing with ASP.NET technologies set.

Web-site schematic view is presented at "Fig. 6". Prerelease web-site views are at "Fig. 7" and "Fig. 8".

For widget system creation we use Dropthings framework, which requires only quick widget creation for web-portal creation.

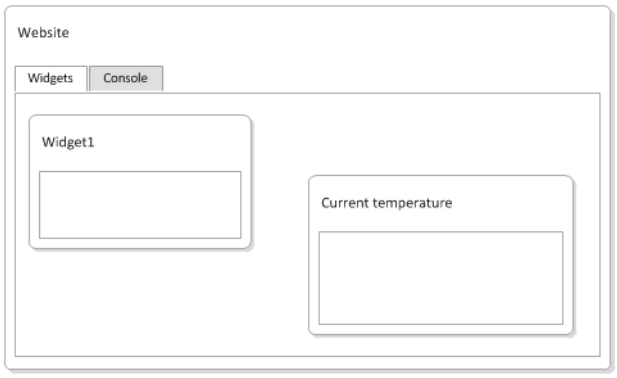


Figure 6. Common web-site scheme

"Widgets" tab at common scheme contents set of various configured and arbitrary placed widgets. Widgets quantity is limited only by self sizes and website workspace size. Every widget can be created several times at one tab workspace with individual settings [3].

Widgets examples:

- Last data, collected from specified device.
- Maps with geographical and/or meters location.
- Indicator of deviation of specified metering device collected information by type and period

Console allows to:

- Create typical reports for different devices or device groups energy consumption
- Manage data collection time intervals
- Etc.

Console is also developed as widget, despite the widget usually executes tiny procedure.

While web-site is developed with Dropthings framework, web-service is developed using common MVC pattern. Scheme is presented at "Fig. 8". United Service class is created for more simple SOAP client using at the web-site side, also it is "View" component. When user asks data for object (city, street or metering device), controller creates this instance of "ObjectInfo" class, that provides self-calculation. After object instance determinates: is it high-level object or counter, it creates instances of child-objects or instance of value provider. If level is high, instance asks children to self-calculate else asks specified value provider to self-calculate. After all data is got, instance calculates it, gets it reliability and returns to controller or high-executed instance. Better to say, it is recursion. Reliability is created for data integrity mark, it is also calculating in value and object provider.

For reporting system we decided to use CrystalReports for it functionality and free MS Visual Studio version. As CrystalReports requires previously defined database structure, but out targets requires to create new reports without recompilation we developed several ways to make datasource for reporting system. Both based on one DataSet table, containing one field. First using parameter-based Crystal Report and dynamic DataSet. Second using XML document for creation Crystal Document and also dynamic Dataset. For simplicity we using second variant. Using report sample we decide needed values, then form special XML document, containing DataSet structure. Using this this XML document Crystal Report is creating and registering at the system. Then, when user asks system for the report, web-site generates DataSet with structure, similar to XML file, fills it with data and executes CrystalReports to generate the report.

While using reports system, we met one more trouble: using Dropthings and CrystalReports with advanced documents led to JavaScript errors. This with solved with blank template.

V. Figures and Tables ASP.NET Website and ASP.NET Webservice Communication

For website and web-service communication is used SOAP protocol, which is fully supported by MS .NET framework.

Information about users, user's sessions, user's success and failed logins is stored at database. While authentication process client sends "login" and "password" pair to the web

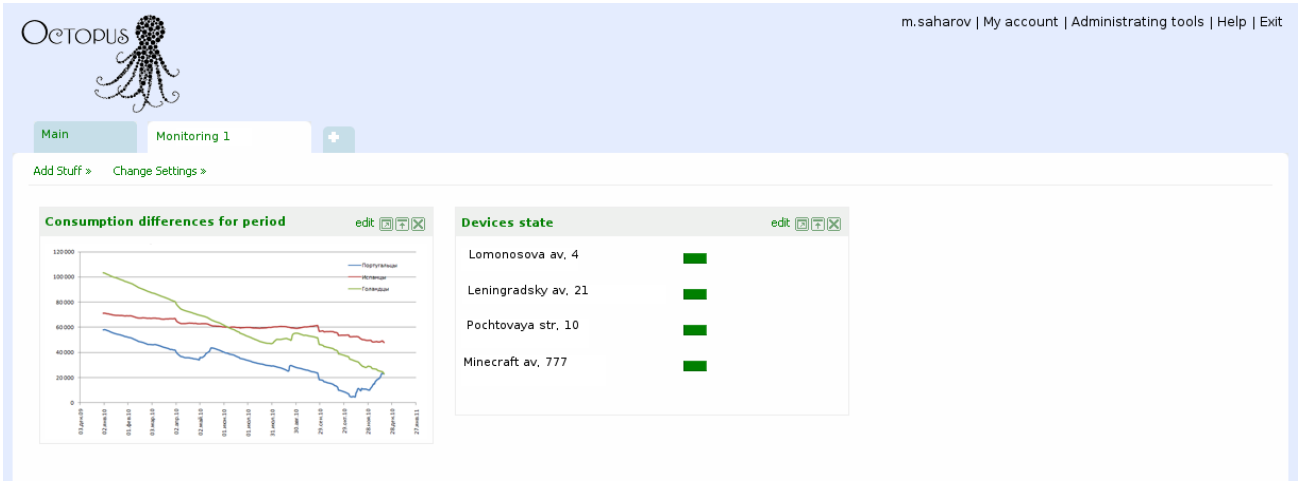


Figure 7. Widgets

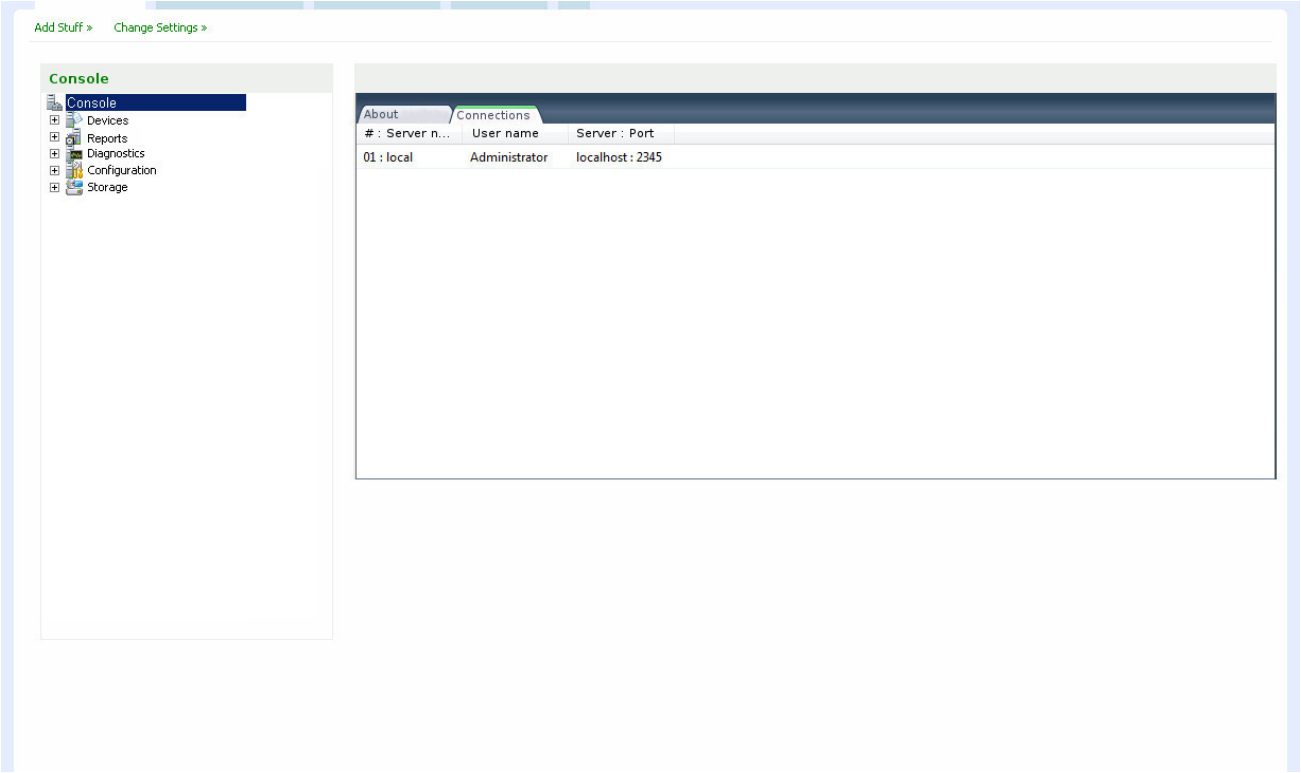


Figure 8. Console

service, who writes session information, session duration and session key in to the DB. Session key is sending with reply to the website. In case of invalid login-password pair client gets "SOAP Exception" message and web services writes information about failed authentication attempt.

After successful authentication, while new .NET session is starting, client sets saved session key, using special web method. Actual session key sets one time per .NET session.

In case of every fail execution client gets SOAP Exception message with details.

VI. Conclusion

System which can provide to the customer previously processed, user friendly consumption information, can save processing results was created at result. Such system allows working with various device types by various vendors and with various data types. This system has a low threshold of entrance, it is easily set up, it is informative for all categories of customers. It doesn't require any installation and setup.

References

[1] Matthew MacDonald, Mario Szpuszta, "Pro ASP.NET 3.5 in C# 2008", ISBN: 1590598938, November 15, 2007
[2] Hannu Krosing, Kirk Roybal, "PostgreSQL Server Programming" ISBN: 1849516987, June 2013
[3] Bill Scott, Theresa Neil, "Designing Web Interfaces", ISBN: 978-0-596-51625-3, 2009

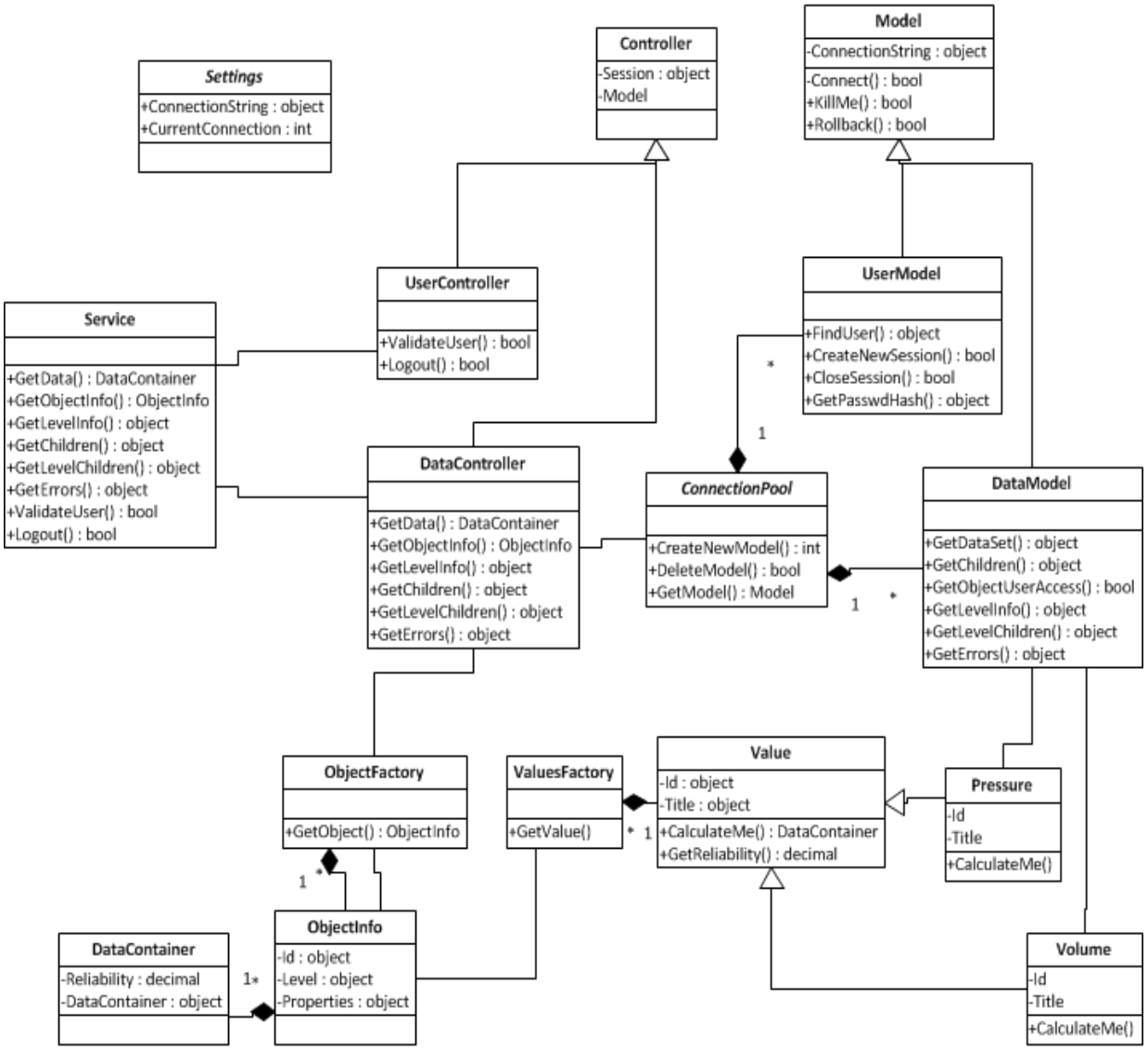


Figure 9. Web-service scheme

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 57–62

A Middleware Architecture for Mobile Social Networking with Intelligent Energy Saving

Luiz M. M. de A. Santos, Maria A. S. N. Nunes, Admilson de R. L. Ribeiro
Computing Departament
Universidade Federal de Sergipe, UFS
São Cristóvão, Sergipe, Brazil
Email: {luizm1000, gutanunes}@gmail.com, admilson@ufs.br

Abstract—In recent years, middleware for social mobile networking has attracted the attention of academia, causing the design and development of various approaches by researchers. This type of middleware facilitates and makes more efficient the development process of mobile social networking applications. Furthermore, middleware solutions also abstract the communication process with other applications, allowing the acquisition, persistence and reuse of social context information and location of users, besides providing an API so that developers can access this information quickly and build new social applications. With a view to supporting this new trend of research, in this paper we propose a middleware architecture for mobile social networking called My-Direct which will make use of Wi-Fi Direct as a solution to improve the connectivity between the nodes of the social network. Along with My-Direct also created an intelligent mechanism for the management of mobile resources in order to reduce energy consumption.

I. Introduction

THE growing interest in social networks and smartphones has resulted in a new trend called MSN (Mobile Social Network). The goal of these social networks is to contribute to human interactions, allowing users who are socially related to use their mobile devices to perform activities of common interest. Some examples of MSN are: Google Latitude[1], Path[2] and Instagram[3].

In recent years, the introduction of a middleware on the development of MSN has attracted the attention of academia, leading the design and development of various approaches by researchers. In [1] it is possible to find some of these solutions, already in [2] describes a solution developed by Brazilian researchers.

A Middleware for MSN (MMSN) should consider the limitations of mobile devices such as limited power, low memory capacity, limited processing power, scalability and heterogeneity [3]. Furthermore, MMSN should provide a layer that provides common services needed by different MSN applications and separate social network management concerns from application requirements [10]. A MMSN should enable efficient operations of MSN application by being self-configuring, self-adapting, self-optimizing and self-protecting.

Being MMSN a topic of recent research, many of the problems mentioned before doesn't have an ideal solution. Thus, current middleware solutions are incomplete, and doesn't exist middleware infrastructure that solves all problems [1].

In order to contribute to this new trend of research, in this paper we propose a middleware architecture for MSN called My-Direct that will make use of Wi-Fi Direct [4] as a solution to improve the connectivity between the nodes of social network. This work also aims to create an intelligent mechanism for the management of mobile device resources to provide a reduction in power consumption.

II. Requirements for a Middleware for MSN

According to [5], an MSN is the social network where one or more individuals with common interests, chat and connect through the use of a cell phone. Also according to [5], there are two possibilities for MSN. The first possibility is that the network be designed from the first moment to be used on mobile devices. The other possibility is a hybrid, that is, the network was first developed for the Web environment and, over time, was migrated to the mobile platform.

Another observation regarding the design of an MSN is related to how it will be accessed by users. Basically, there are two approaches each of which has its positives and negatives. The first would be the distribution of client software, this approach has the benefit of easy access information persisted in the user's device, but it ends up consuming more resources and requires compatible versions with various mobile platforms. The second approach would be the availability of social networking on the Web, where the user would use the browser to access your profile. However, despite access the browser extend the use of different network platforms, contextual information about the user will not be easily extracted and the use of native

[1] Google Latitude. http://www.google.com/mobile/latitude/
[2] Path. http://www.path.com
[3] Instagram. http://www.instagram.com

features of each mobile platform, such as the camera, is restricted.

The MSN must also submit at least two context information related to users: the social context and location context [6]. Through these information, users have the possibility to realize the proximity of friends, members of a group, or profiles that have similarity to user in real time.

The introduction of a middleware facilitates and makes more efficient MSN application development process. Furthermore, the middleware also eases access to data, so that heterogeneous applications have access to them. However, developing a middleware layer for mobile social applications is not an easy task, because it presents a number of issues and challenges that need to be taken into consideration, which, according to [1], are as follows:

- **Simplifying the development process:** A MMSN should be responsible for simplifying the process of developing an application MSN. The simplification of this process can be obtained by providing high-level abstractions with lightweight interfaces to mobile application developers. Facilitate application integration and reuse must also be functions of the middleware.
- **Energy Efficiency:** A MMSN should provide mechanisms to efficiently use the resources of the battery and ensure that the application performs well on mobile devices with limited resources.
- **Privacy:** A MMSN should provide a simple technique that consumes fewer resources to implement appropriate control policies on the exchange of social data, to ensure user privacy.
- **Scalability:** MSN applications suffers increased number of users constantly. This number of nodes should not affect application performance. Thus, a MMSN must be flexible, so that it can manage the increasing number of nodes without compromising system performance.
- **Fully distributed architecture:** A MMSN should be designed to be fully distributed without centralized control. Specifically, it should be built to be used in ad hoc networks without assuming centralized servers.
- **Heterogeneity and dynamic nature of mobile devices:** A MMSN should be designed in a fully distributed fashion, considering the heterogeneous and dynamic nature of mobile devices, as well as privacy concerns. A MMSN should hide the heterogeneity of applications and allow adaptation to dynamic environments.

III. The Middleware Architecture My-Direct

The construction of mobile social networks is a complex task. The introduction of middleware in the field of mobile social networking aims to facilitate the development of this type of network by providing features that help developers manage the users, the maintenance of social relationships between users, in data privacy and social communication between network devices.

With the emergence of new technologies for communication between devices such as Wi-Fi Direct, becomes interesting introduction and evaluation of these new solutions in the middleware environment for MSN. Thus, this work aims to develop the My-Direct, a middleware for MSN that will make use of Wi-Fi Direct, aiming to bring independence of access infrastructure and improvements in the coverage of the social network and in the transmission rate of data between nodes.

The Wi-Fi Direct specification was developed by the Wi-Fi Alliance and operates in 802.11 devices, but is not linked to any specific standard 802.11. This specification introduces the ability to direct connection to millions of devices already have Wi-Fi deployed [4].

According to [4], the introduction of Wi-Fi Direct devices extends the Wi-Fi in order to provide a new connectivity experience. This technology increases the portability of content and applications across all devices of the user through a single and common specification, allowing users to access movies, music and photos point-to-point. Equipment vendors and content providers also benefit from the development of multiple applications that can be seamlessly interconnected wherever you go.

The Wi-Fi Direct is also based on the strengths of Wi-Fi such as performance, security, ease of use and ubiquity, and it adds features such as no need for access to a network infrastructure [4]. Instead of first connect to a network infrastructure and then connect to another device on the network, users can connect directly to devices that offer the services they need. This allows, for example, a user show the photos on your smartphone to your friends by connecting to a television and viewing the images, regardless of the presence of a network infrastructure that is available for both devices.

Although the My-Direct already have their specified architecture, it is still in the implementation phase. In this section we will only describe its layers, its main features and their modules. We will leave the complete encoding and the results for another occasion.

A. Architecture

The middleware My-Direct is being implemented in a P2P architecture targeted the Android platform. The fact that it targeted the Android platform implies the choice of a minimum version of operating system support. For this, we chose to develop the My-Direct to version 4.0 of Android, since the manipulation of resources and support Wi-Fi Direct are present in their SDK (Software Development Kit).

The Fig. 1 shows the My-Direct architecture. Here, we can see that the architecture is composed of four layers: (i) interface, where will stay the set of classes responsible for building the GUI (Graphic User Interface), (ii) communication, where will get located the API that will assist in linking between devices, (iii) privacy, that will

serve to identify users and verify their degree of friendship (iv) modules, which will serve to extend the middleware.

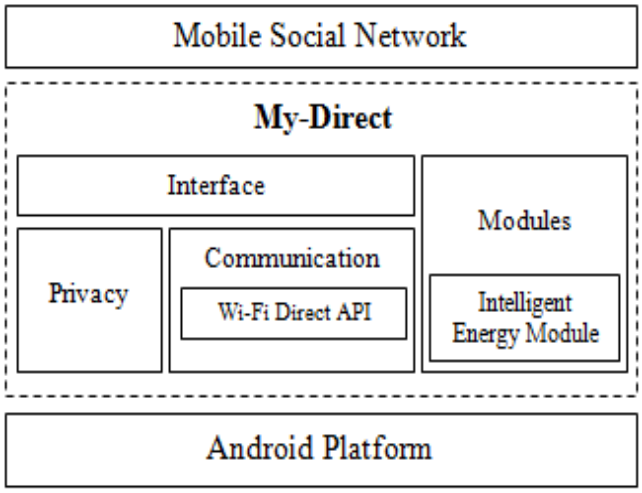


Fig. 1. The My-Direct architecture.

1) Interface Layer

As mentioned earlier, the My-Direct will run on the Android platform. This implies that all classes of the interface will have to follow the pattern of this platform. Thus, the interface layer will contain the implementation of classes that extend the Activity class of the Android SDK. These classes will be responsible for the interaction between the user and My-Direct. Thus, all functions of the My-Direct will be available on the screen and the user only has to select one of them and see the result.

It is important to remember that in the Android platform the design of classes is defined in XML (eXtensible Markup Language). These XML also serves to construct menus, the definition of strings and assigning images to display an Android application. Therefore, in the My-Direct interface layer will be stored implementations of the GUI classes and their respective XML.

2) Communication Layer

The communication layer consists of the classes responsible for the association and effective communication between devices. To perform these actions, classes of this layer will make use of Wi-Fi Direct API, which is provided by the Android SDK. This layer will possess key features like detection of mobile devices with Wi-Fi Direct, the association between them, extracting information (name, IP and MAC) of these devices and effective data exchange.

When occurs the proximity between two devices, the communication layer will check if they are with Wi-Fi Direct enabled. If so, will allow the devices to perform the association and begin the process of information exchange. This whole process will be controlled and monitored by users via interface layer.

3) Privacy Layer

Since the goal of My-Direct is to provide communication between nodes in an MSN regardless of the presence of an access infrastructure in your project we need to take into account the association between mobile devices and their owners have some affinity before starting social activities in the network.

There are several works in the literature such as [7] and [8] that address the issue of how to infer social relationships in a mobile environment. In the case of My-Direct, the independence with respect to the access infrastructure does not allow access to online social networks for information extraction, for example. Thus, it would be wiser to evaluate the affinity between users through information already contained in the mobile device. For this, we chose to identify users using a tuple containing the MAC address of the device, the name that identifies the device during connection and the bond between users. The schema of the tuple can be seen below:

(MAC, DEVICE NAME, BOND)

The MAC address was chosen because it is a unique identifier, which will avoid repetition of information associated with a particular mobile device. As the MAC address is not simple information to the user decorating, we also chose to store and use it along with the device name as information to be displayed on the screen of middleware.

The bond is nothing more than the degree of affinity between two individuals. It works the same way as the concept of circles used in the social network Google Plus[4]. We decided to use this trick because we noticed that the middleware solutions for MSN there were no reference to a management mechanism affinity allowing the user to classify a person according to the degree of affinity. Thus, the user can specify whether a person belongs to the family circle, circle of friends, circle of colleagues or circle of acquaintances. Based on this classification, the user can, for example, share data only with individuals of a given circle.

The information contained in the tuple will be persisted to the mobile device through files. The use of files justifies our choice to keep only three pieces of information, since we know the restrictions of mobile storage. It is also important to note that the act of communication these files will not be transmitted. The information contained therein will be obtained from your own Wi-Fi Direct. The operation of the privacy mechanism My-Direct can be better understood through the following steps:

a. The device D1 is within reach of D2. So at any given time D1 calls D2 to the association.
b. From Wi-Fi Direct, the device D2 gets the MAC and name of D1. With the MAC, D2 checks on your file if there is some input tuples related to this address.
c. How is the first time that D1 finds D2, a new line in the tuple is added. This line is filled with the MAC, the name of D1 and the bond of friendship. For the

[4] Google Plus. http://plus.google.com

bond, the middleware asks through a dialog box to D2 which friendship circle D1 belongs.

d. Steps b and c also happen to D1. Thus, D1 can classify D2 as coworker, D2 can embed D1 in your circle of friends. Thus, it is explicit that the characterization of bond is private, so neither D1 or D2 know how to were classified.

e. Thereafter social activities can be performed.

f. After the end of the communication, if the devices re-find, the information already stored locally will expedite the process of association.

The privacy layer will be responsible for managing the tuples of user data in order to use them as input for mechanism privacy. Thus, with these data the privacy mechanism determines, for example, if a device found in the neighborhood belongs to a friend of the user.

4) Modules Layer

Another goal of My-Direct is to facilitate the development of MSN. For this, the layer modules can be used by developers to introduce new features to the My-Direct. This will allow developers to create different types of MSN.

a) Intelligent Energy Module

In studies such as [16] and [17], it is possible to observe the energy cost of each feature of a mobile device. The application of energy efficiency in mobile devices is very important to their use. Thus, the management of energy consumption can give the user a long period of use of your phone. Thus, the first module developed for the My-Direct is the Intelligent Energy Module (IEM). This module has the responsibility to manage resources of the mobile device as wireless networks and the brightness of the screen, so that the battery usage is optimized.

The IEM was designed to make use of intelligent agents. The multi-agent approach was chosen because agents are autonomous, that is, have the ability to dynamically react to events occurring in the environment in which they were entered. Furthermore, the use of multi-agent in this context allows, for example, every available resource on the mobile device can be controlled by an agent, that allows the implementation of different algorithms for each agent, aiming at optimal solution to the problem.

The agents of IEM are responsible for the execution of pre-defined actions aimed at optimizing the use of the battery. These agents are being implemented through an API called Andromeda (ANDROid eMbeddED Agent platform)[5].

The Andromeda is a platform developed specifically to provide the use of agents with the Android operating system. The agents developed by the Andromeda acts as services in Android. For the specification of behavior and task to be performed by the agent are used Java classes.

The Fig. 2 shows an excerpt of the code of an agent of IEM implemented by Andromeda. It is possible to see that the agent has a behavior (Behaviour class) called "Monitoring". Besides behavior, is also defined capacity (Capability class) named "Turn Off 3G", which will be the trigger for the task associated with it. The trigger fires when the condition contained in the expression method implemented in the Condition class is met. In this case, the action is to turn off the 3G network.

```
public class NetworkAgent extends Agent {

    @Override
    public void init() {

        Behaviour behaviour = new Behaviour("Monitoring");

        Capability mCapac1 = new Capability("Turn Off 3G",
                "intelligentenergysaving.TurnOff3g", new Condition() {

                    @Override
                    public boolean expression(Intent arg0) {
                        if (arg0.getAction().equals(
                                "intelligentenergysaving.TurnOff3g")) {
                            if (wifiIsConnected()) {
                                return true;
                            } else return false;
                        } else return false;
                    }
                });
        mCapac1.addTaskRun(task.class);
        behaviour.add(mCapac1);
```

Fig. 2. Excerpt of the code of an agent of IEM.

The IEM has two agents. One for the management of wireless networks (NetworkAgent) and another for managing the display (DisplayAgent). When the My-Direct is active, the DisplayAgent will periodically check the battery level of the mobile device and the current set the screen brightness according to this information. Meanwhile, the NetworkAgent will turn off or on the wireless networks, with the exception of Wi-Fi Direct, according to pre-established rules, to minimize battery consumption.

B. Example of Use

When the My-Direct get ready, developers will be able to build social P2P applications such as chat and photo sharing, in addition to can extend it through modules. Thus, in a simple chat application developed with My-Direct, for example, the user will be able to use the GUI (interface layer) to see your list of partners, the bond with them, and which theirs is available to connect. If some partner has not yet been classified, the user will have the option to add the bond (privacy layer). When choosing the partner, the communication layer, through Wi-Fi Direct, will go to associate the mobile devices of users. Thereafter, the chat can be performed.

[5] Andromeda. http://www.gti-ia.upv.es/sma/tools/Andromeda/

IV. Related Works

Although middleware for MSN is a new area of study, there are several works in order to resolve issues related to this topic. Each approached the problem from a point of view, however, some of them has similar characteristics.

In [1] it is possible to find some of these solutions. Among them, the most common type of architecture is P2P as in [9], [10], [11], [12], [14] and [15], but there are also those that use centralized architecture as [11] or hybrid as [13]. With regard to network technology used for communication between nodes, the most cited is the Bluetooth. However, there are also works that make use of Wi-Fi (infrastructure mode).

In mobile environment, the user wants to access information anytime and anywhere. Thus, the construction of middleware with a centralized architecture like [11] is not appropriate. The best option for middleware architecture aimed at mobile environment is P2P [1], where there is not centralized control and communication is done directly between devices.

The decentralized architecture requires network technologies that support interaction P2P. Most of the aforementioned studies, with the exception of [11], make use of Bluetooth for communication between nodes of mobile social network. However, the use of Bluetooth requires that the devices are relatively close so that the communication occurs. Furthermore, transmission rate of this technology is limited, which influences the performance of the middleware.

MMSN also deal with sensitive data such as social relationships, activities, and user preferences that can be used to infer other confidential information about the user over time. Thus, as in [12], [13], [15] and [2], a MMSN must manage user data and possess adequate control policies on the exchange of such information, to ensure the user's privacy. However, none of these works allows the user to have the option to sort the people you want to relate according to the degree of affinity.

It is important to note that a MMSN must also be able to minimize battery consumption during its execution. Among the solutions found in the literature only in [11] has this feature.

From the analysis of these works can be seen that the mobile environment requires a decentralized architecture. So to connect the nodes of the network is necessary to use a type of wireless network, independent of central infrastructure, which allows for a fast and efficient communication between nodes of the social network. During this process of communication the battery consumption should also be considered and a solution for energy saving becomes necessary. Furthermore, it is also clear that a user may want to classify one or more persons that relates in a different way, as a friend and colleague for example.

Therefore, to correct these deficiencies, this paper proposes the My-Direct a middleware architecture that will make use of the Wi-Fi Direct and network solution for the relationship between us and a privacy mechanism that will allow the user to sort their peers according to the degree of affinity. My-Direct also allow the use of modules, which may be developed to meet the specific functionality such as managing energy consumption.

V. Conclusion

MSN is a new and rapidly evolving field, and has attracted the attention of both the market and the academy. This makes appear several middleware solutions with different characteristics. Among the main features these solutions are: privacy of users' data, inference unknown social standards, application development and integration with social networking sites.

The middleware solutions in the literature show that there is an approach which satisfies all the requirements that the development of an MSN requires. Furthermore, within this set of approaches each author tries to solve one or more problems, present in a mobile social network environment, using various techniques, methods and tools. Given this middleware solutions heterogeneity, this paper proposed a new architecture called My-Direct. This solution aims to bring new resources to the area middleware for MSN as the introduction of the use of Wi-Fi Direct for P2P communication between mobile devices, a privacy mechanism that will help users to establish social relationships with people of interest and a module for energy saving.

At the end of the design and implementation of the My-Direct, we plan to provide a tool to assist developers in the process of building social networks P2P.

References

[1] A. Karan and N. Mohamed, "Middleware for Mobile Social Networks: A Survey," System Science (HICSS), 2012 45th Hawaii International Conference, pp.1482-1490, Jan. 2012.

[2] M. Holanda, A. Araújo, A. Queiroz, and J. de Castro; "MobiSN: A framework for developing mobile social networks in MANET". Information Systems and Technologies (CISTI), 2012 7th Iberian Conference, pp.1-5, June 2012.

[3] C. Borcea, A. Gupta, and A. Kalra. “The mobisoc middleware for mobile social computing: challenges, design, and early experiences”, 2008.

[4] WI-FI Alliance; “Wi-Fi Certified Wi-Fi Direct, Personal, portable Wi-Fi technology”. 2010. Available: http://www.wi-fi.org/knowledge-center/white-papers/wi-fi-certified-wi-fi-direct%E2%84%A2-personal-portable-wi-fi%C2%AE-technology-2010.

[5] H. Nguyen, “Mobile Social Network and Its Open Research Problems”. Available: http://www2.cs.uh.edu/~rzheng/course/COSC7388sp11/MSN.pdf, 2010.

[6] D. B. A. Santos, “Infraestrutura para o Desenvolvimento de Aplicações Pervasivas Cientes de Redes Sociais”. Dissertação de Mestrado. Universidade Federal de Campina Grande. Centro de Engenharia Elétrica e Informática. Campina Grande. Maio de 2011.

[7] G. Castelli, A. Rosi, and F. Zambonelli, "Design and implementation of a socially-enhanced pervasive middleware," Pervasive Computing and

Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference, pp.137-142, March 2012.

[8] M. Endler, A. Skyrme, D. Schuster, and T. Springer, "Defining situated social context for pervasive social computing" *in Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference*, pp. 519 –524, March 2011.

[9] S. Kern, P. Braun, and W. Rossak, "MobiSoft: An Agent-Based Middleware for Social-Mobile Applications," *in On the Move to Meaningful Internet Systems: OTM 2006 Workshops*, vol. 4277, pp. 984-993, Springer Berlin/Heidelberg, 2006.

[10] D. Bottazzi, R. Montanari, and A. Toninelli, "Context-Aware Middleware for Anytime, Anywhere Social Networks," *in IEEE Intelligent Systems*, Vol. 22, pp.23-32, 2007.

[11] A. Gupta, A. Kalra, D. Boston, and C. Borcea, "MobiSoC: A Middleware for Mobile Social Computing Applications," *Mobile Networks and Applications*, Vol. 14, pp. 35-52, 2009.

[12] D. N. Kalofonos, Z. Antoniou, F. D. Reynolds, M. Van-kleek, J. Strauss, and P. Wisner, "MyNet: A Platform for Secure P2P Personal and Social Networking Services," *in Proceedings of the 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, pp. 135-146, March 2008.

[13] D. N. Kalofonos and Z. Antoniou, "A Hybrid P2P/Infrastructure Platform for Personal and Social Internet Services". *Proceedings of the 19th IEEE Personal, Indoor, and Mobile Radio Communications Symposium - Applications, Services & Business Track (PIMRC '08)*, Cannes, France, pp.1-6, September 2008.

[14] A. Pietilainen, E. Oliver, J. Lebrun, G. Varghese, and C. Diot, "MobiClique: Middleware for Mobile Social Networking," *in Proceedings of the 2nd ACM Workshop on Online Social Networks*, Barcelona, Spain, pp. 49-54, 2009.

[15] M. Basuga, R. Belavic, A. Slipcevic, V. Podobnik, A. Petric, and I. Lovrek, "The MAgNet: Agent-based Middleware Enabling Social Networking for Mobile Users," *in Proc. 10th International Conference on Telecommunications (ConTEL)*, pp. 89-96, 2009.

[16] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," *Proceedings of the 2010 USENIX Conference*, 2010.

[17] G. P. Perrucci, F. H. P. Fitzek, and J. Widmer, "Survey on Energy Consumption Entities on the Smartphone Platform," *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1–6, May 2011.

Position papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 63–66

Usage of RBF Networks in prediction of network traffic

Maciej Szmit
Telekomunikacja Polska
Orange Labs
7 Obrzeżna Street
02-691 Warsaw Poland
Email: maciej.szmit@orange.com

Anna Szmit
Technical University of Lodz
Depatment of Management
266 Piotrkowska Street
90-924 Łódź Poland
Email: agorecka@p.lodz.pl

Marcin Kuzia
Telekomunikacja Polska
Orange Labs
7 Obrzeżna Street
02-691 Warsaw Poland
Email: marcin.kuzia@orange.com

***Abstract*—Prediction of future time series values is area of statistics and computer science research related to pattern recognition. Especially possibility of prediction of the future computer network traffic may be usable in detection of abnormal situations like DoS attacks or occurrence of problems with network infrastructure. The article is devoted to usage artificial neural networks, with radial basis activation function for prediction of network traffic in sample local area networks.**

I. Introduction

IN MANY areas the occurrence of atypical value (named anomaly or outlier) of the particular variable may indicate existence of undesirable phenomenon (especially threat), like symptom of the disease in medicine, change the characteristics of the process in engineering or security incident in telecommunication networks. Anomaly Detection (AD) also called Outlier Detection, is area of statistics and computer science research (see e.g. [1]) related to pattern recognition.

Table I.
Investigated networks description. Source: own research.

Symbol	**Description**
W1	Amateur campus network consisting of circa 25 workstations. Snort has worked on the router which acts also as the gateway to the Internet and as FTP, www, SAMBA and TeamSpeak servers. Data were collected from 13th September to 5th December 2006 with a ten minute interval (a total of 11 969 measurements).
T2	Campus network provided by a mid-size Internet Access Provider – (about 400 clients). Data were collected from 3rd January to 16th March 2007 with ten-minute intervals (a total of 10 001 measurements) on the link between the network and the Internet in housing estates.
T3	A network in a block of flats; one of the subnetworks mentioned in the examples T2 containing about 20 clients. The data were collected from 20th November 2006 to 16th March 2007 with ten-minute intervals (a total of 16402 measurements) on the same link as above (T2) but with address filtering.
MM	Home network connected to the campus amateur network (with maximum speed of inbound traffic set on the bandwidth manager to 4 Mbps. Home network consists of five computers protected by corporate firewall and two intranet servers (ftp and PrintServer). The network has no servers providing outside services and there is no remote access to the home network from the outside. IDS was placed on the link to the campus network before the firewall. The data were collected from 12th February to 1st July 2011 (a total of 20113 measurements).
II	Local Area Network in small company (about 40 computers, two intranet servers). The data were collected from 3rd February 2011 to 4th July 2011 (a total of 21747 measurements).

Anomaly Detection approach needs firstly recognize a pattern of system behavior and next - find observations that differs from the expected ones. The first task may be done using various types of models form classic mathematical and statistical models up to Artificial Intelligence based ones.

Detection of anomalies in computer security domain is one of three approaches used in Intruder Detection Systems (among misuse detection systems and integrity verification – see e.g. [4], [25]). Especially the analysis may concern behavior of single hosts (HBAD – Host Behavior Anomaly Detection) or computer networks (NBAD – Network Behavior Anomaly Detection). NBAD research may focuses on single packets structure anomalies (monitoring phenomena like untypical flag sets in TCP packets, incorrect fragmentation of IP packets etc. – see e.g. [26]) or on network traffic flows (see [21]-[23]).

The previous works of us (see e.g.: [4]-[11], [18]) were according to time series modelling and forecasting using econometric and Artificial Intelligence methods on application of selected time series prediction models in computer networks security area and to methods of detection anomalies of network traffic at the packet level measurements (see e.g.: [12], [13], [19], [24]) using confidence band-based algorithms (see e.g.: [1]-[3]). We focus on classical statistical models like naïve method, autoregression-based (AR, SARIMA), exponential smoothing (Holt-Winters model and its modification) etc. Especially in the article [10] we tried to use multilayer perceptron (MLP) artificial neural networks for modelling and forecasting of network traffic in some local area and campus networks. The current article is devoted to usage of the other kind of neural networks, with radial basis activation function (called RBF Networks) in computer network traffic prediction.

II. The objectives and the method of the research

Like in previous researches we use time series with network traffic data collected from five computer networks described in Table i (detailed information about these networks and descriptive statistics of collected time series are described in [9]).

We collect several time series, containing data about network traffic (overall number of packets received by network probe in five minutes period) according to the three most popular network protocols: TCP, UDP and ICMP and next modeled these time series using RBF networks.

A radial basis functions (see e.g.: [14]-[16]) are a real-valued functions whose value depends only on the distance from the some other point c called a center

$$f(x,c) = f(\|x-c\|) \tag{1}$$

The norm is often euclidean distance, however other distance functions are also possible.

Radial basis function networks are artificial neural networks uses radial basis functions as activation functions. Typically RBF networks typically contains from three layers: input layer which is designed only to sends input signals for all neurons in hidden layer, a hidden layer with radial activation function and a linear output layer so the output of the network are a linear combination of radial basis functions of the inputs and neuron parameters (see Fig. 1).

The value on j output neuron network may be described as a function of the input vector x given by the equation

$$y_j = \sum_{i=1}^{N} w_i f\left(\|x-c_i\|\right) \tag{2}$$

where:

N is the number of neurons in the hidden layer,

x is the input vector,

c_i is the center vector for neuron i and

w_i is the weight of neuron i in the linear output neuron.

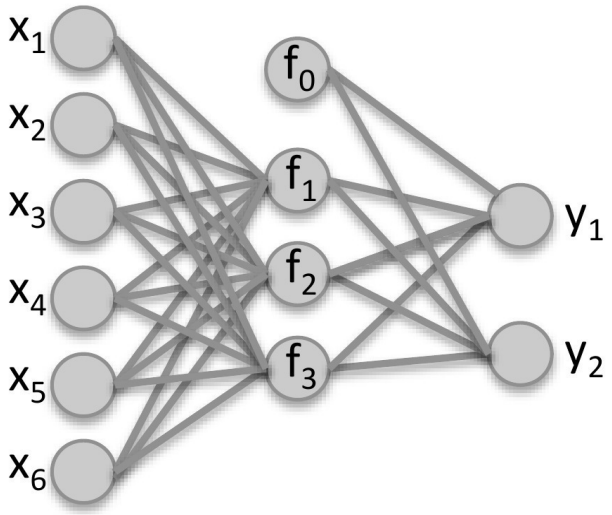


Fig. 1. Sample RBF Network. Source: own research.

We used RBF ANN simulator build in Statistica packet with default algorithm of network structure determination.

Like in previous research we decided to use MAE (Mean Absolute Error), rather than MSE (Mean Squared Error) based measure because there were a lot of so-called outliers are noted in the analysed samples and MSE-based measure can be oversensitive in those cases (see e.g.:[17]). As a meter of model fit we use a quotient:

$$\frac{Mean\_Absolute\_Error}{Mean}$$

expressed as a percentage.

III. Results and conclussions

For each time series we tried to predict single value (number of particular packets in the next 5 minutes), so all of the networks has one neuron in output layer. On the input ANNs can get information about previous value of time series. In all of cases the explanatory variable were value of the series delayed by 1, 2 and 3 periods, and in the most of cases additionally the explanatory variables were value of the series delayed by 144 and 145 periods (that mean delayed by one

TABLE II.
RBF AND MLP ANNS STRUCTURES AND FIT. SOURCE: [10], OWN RESEARCH.

Series	Protocol	MLP Topology	MLP MAE/M	RBF Topology	RBF MAE/M
W1	TCP	2-1-1 (-1,-3)	**46,18%**	4-7-1	47,33%
W1	UDP	1-2-1 (-1)	**31,73%**	3-7-1	33,28%
W1	ICMP	4-2-1 (-1, -2, -3, -144)	**34,54%**	6-7-1	36,63%
T2	TCP	1-1-1 (-1)	**4,23%**	3-10-1	4,28%
T2	UDP	5-1-1 (-1, -2, -3, -144, -1009)	**15,41%**	5-2-1	17,07%
T2	ICMP	2-2-1 (-1, -2)	**8,66%**	3-12-1	9,65%
T3	TCP	1-1-1 (-1)	**4,07%**	3-7-1	4,31%
T3	UDP	1-1-1 (-1)	**15,05%**	5-2-1	22,13%
T3	ICMP	3-1-1 (-1, -3, -1008)	**8,91%**	5-2-1	21,86%
MM	TCP	2-2-1 (-1, -2)	**75,72%**	6-10-1	82,64%
MM	UDP	4-1-1	30,12%	3-7-1 (-1,-2,-3)	**29,11%**
MM	ICMP	3-1-1	10,93%	3-13-1 (-1,-2,-3)	**10,77%**
II	TCP	2-1-1	41,14%	4-10-1 (-1, -2, -3, -1008)	**38,97%**
II	UDP	5-1-1 (-1, -2, -3, -144, 1008)	**48,42%**	5-8-1	49,24%
II	ICMP	1-1-1	116,44%	3-7-1 (-1,-2,-144)	**114,43%**

day and one and five minutes) and in several cases the algorithm select also values delayed by 1008 and 1009 that means week and week and five minutes).

The Table ii includes number of neurons in input, hidden and layer in each RBF ANN. presents fits of the model (MAE/M) compared with results gets by Multilayer Perceptron (MLP) ANN described in article [10]. For winner ANNs its topology and explanatory variables are described.

As one can see RBF ANNs structure determination algorithm tend to choose a larger number of explanatory variables comparing the analogous one for MLP ANNs, but it does not often lead to better model fit. Only in TCP in II network traffic time series the results of RBF were significantly better then MLP.

REFERENCES

[1] Markou M., Singh S.: Novelty detection: a review part 1: statistical approaches, Signal Processing, vol. 83, pp. 2481 – 2497, Decemeber 2003;

[2] Brutlag J. D.: Aberrant behavior detection in time series for network monitoring, Proceedings of the 14th System Administration Conference, pp. 139–146, New Orleans, Fla, USA, 2000.

[3] Liu, W; Lin S., Piegorsch W. W.: Construction of Exact Simultaneous Confidence Bands for a Simple Linear Regression Model, International Statistical Review 76 (1): 39–57. doi:10.1111/j.1751-5823.2007.00027.x.

[4] Szmit M., Adamus S., Bugała S., Szmit A.: Implementation of Brutlag's algorithm in Anomaly Detection 3.0, Federated Conference on Computer Science and Information Systems, Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 685–691, PTI, IEEE, Wrocław 2011

[5] Szmit M., Szmit A.: Use of Holt-Winters method in the analysis of network traffic. Case study, Springer Communications in Computer and Information Science vol. 160, 18th Conference Computer Networks, 2011, s. 224-231, ISSN: 1865-0929; ISBN: 978-3-642-21770-8, DOI: 10.1007/978-3-642-21771-5_24

[6] Szmit M., Szmit A.: Usage of Modified Holt-Winters Method in the Anomaly Detection of Network Traffic: Case Studies, Journal of Computer Networks and Communications, vol. 2012, DOI:10.1155/2012

[7] Szmit M., Szmit A.: Usage of Pseudo-estimator LAD and SARIMA Models for Network Traffic Prediction. Case Studies, Communications in Computer and Information Science, 2012, Volume 291, 229-236, DOI: 10.1007/978-3-642-31217-5_25

[8] Szmit A., Szmit M.: O wykorzystaniu modeli ekonometrycznych do prognozowania ruchu sieciowego, Zarządzanie rozwojem organizacji, Spała 2013 (accepted for publication)

[9] Szmit M.: Využití nula-jedničkových modelů pro behaviorální analýzu síťového provozu, Internet, competitiveness and organizational security, TBU, Zlín 2011

[10] Szmit M., Szmit A., Adamus S., Bugała S.: Usage of Holt-Winters Model and Multilayer Perceptron in Network Traffic Modelling and Anomaly Detection, Informatica Vol. 36, Nr 4, pp. 359-368

[11] Jašek R., Szmit A., Szmit M.: Usage of Modern Exponential-Smoothing Models in Network Traffic Modelling, Advances in Intelligent Systems and Computing Volume 210, 2013, pp. 435-444 (Nostradamus 2013: Prediction, Modeling and Analysis of Complex Systems), DOI:978-3-319-00542-3_43

[12] Münz G.: Traffic Anomaly Detection and Cause Identification Using Flow-Level Meas- urements, TUM, Müchen 2010, http://www.net.in.tum.de/fileadmin/TUM/NET/NET- 2010-06-1.pdf

[13] Wang Y.: Statistical Techniques for Network Security: Modern Statistically-Based Intrusion Detection and Protection, IGI Global 2009

[14] Broomhead D. S. LoweD.: Radial basis functions, multi-variable functional interpolation and adaptive networks, Technical report 4148, RSRE 1988, http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA196234

[15] Yanwei F., Yingying Z., Haiyang Y.: Study of neural network technologies in intrusion detection systems, Proceedings of the 5th International Conference on Wireless communications, networking and mobile computing (WiCOM'09). IEEE Press, NJ 2009 pp. 4454-4457

[16] Yang Z.: An intrusion detection system based on RBF neural network, Proceedings of the Ninth International Conference on Computer Supported Cooperative Work in Design, 2005, pp. 873 - 875 Vol. 2

[17] Rousseeuw P. J., Leroy A. M.: Robust Regression and Outlier Detection, Wiley, 1987

[18] Szmit A., Szmit M.: O wykorzystaniu modeli ekonometrycznych do prognozowania modeli ruchu sieciowego, Zeszyty Naukowe Organizacja i Zarządzanie Nr 1154 (55), Politechnika Łódzka, Łódź 2013, pp. 193-201, ISSN: 0137-2599

[19] ITU-T E.507 Models for Forecasting International Traffic, ITU 1998, http://www.itu.int/rec/T-REC-E.507-198811-I/en

[20] Villén-Altamirano M.: Overview of ITU Recommendations on Traffic Engineering, Paper presented in the ITU/ITC workshop within 17th ITC.

[21] Rajahalme J., Conta A., Carpenter B., Deering S.: IPv6 Flow Label Specification, RFC 3697, Network Working Group 2004

[22] Quittek J., Zseby T., Claise B., Zander S.: Requirements for IP Flow Information Export (IPFIX), RFC 3917, Network Working Group 2004

[23] Brownlee N., Mills C., Ruth G.: Traffic Flow Measurement: Architecture, RFC 2722, Network Working Group 1999

[24] ITU-T Recommendation E.490.1: Overview of Recommendations on traffic engineering, ITU 2003, http://www.itu.int/rec/T-REC-E.490.1-200301-I/en

[25] Vala R., Malanik D., Jašek, R. "Usability of Software Intrusion-Detection System in Web Applications", Advances in Intelligent Systems and Computing, Vol. 189, pp. 159-166, Springer-Verlag, Berlin 2013

[26] V. A. Siris and F. Papaglou, "Application of anomaly detection algorithms for detecting syn floodinfg attacks," in Proceedings of the IEEE Global Telecommunications Conference, vol. 4, pp. 2050–2054, 2004

International Conference on Wireless Sensor Networks

A FEW years ago, the applications of WSN were rather interesting example than a powerful technology. Nowadays, this technology attracts still more and more scientific audience. Theoretical works, where the WSN principles were investigated, grew into the interesting practical applications integrated by this time in a real life. Various application fields, from military to healthcare, are already covered by WSN. Hand in hand with WSN application coverage expansion, still new and new tasks bringing along the new interesting problems occur. Therefore such application practices stimulate the progress of WSN theory as well as unlock new application possibilities.

Wireless sensor networks, as the spatially distributed networks consisted of relatively simple components interconnected mutually, provide quite wide application potential to be utilized in military, industry, transport, agriculture, healthcare and many other branches. However, in the near future, even higher growth of WSN application coverage could be expected. This expansion is nevertheless conditioned by solving questions related to the communication protocols standardization, to the lack of effective energy sources enabling network nodes working time prolongation, as well as to the progress in the field of ultra-low-power microelectronic components industry. An integration of WSN within the public data networks as well as within the domains where confidential and private data are proceed (e.g. E-Health) brings about problems related to the ethical and legal questions too.

The WSN is one of actual affairs getting to the fore in the European Research Area since the issue of Future network technologies research and innovation is planned to be included in the new HORIZON2020 FP7 program.

It is therefore necessary to create an experience-sharing platform for scientific researchers and experts from research institutes and WSN occupied companies that employ benefits of this modern technology and to the exchange of skills, experiences and new ideas from the WSN problematic within the all participants of iNetSApp.

Topics

Original contributions, not currently under review to another journal or conference, are solicited in relevant areas including, but not limited to, the following:

Development of sensor nodes and networks

- Sensor Circuits and Sensor devices – HW
- Applications and Programming of Sensor Network – SW
- Architectures, Protocols and Algorithms of Sensor Network
- Modeling and Simulation of WSN behavior
- Operating systems

Problems dealt in the process of WSN development

- Distributed data processing
- Communication/Standardization of communication protocols
- Time synchronization of sensor network components
- Distribution and auto-localization of sensor network components
- WSN life-time/energy requirements/energy harvesting
- Reliability, Services, QoS and Fault Tolerance in Sensor Networks
- Security and Monitoring of Sensor Networks
- Legal and ethical aspects related to the integration of sensor networks

Applications of WSN

- Military
- Health-care
- Environment monitoring
- Transportation & Infrastructure
- Precision agriculture
- Industry application
- Security systems and Surveillance
- Home automation
- Entertainment – integration of WSN into the social networks
- Other interesting applications

Event Chairs

Fouchal, Hacene, University of Reims Champagne-Ardenne, France

Hodon, Michal, University of Žilina, Slovakia

Kapitulík, Ján, University of Žilina, Slovakia

Miček, Juraj, University of Žilina, Slovakia

Segal, Michael, Ben-Gurion University of the Negev, Israel

Ševcik, Peter, University of Žilina, Slovakia

Stojmenovic, Ivan, University of Ottawa, Canada

Xiao, Yang, The University of Alabama, United States

Program Committee

Gu, Yu, National Institute of Informatics, Japan

Scholz, Bernhard, The University of Sydney, Australia

Shu, Lei, Osaka University, Japan

Staub, Thomas, Data Fusion Research Center (DFRC) AG, Switzerland

Wang, Zhonglei, Karlsruhe Intitute of Technology, Germany

Al-Anbuky, Adnan, Auckland University of Technology, New Zealand

Baranov, Alexander, Russian State University of Aviation Technology, Russia

Chaczko, Zenon, University of Technology Sydney, Australia

D'Innocenzo, Alessandro, University of L'Aquila, Italy

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 69–74

Resources placement in the 4-dimensional fault-tolerant hypercube processors network

Jan Chudzikiewicz
Military University of Technology
ul. Kaliskiego 2,
00-908 Warszawa, Poland
Email: jchudzikiewicz@wat.edu.pl

Zbigniew Zieliński
Military University of Technology
ul. Kaliskiego 2,
00-908 Warszawa, Poland
Email: zzielinski@wat.edu.pl

***Abstract*—In the paper some properties of the perfect resources placement in the 4-dimensional hypercube network with soft degradation are considered. The methods of determining some kind of perfect placement are discussed. Two algorithms for the resources deployment are proposed. Moreover, some characteristics of the network degradation are determined. The average number of processors working of degraded network with given order for a specific type of resource placement is determined. This value characterizes the loss of the network computing capabilities resulting from the increase of the degree of network degradation. Also, the probability distribution that for degraded logical structures of the hypercube with given order there exists a specific type of resources placement is evaluated.**

I. Introduction

A REGULAR structure of processors network as torus or hypercube ([1],[2]) could be implemented in many kind of networks. An interconnection network with the hypercube logical structure is a well-known interconnection model for multiprocessor systems. Such networks possess already numerous applications in critical systems ([3], [4]) and still they are the field of interest of many theoretical studies concerning system level diagnostics ([5], [6], [7], [8], [9]) or resource placement problem, which has intensively been studied in [10]-[16].

We could observe the usage processors networks in critical application areas (military, aerospace or medical systems etc.). Such networks are (mostly) used in real-time systems, which require a very high data reliability processing throughout the life of the network, which is achieved with the use of real-time system diagnostic techniques. It causes increase of specific requirements concerning the dependability of such systems. The increase of the dependability of a processors network could be achieved by using effective methods of diagnosis and reconfiguration after the faulty processor identification. Designing and exploitation of such networks is a comprehensive project that requires addressing a number of theoretical and practical problems. One of the problem is skillful resources deployment in the network and modification of resource deployment after each phase of the network degradation. Wherein by the resource is to be understood for example: database, I/O port, data files etc.

We investigate the case when network of processors has the logical structure of 4-dimensional hypercube. A topology of hypercube may be represented by an ordinary consistent graph whose nodes are described by 4-dimensional binary vectors such that the Hamming distance between vectors (labels) of the adjacent nodes equals one.

In networks with soft degradation a processor identified as faulty is not repaired (or replaced) but access to it is blocked. New (degraded) network continues work under the condition that it meets specific requirements [17]-[19]. The method of a hypercube network reconfiguration after identifying faulty processors, based on the determination of the set of fit coherent sub-cubes of hypercube with maximum dimensions was presented by Chen and Tzeng in [11]. In the work [20] an analysis of the different patterns of reconfiguration in networks with soft degradation was conducted, these schemes were divided into software and hardware.

We assume that processors of a network could be divided into processors working, and resources processors (e.g. distributed database, I/O devices). The execution of some tasks by a processor working requires an access to resources, also some results obtained by processors working must be submitted to the resources [21].

A generalized cost of a network traffic with a specified resources deployment and workload of network are usually tested with experimental methods or simulation. We try to express it analytically for a given perfect deployment, which is defined as the minimum number of resources processors which are accessible by processors working. Each processor working has access to the at least m resources processors at a distance of not more than d. Such an approach is known in relation to the network of regular logical structure [12], [13], [14], [15], [16]. The definition *(m,d)*-perfect deployment is a characteristic of a value of the generalized cost of information traffic in the network at a given load of tasks.

The goal of this paper is to apply the above presented approach to a processors network of a 4-dimensional cube type logical structure along with its soft degradation process.

The paper consists of four sections and a summary. The second section provides a basic definitions and properties of working structures which are induced by the network in the process of its degradation. There are defined a concept of perfect deployment of processors with resources in the working structure of network G (definition 3). In the third section are considered the conditions of existence of a specific type of perfect deployments and the algorithms for determining them for working structures. In the fourth section the characteristics of the network degradation as the probability distribution of the number of consecutive hypercube processors failures, after which it loses possibility of perfect resources deployment of type $(m,1)$.

In summary there are formulated conclusions from the results presented in the paper.

II. THE BASIC DEFINITIONS AND PROPERTIES

Definition 1. The logical structure of processors network we call the structure of 4-dimensional hypercube if it is described by coherent ordinary graph $G = < E, U >$, (E-set of processors, U-set of bi-directional data transmission lines), which nodes can be described (without repetitions) by 4-dimensional binary vectors (labels) in the following way

$$[H(\varepsilon(e'), \varepsilon(e'')) = 1] \Leftrightarrow [(e', e'') \in U], \qquad (1)$$

where $H(\varepsilon(e'), \varepsilon(e''))$ is Hamming distance between the labels of nodes e' and e''.

If $|E| = 2^4$ and $|U| = 2|E|$ then graph of type of 4-dimensional cube we denote by H^4 and we call it (non labeled) 4-dimensional the cube.

Let $\tilde{H}^4 = \langle H^4; \{\varepsilon(e) : e \in E(H^4)\}\rangle$ indicates 4-dimensional cube with labeled nodes, and $\breve{G}_p^C(H^4)$ and $\breve{G}_p^C(\tilde{H}^4)$ denote coherent sets of subgraphs of graphs H^4 and $\tilde{H}^4$ of order p.

Let $\nu(G)$ $(G \in \breve{G}_p(H^4), p \ge 6)$ indicates the number of ways to assign labels for nodes of graph G satisfying the formula 1 and thus the number of subgraphs of graph $\tilde{H}^4$ which after removal of the labels of the nodes are isomorphic with graph G.

The graph G is a regular graph of degree 4 with the degree $\mu(e) = |E(e)|$ ($E(e)$ - a set of nodes adjacent to a node $e \in E$) -of each node of the graph G is equal to 4.

Geometric forms of working structures $G \in \breve{G}_p(H^4)$ of order $p \ge 6$ induced by the network in the process of its degradation are presented in the form of a possibly minimum number of intersecting edges ([9], [22], [23]). It should be noticed that this form of presentation makes it easier to analyze the structures properties. The numbers $\nu(G)$ are determined by the method of structures composition [22][23]. Some examples of such structures are presented in Fig. 1.

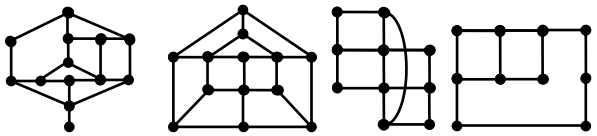


Fig. 1 Examples of geometric forms of working structures of the network (specified the numbers $\nu(G)$) [9]

From many properties of working structures of the network we will discuss only properties of a cycle.

The cycle will be treated as a subgraph of graph G and not as a cyclical chain [24], [25].

Definition 2. A cycle C in the graph G we call such a coherent subgraph of G that $\forall_{e \in E(C)} : \mu(e|C) = 2$.

Property 1. $\forall_{C \in C(G)} : |E(C)| \in \{4, 6, 8\}$ ($C(G)$ **-**the set of cycles in the graph G), because the graph C is a graph of class H^4 and thus (in accordance with definition 1)

$$\forall_{e \in E(C)} : [\{e', e''\} = E(e)] \Rightarrow$$
$$\Rightarrow [(H(\varepsilon(e), \varepsilon(e')) = H(\varepsilon(e), \varepsilon(e'')) = 1) \wedge$$
$$\wedge (\forall_{e''' \in \{E(C) \setminus \{E(e) \cup \{e\}\}\}} : H(\varepsilon(e), \varepsilon(e''')) > 1)]$$

($H(\varepsilon(e), \varepsilon(e'))$ **-**Hamming distance between the labels of nodes e and e'), what is satisfied only if the string lengths of the cycle binary labels are equal to 4, 6 or 8 (there is no need to justify that).

Property 2.

$$[\exists_{\{C', C''\} \subseteq C(G)} : \{E(C') \cap E(C'')\} = \varnothing] \Rightarrow$$
$$\Rightarrow [\neg \exists_{u \in U(G)} : |\{E(u) \cap E(C')\}| =$$
$$= |\{E(u) \cap E(C'')\}| = 1].$$

The deployment of data resources in the processors network of 4-dimensional cube type logical structure and soft degradation we will regard as a labeled graph $\langle G; \dot{E}\rangle$ $(G \in \breve{G}_p(H^4),\ p \ge 6, \dot{E} \subset E(G),\ \dot{E}$ - set of database processors, $\{E(G) \setminus \dot{E}\}$ - set of the working processors of the network G) [26].

Let $d(e', e''|G)$ denotes the distance between nodes e' and e'' in a coherent graph G, that is the length of the shortest chain (in the graph G) connecting node e' with the node e''.

Of course $d(e', e''|\tilde{H}^4) = H((\varepsilon(e'), \varepsilon(e''))$.

Definition 3. We say ([12], [13], [14], [15], [16]) that the labeled graph $\langle G; \dot{E}\rangle$ $(|\dot{E}| \ge 1)$ is $(m,d|G)$ **-perfect placement** $(m \in \{1,\ldots,\mu(G)\},\ d \in \{1,\ldots,D(G)\},$ $D(G)$ - diameter of the graph G) a set $\dot{E}$ of resources processors in the network G if there exists set $\dot{E}$ of minimum cardinality satisfying the following condition

$$[\forall_{e\in\{E(G)\setminus\dot{E}\}}:|\{e'\in\dot{E}:d(e,e'|G)\le d\}|\ge m]\wedge$$
$$\wedge[\forall_{\{e^*,e^{**}\}\subset\dot{E}}:d(e^*,e^{**}|G)>d]\wedge \quad (2)$$
$$\wedge[(\mu(e''|G)=1)\Rightarrow(e''\notin\dot{E})].$$

$(m, d|G)$- **perfect placement** will be denoted by $(m, d|G)d$ to distinguish from the placement of type $(m, d|G)$ in which the set $\dot{E}$ does not need to be of minimum cardinality.

Note that (according to the definition 3) a set $\dot{E}$ in deployment $(m, d|G)$ is a particular kind of an externally stable set.

Let $F_{(m,d)d}(G)$ denote the set of $(m, d|G)$-perfect deployment in network G.

We are interested in the conditions of existence perfect deployments of type $(m, 1|G)$, the algorithms for their determination and the average number $\Re_{(m,d|G)d}(p)$ of working processors of the network of order p with $(m, d|G)$-perfect deployment of resources. The average number $\Re_{(m,d|G)d}(p)$ characterize the loss of the network computing potential (the number of working processors) together with increasing its degradation with the use of the specified $(m, d|G)$-perfect deployment.

III. PERFECT PLACEMENTS IN THE NETWORK WORKING STRUCTURES

Let $F_{(m,d)}(G)$ denotes the set of $(m, d|G)$-perfect deployments in the network G and $\dot{E}(f)$-set of resources processors of the network G for the placement $f\in F_{(m,d)}(G)$.

Let us consider the methods and the algorithms for determining perfect deployments $(1, 1|G)$ for $G\in\breve{G}_p(H^4)$ $(p\ge 6)$ and we determine (for these deployments) the average number

$$\Re_{(m,d|G)d}(p)=\left|\breve{G}_p(H^4)\right|^{-1}\sum_{G\in\breve{G}_p(H^4)}\nu(G)(p-\left|\dot{E}_{(m,d)d}(G)\right|)$$
$$(p\ge 6)$$

of working processors of the network of order p with $(m, d|G)$-perfect deployment of data resources (with distinction of cyclic and acyclic structures).

We denote $E^{(d)}(e|G)=\{e'\in E(G):d(e,e'|G)\le d\}$ $(d\in\{1,\ldots,D(G)\})$ and $\hat{E}^1(G)=\{e\in E(G):(\exists_{e'\in E(e)}:\mu(e')=1)\}$.

Let us consider $(1, 1|G)$-perfect deployment.

Note that $[f\in F_{(1,1)}(C)]\Rightarrow[\left|\dot{E}(f)\right|=\left\lceil 3^{-1}|E(C)|\right\rceil]$ $(C\in C(G))$, and cyclic sequences $\lambda_{(1,1)}(|E(C)|)$ of numbers $d(e',e''|C)$ $(\{e',e''\}\subseteq\dot{E}(f))$ have the form of $\lambda_{(1,1)}(4)=(2,2)$, $\lambda_{(1,1)}(6)=(3,3)$ and the $\lambda_{(1,1)}(8)=(3,3,2)$.

Property 3. $[G\in\breve{G}_p^C(H^4)(p\ge 6)]\Rightarrow[F_{(1,1)d}(G)\ne\varnothing]$ because $[G\in\breve{G}_p^C(H^4)(p\ge 6)]\Rightarrow[F_{(1,1)}(G)\ne\varnothing]$ because such a placement $\langle\tilde{H}^4;\dot{E}\rangle$ that $\dot{E}=\{e\in E(\tilde{H}^4):\delta(\varepsilon(e))\in\{0,2,4\}\}$ ($\delta(\varepsilon(e))$- is the number of ones in the binary label $\varepsilon(e)$ of a node $e\in E(\tilde{H}^4)$) is placement of type $(1,1|\tilde{H}^4)$, and thus the placement $\langle G;E(G)\cap\dot{E}\rangle$ is an element of the set $F_{(1,1)}(G)$ although it does not need to be an element of the set $F_{(1,1)d}(G)$ (see Fig. 2).

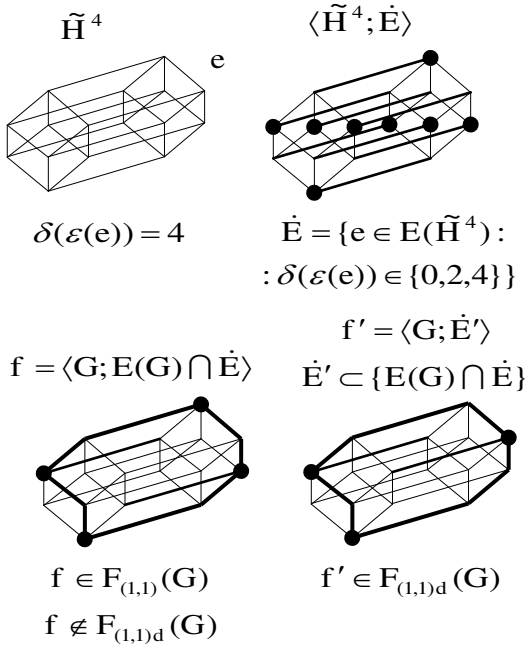


Fig. 2 Illustration of the property 3

Property 4. $[F_{(1,1)}(G)=\varnothing]\Leftrightarrow[\exists_{\{e',e''\}\subset E(G);\mu(e')=\mu(e'')=1}:$ $:(E(e'),E(e''))\in U(G)]$ because (according to the formula 2) nodes $E(e')$ and $E(e'')$ must belong to the set $\dot{E}$, which contradicts that it is an externally stable set.

The methods of determining $(1, 1|G)$-perfect deployment:

1. If $G\in\breve{G}^C(H^4)$ then as the first node of a set to be found $\dot{E}(f)$ $(f\in F_{(1,1)d}(G))$ we choose such a node $e_i\in E(G)$ with the greatest degree that the subgraph $\overline{G}^{(1)}(G,e_i)=\langle\{E(G)\setminus E^{(1)}(e_i)\}\rangle_G$ has the smallest number of components of coherence and we determine a placement for every of these components of coherence,

wherein if a component of coherence is one-node it belongs to the set $\dot{E}(f)$.

2. If $G \in \breve{G}^{C}(H^{4})$ then we determine the subgraph of G which is a cycle (in sense of definition 2) C of the greatest order and so we choose nodes of set $\dot{E}(C)$ so that they form a cyclic sequence $\lambda_{(1,1)}(|E(C)|)$ and the expression $\sum_{e \in \dot{E}(C)} \mu(e|G)$ reached the maximum value. If $\exists_{e' \in \{E(G) \setminus E(C)\}} : \{E(e'|G) \cap \dot{E}(C)\} = \varnothing$ then $e' \in \dot{E}(G)$.

A. The algorithm I - determining the resource placement based on the first method

Step 1.

Choose a node $e_i \in E(G)$ such that:

- the degree of $\mu(e_i) = \max_{e \in E(G)} \mu(e)$;
- subgraph $\overline{G}^{(1)}(G, e_i)$ has the smallest number of components of coherence.

Add the node e_i to the set $\dot{E}(f)$.

Step 2.

Check if a component of coherence of the subgraph $\overline{G}^{(1)}(G, e_i)$ is one-node or $\overline{G}^{(1)}(G, e_i) = \varnothing$.

YES

If $\overline{G}^{(1)}(G, e_i) \neq \varnothing$ add all nodes of $\overline{G}^{(1)}(G, e_i)$ to the set $\dot{E}(f)$.

Go to step 3.

NO

Assume that the $\overline{G}^{(1)}(G, e_i)$ is a new graph G.

Return to step 1.

Step 3.

The end of the algorithm.

An illustration of the algorithm work is presented in Fig. 3. One of the possible $(1,1|G)$-perfect deployment for the structure G (chosen by the algorithm) is shown in Fig. 3a.

B. The algorithm II - determining the resource placement based on the second method

Step 1.

Determine the subgraph of G which is a cycle C such that: $|E(C)| = \max_{C' \subset G} |E(C')|$. If a cycle C not exist in G go to step 3.

Step 2.

Choose nodes of set $\dot{E}(C)$ such that:

- the nodes form a cyclic sequence $\lambda_{(1,1)}(|E(C)|)$ ($\lambda_{(1,1)}(4) = (2,2)$, $\lambda_{(1,1)}(6) = (3,3)$, $\lambda_{(1,1)}(8) = (3,3,2)$).
- the expression $\max(\sum_{e \in \dot{E}(C)} \mu(e|G))$.

Step 3.

The end of the algorithm.

The algorithm working correctly for the structure G which is a cyclic graph. An illustration of the algorithm work is presented in Fig. 4. One of the possible $(1,1|G)$-perfect deployment for the structure G (chosen by the algorithm) is shown in Fig. 4a.

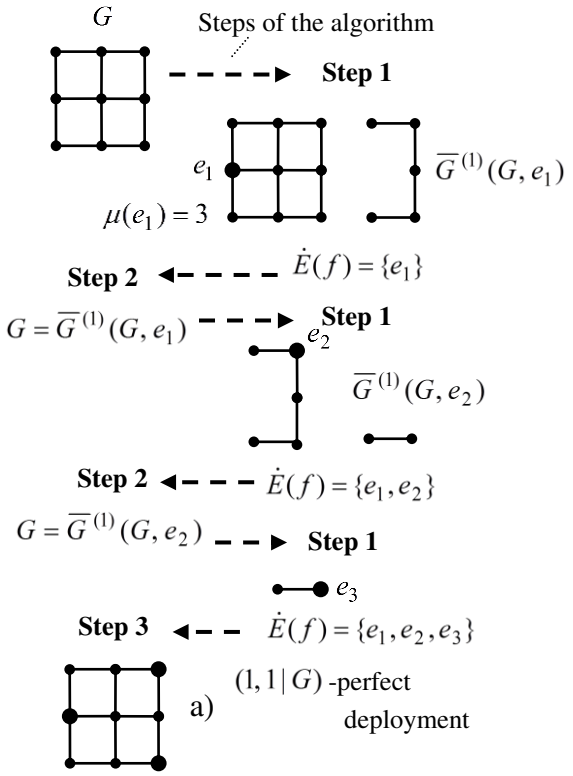


Fig. 3 An illustration of the algorithm I steps

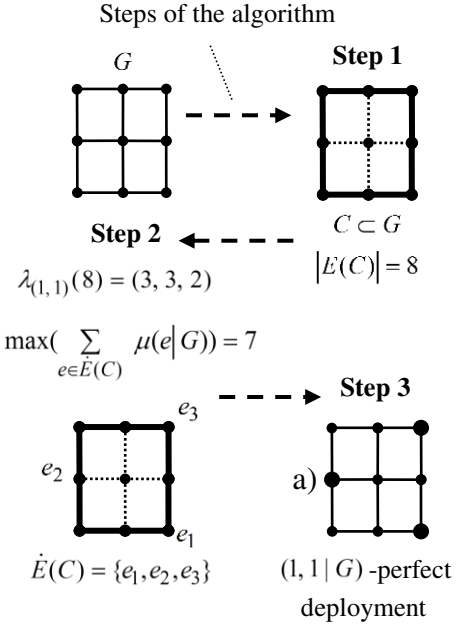


Fig. 4 An illustration of the algorithm II steps

IV. CHARACTERISTICS OF THE NETWORK DEGRADATION

A degraded structure G' of the structure $G \in \breve{G}_p^C(\tilde{H}^4)$, $(p > 6)$, after a processor $e \in E(G)$ failure is the most numerous (in terms of number of nodes) such a cyclic consistent component $G' = \langle E(G) \setminus \{e\} \rangle_G$ of the graph G that there exists a placement $(m, 1 | G')$ for $m \geq 1$. If the graph G' does not exist, then processor e is called a critical processor of the structure G [18]. The critical processor failure causes the unfitness of the network for further usage. A structure, in which a critical processor exists, is called a critical structure.

It could be easily observed (on the base of properties 3-4) that verifying the condition of the existence of placement $(m, 1 | G')$ for degraded structure G' could be reduced to checking the following condition: $(G' \in \breve{G}^C(H^4)) \wedge (|E(G')| \geq 6)$.

Let us determine

$$p_j^i(G) = \left|\{e \in E(G) : \langle E(G) \setminus \{e\} \rangle_G \in \breve{G}_j^C (j \geq 6)\}\right| \; |E(G)|^{-1}$$

$$G \in \breve{G}_i^C(H^4), i > j$$

and

$$p_\bullet^i(G) = \left|\{e \in E(G) : \langle E(G) \setminus \{e\} \rangle_G \notin \breve{G}_j^C (j \geq 6)\}\right| \; |E(G)|^{-1}.$$

Degradation characteristics $\pi^C(G)$ of structure $G \in \breve{G}_i^C(H^4)$ is called a pair $\langle p^i(G); p_\bullet^i(G) \rangle$, where $p^i(G) = (p_{i-1}^i(G), p_{i-2}^i(G), ..., p_6^i(G))$, $7 \leq i \leq 16$, whereas $p_\bullet^6(G) = 1$.

For example, structure G' in Fig. 5 Examples of specifying degradation characteristics of structure is not a critical structure and has a degraded characteristic $\pi^C(G') = \langle (2/10, 4/10, 0, 4/10); 0 \rangle$, and structure G'' is a critical structure with four (indicated) critical processors as well as $\pi^C(G'') = \langle (0, 4/8); 4/8 \rangle$.

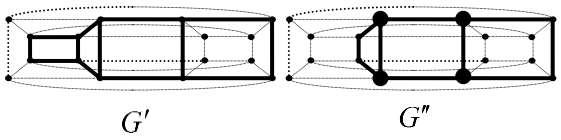


Fig. 5 Examples of specifying degradation characteristics of structure ($\pi^C(G') = \langle (2/10, 4/10, 0, 4/10); 0 \rangle$ and $\pi^C(G'') = \langle (0, 4/8); 4/8 \rangle$) (critical nodes of the structures were indicated)

The set $\{\langle p^i(\tilde{H}^4); p_\bullet^i(\tilde{H}^4) \rangle : i \in \{7, ..., 16\}\}$, $(p^i(\tilde{H}^4) = (p_{i-1}^i(\tilde{H}^4), ..., p_6^i(\tilde{H}^4)))$, is called a degradation characteristic of a processor network, whereas

$$p_j^i(\tilde{H}^4) = \left|\breve{G}_i^C(\tilde{H}^4)\right|^{-1} \sum_{G \in \breve{G}_i^C(H^4)} p_j^i(G)\, \nu(G) \qquad (3)$$

$$p_\bullet^i(\tilde{H}^4) = \left|\breve{G}_i^C(\tilde{H}^4)\right|^{-1} \sum_{G \in \breve{G}_i^C(H^4)} p_\bullet^i(G)\, \nu(G) \qquad (4)$$

Notice that if the random variables describing the reliability state of each network processor are mutually independent as well as the probability of damage of any processor is the same, then $p_j^i(\tilde{H}^4)$ as well as $p_\bullet^i(\tilde{H}^4)$ indicate the possibility of occurrences that the structure $G \in \breve{G}_i^C(H^4)$ with order i will be degraded (after the processor failure of this structure) to the structure with order j and that the network would become unfit for further usage accordingly.

On the base of formulas 3 – 4 and results from section III as well as geometric forms of hypercube network working structures given in [9], the characteristics of degradation of the cyclic network with an initial H^4-type logical structure were calculated (Table 1).

Table 1 $\{\langle p^i(\tilde{H}^4); p_\bullet^i(\tilde{H}^4) \rangle : i \in \{5, ..., 13\}\}$

i \ j	i−1	i−2	i−3	i−4	i−5	$p_\bullet^i$
16	1					0
15	1					0
14	0,870	0,130				0
13	0,796	0,204				0
12	0,604	0,267	0,119	0,005	0,005	0
11	0,541	0,247	0,138	0,034	0,033	0,007
10	0,384	0,218	0,149	0,139		0,112
9	0,316	0,103	0,226			0,355
8	0,032	0,323				0,645
7	0,245					0,755
6						1

By using Table 1 it is easy to determine the probability distribution of $P_{H^4}\{\xi = k\}, (k \in \{1, 2, ...\})$ numbers of consecutive failures of the network processors with an initial logical structure H^4, after which in accordance with the accepted rule of degradation it ceases to be a cyclic network of at least six processors, with the assumption that processors failures are mutually independent random variables with the equal probability (see Fig. 6).

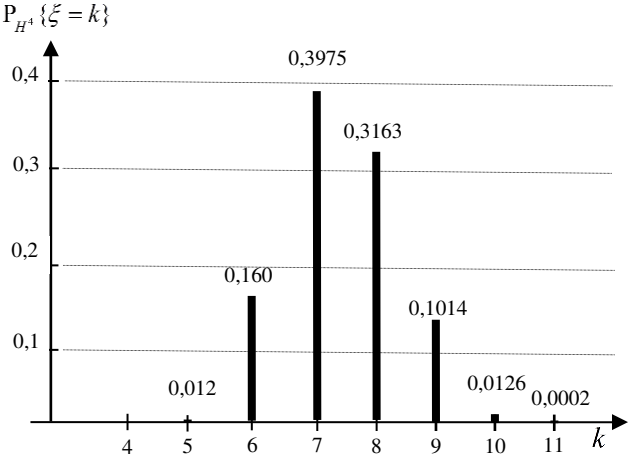


Fig. 6 Probability distribution of the number of consecutive hypercube processors failures, after which it ceases to be a cyclic network with at least six processors

Notice that to state the value of $P_{H^4}\{\xi = k\}$ it is sufficient to just calculate the probability sum of changes of the network structure from the initial state to the final state of all paths with the length of k.

V. CONCLUSION

On the base of the conditions of the existence of certain types of perfect deployments in hypercube structures established in Section III and the catalog of geometric forms of the hypercube network working structures given in [9], [22], [23] the average number of working processors for selected resources deployments and given degree of network degradation was determined. It characterizes of a loss of the computational potential of the degradable hypercube network for the specific perfect deployment along with the process of its degradation.

Obtained characteristics of the impact of network degradation on the average number of processors working in the network and the probability distribution of the number of consecutive processors failures, after which it loses possibility of perfect resources deployment of type $(m,1)$ indicates that 4-dimensional hypercube network characterized by a high resistance for processors failures. It turned out that if the number of faulty processors is less than 6, the probability that for the network exists the perfect placement $(m,1)$ is equal 0,998 and the average number of processors working is equal 8,17. (Table 2).

Table 2 The average numbers of working processors for $(1,1|G)$ -perfect placement

p	16	15	14	13	12	11	10	9	8	7
$\Re^{C}_{(1,1\|G)}(p)$	12	11	10	9,19	8,46	8,17	6,92	6	5	4,57

Generalized cost of information traffic in a network for a given deployment of resources processors depends on the nature of the tasks performed by the network. Such an analysis is not the subject of this paper. This is another problem which can be examined by using simulation methods for a specified (m,d) -perfect deployment and a particular type of task load of the network.

REFERENCES

[1] I. Cohen, O. Rottenstreich, I. Keslassy, Statistical Approach to Networks-on-Chip, IEEE Transactions on Computers, Vol. 59, No. 6, 2010, 748–761.

[2] S. Murali, D. Atienza, P. Meloni, S. Carta, L. Benini, G. De Micheli, L. Raffo, Synthesis of Predictable Network-on-Chip-Based Interconnect Architectures for Chip Multiprocessors, IEEE Trans. Very Large Scale Integration Systems, Vol. 15, No. 8, 2007, 869–880.

[3] T. Ishikawa, Hypercube multiprocessors with bus connections for improving communication performance, IEEE Trans. Computers, Vol. 44, No. 11, 1995, 1338–1344.

[4] A. B. Izadi, F. Özunger, A Real-Time Fault_Tolerant Hypercube Multiprocessor, IEE Proceedings – Computers and Digital Techniques, Vol. 149, No. 5, 2002, 197–202.

[5] D. Wang, Diagnosability of Hypercubes and Enhanced Hypercubes under the Comparison Diagnosis Model, IEEE Transactions on Computers, Vol. 48, No. 12, 1999.

[6] P. Santi, S. Chessa, Reducing the Number of Sequential Diagnosis Iterations in Hypercubes, IEEE Transactions on Computers, Vol. 53, No. 1, 2004.

[7] S.-Y. Hsieh, C.-W. Lee, Pancyclicity of Restricted Hypercube-Like Networks under the Conditional Fault Model, SIAM J. Discrete Math.", Vol. 23, No. 4, 2010, 2100–2119.

[8] J. Chudzikiewicz, Sieci komputerowe o strukturze logicznej typu *hipersześcianu*, Instytut Automatyki i Robotyki, Wydział Cybernetyki, WAT, Warszawa 2002, str. 113. (in Polish)

[9] Z. Zieliński, *Podstawy diagnostyki systemowej sieci procesorów o łagodnej degradacji i strukturze hipersześci*anu, Wojskowa Akademia Techniczna, Warszawa, 2012, pp. 179. (in Polish)

[10] N. Tzeng, G. Feng, Resource allocation in cube network systems based on the covering radius, IEEE Trans. Parallel Distributed Systems 7 (4) (1996) 328–342.

[11] H. Chen, N. Tzeng, Efficient resource placement in hypercubes using multiple-adjacency codes, IEEE Trans. Comput. 43 (1) (1994) 23–33.

[12] B. F. AlBdaiwia, B. Bose, On resource placements in 3D tori, Journal of Parallel Distributed Computer vol. 63, 2003, pp. 838–845.

[13] B. F. AlBdaiwia, B. Bose, Quasi-perfect resource placements for two-dimensional toroidal networks, Journal of Parallel Distributed Computer vol. 65, 2005, pp. 815-831.

[14] M. M. Bae, B. Bose, Resource Placement in Torus-Based Networks, IEEE Transactions on Computers, vol. 46, no. 10, October 1997, pp. 1083-1092.

[15] N. Imani, H. Sarbazi-Azad, A.Y. Zomaya, Resource placement in Cartesian product of networks, Journal of Parallel Distribiuted Computer 70, 2010, pp. 481-495.

[16] P. Moinzadeh, H. Sarbazi-Azad, N. Yazdani, Resource Placement in Cube-Connected Cycles, The International Symposium on Parallel Architectures, Algorithms, and Networks, IEEE Computer Society, 2008, pp. 83-89.

[17] J. Chudzikiewicz, Z. Zieliński, Reconfiguration of a processor cube-type network, PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), R. 86, NR 9/2010, pp. 149-153.

[18] R. Kulesza, Z. Zieliński, *The life period of the hypercube processors'* network diagnosed with the use of the comparison method, Monographs of System Dependability - Technical Approach to Dependability, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 2010, pp. 65-78.

[19] Z. Zieliński, Ł. Strzelecki, R. Kulesza, Diagnosability characterization of the 4-dimensional cube type soft degradable *processors' network*, Monographs On System Dependability – Problems of Dependability and Modeling, Oficyna Wydawnicza Politechniki, Wrocław, 2011, pp. 283-296.

[20] A. Reddy, P. Banerjee, Design, analysis, and simulation of I/O architectures for hypercube multiprocessors, IEEE Trans. Parallel Distributed Systems 1 (2) (1990) 140–151.

[21] J. Chudzikiewicz, Z. Zieliński, Determining a non-collision data transfer paths in hypercube processors network, Embedded Systems - High Performance Systems, Applications and Projects, InTech, 2012, pp. 19-34.

[22] R. Kulesza, Z. Zieliński, Metoda generowania struktur logicznych *sieci procesorów o łagodnej degradacji typu 4*-wymiarowego *sześcianu*, Biuletyn WAT, Vol. LX, Nr 4, 2011. (in Polish)

[23] R. Kulesza, Z. Zieliński, Wyznaczanie struktur logicznych sieci *procesorów o łagodnej degradacji i strukturze 4*-wymiarowego *hiepersześcianu*, Biuletyn WAT, Vol. LXI, Nr 4, 2012, pp.1-14. (in Polish)

[24] B. Korzan, *Elementy teorii grafów i sieci. Metody i zastosowania*, Wydawnictwo Naukowo-Techniczne, Warszawa, 1978. (in Polish)

[25] J. Kulikowski, *Zarys teorii grafów*, Wydawnictwo Naukowo Techniczne, Warszawa, 1986. (in Polish)

[26] R. Kulesza, Podstawy diagnostyki sieci logicznych i komputerowych, Instytut Automatyki i Robotyki, Wydział Cybernetyki WAT, Wydanie II, Warszawa 2000, str. 222. (in Polish)

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 75–78

Wireless Sensor Based Monitoring and Content Management System for Agricultural Application

Wen-Yaw Chung[1], Meo Vincent C. Caya[1,2], Chien-Lin Chen[1]
[1] Department of Electronics Engineering, Chung Yuan Christian University,
Zhongli, Taiwan, R.O.C.
[2] Department of Electrical, Electronics and Computer Engineering,
Mapua Institute of Technology, Manila, Philippines
Email: eldanny@cycu.edu.tw, mvccaya@mapua.edu.ph, g9876606@cycu.edu.tw

***Abstract*—This paper presents a Zigbee-based wireless sensor network system with an integrated hardware implementation and content management function for web-based monitoring in a greenhouse environment. The Zigbee is based on the TI CC2530 System on Chip and composed of key sensors such as a temperature, relative humidity and light intensity sensors. The content management system is based on open source software Joomla. The system provides the semi-real time monitoring of the environmental parameters with an epoch time of five minutes. In addition, it also provides a monitoring graph and the mean value of the parameter being measure with its High and Low value for a particular day monitoring. The proposed system has been implemented and tested in Dai-Yun Organic Farm in Taiwan that cultivate vegetable crops, the system performance showed the potential applications for precision agriculture and other fields.**

I. Introduction

AGRICULTURE is considered one of the most important sources of income and food production worldwide. Agricultural activities are affected by environmental factors such as ambient and surface temperature, relative humidity and light intensity; in turn, these factors have significant impact in the production and quality of agricultural crops. Therefore, it is important to monitor environmental parameters and make effective use of the resources through the design and implementation of a greenhouse monitoring systems.

Greenhouse production systems have become the alternative production practice to satisfy the consumer demand of healthier, safer and higher quality produce in a year-round manner, while implementing environmentally friendly methods that make efficient use of resources such as land, water, labor, capital and energy. However, they are highly dependent on energy, skilled labor, effective management and increased knowledge of growing specific crops [1], [2]. The main purpose of such systems is to produce a high quality produce at high production rates, consistently, in an economic, efficient and environmentally sound manner. To achieve this level of productivity, accurate monitoring and control of several aspects of the crop including plant health, growth rate, development and plant growth-mode, must be implemented during the production cycle.

In terms of network technology, many studies consider the circuit implementation of the Zigbee technology, the technology of choice for wireless sensor network due to its low-power consumption requirement, flexible network scale and cost effective design specifications. Reference [3] used CC2430 System on chip (SoC) based on Zigbee technology and integrates environmental sensors that monitor soil temperature, soil humidity, air temperature and relative humidity, ambient light, and carbon dioxide for greenhouse application. Reference [4] used wireless microcontroller JN5121 that implements a Zigbee compliant sensor node, and integrated SHT1x, TSL2550, M25P10 that measure temperature, relative humidity, and luminance respectively. Gathered data from a greenhouse unit were stored in a flash memory in sleep-mode and sent back to the coordinator after waking up. The use of Wi-Fi technology was also explored. However, the protocol usually utilized a PC-based system intended as a substitute for wired LAN with a rather high power consumption requirement for a wireless sensor node implementation [5], [6].

II. SYSTEM IMPLEMENTATION

A. System Architecture

The greenhouse environmental monitoring system based on a wireless sensor network consists of the sensor nodes, the coordinator node, the remote computer, and the webserver. The sensor node consists of the environmental sensor such as the temperature, relative humidity and light intensity sensor that measures particular parameters inside the greenhouse. The coordinator node serves as the communication bridge between the remote computer and the wireless sensor nodes. It is also responsible in the creation of the Zigbee network that adds other nodes in the system.

The data from the sensor nodes are stored in the remote computer database which is based on the MySQL platform and the system is being hosted in a remote webserver which consists of an Apache Webserver, MySQL database and PHP: Hypertext Preprocessor. The system block diagram is shown in fig. 1.

This work was supported by National Science Council, Taiwan for funding this project (NSC 102-2221-E-033-066)

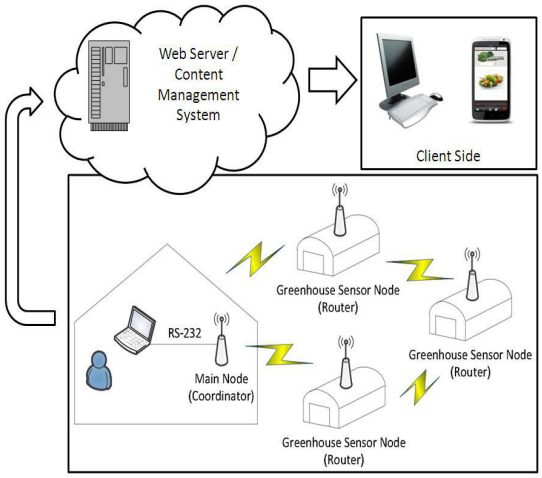


Fig. 1 System Block Diagram

In the webserver, a content management system (CMS) has been installed to manage the publishing, editing, and modifying content as well as maintenance from a central interface [7]. Joomla CMS platform has been installed in the webserver.

B. Wireless Sensor Node Components

The wireless sensor node shown in fig. 2 is composed of SHT10 digital humidity and temperature sensor, ISL29023 an integrated ambient and infrared light to digital converter, and the main core component TI CC2530, which is a true system-on-chip (SoC) solution for IEEE 802.15.4, Zigbee and RF4CE applications [8].

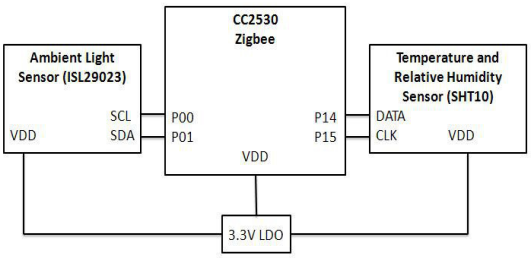


Fig. 2 Sensor Node Block Diagram

SHT10 is a temperature and relative humidity sensors manufactured by Sersirion[9], which measure temperature with a range from minus 55 to plus 125 degree Celsius, and measure relative humidity range from 0 to 100 percent. ISL29023 is an ambient light sensor manufactured by Intersil Company [10], which can measure illumination with range from 1 to 64000 lux. It can be controlled by CC2530 through I2C bus interface.

Coordinator node consists of CC2530 and PL2302. PL2303 is USB to UART transceiver. The coordinator node was used to create a Zigbee network, receive data from and send data to greenhouse sensor nodes, then forward to remote computer. The coordinator node also uses the sensor node PCB but instead of environmental sensor, the USB to UART transceiver has been integrated to serve as a link between the remote computer and the WSN.

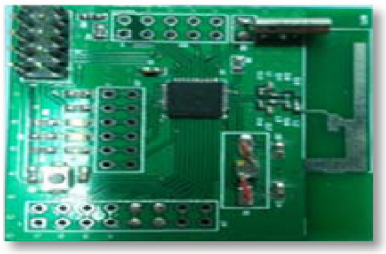


Fig. 3 Actual PCB of the Sensor Node

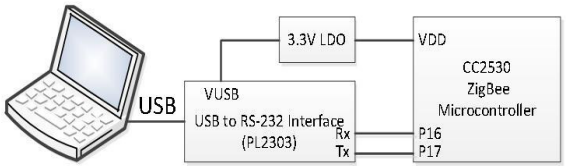


Fig. 4 Coordinator Node Block Diagram

C. Monitoring System Application

The desktop monitoring system application shown in fig. 5, running in the remote computer gets the data coming from the sensor node via the coordinator node and the serial RS-232 interface connected to the computer. It is also responsible in storing of data in the database server. The application program is made from Visual Basic Express 2010.

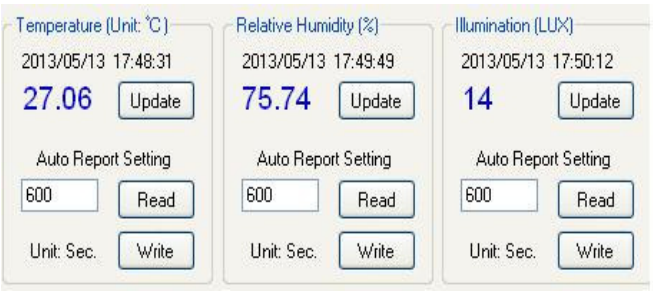


Fig. 5 Monitoring System Application Controls

The application controls in fig. 5 shows that the epoch time for the sensor node can be adjusted through the auto report setting controls. The read button check the current epoch time, or check whether the set reporting has been successfully updated in the sensor node. The write button updates the epoch time embedded in the sensor node.

D. Content Management and Monitoring System

The content management system shown in fig. 1 is based on the Joomla CMS platform. It consist of frontend showed in fig. 6, which is readily available for client machines or users of the website, and the backend showed in fig. 7, wherein the maintenance, publishing, editing and modifying of the website content has been made.

The Joomla CMS platform advantages are the ease-of-use, extensibility and most of all it is an open source solution that is freely available to everyone [11].With this

regard, integration of the CMS platform became a vital component of the Wireless Sensor Network Monitoring System.

Internet connection is a requirement in accessing the website in a client machine. Internet disconnection does not in any way affect the WSN monitoring, because internet connection is not a requirement for WSN and the remote computer transmission of environmental data.

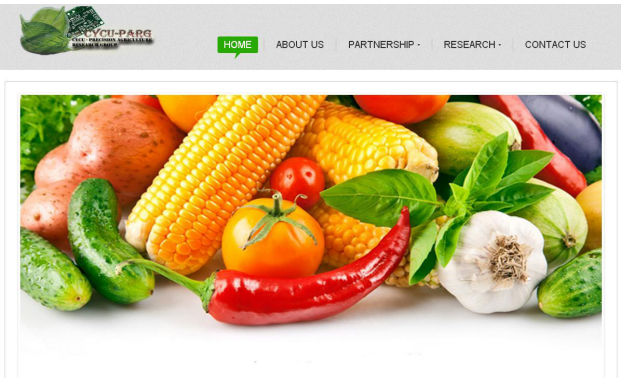


Fig. 6 Website Frontend

The website frontend page shown in fig. 6 provides a user browsing interface and the administrator backend shown in fig.7 offers the necessary tools and applications for the maintenance and editing of the website. The administrator can create users that are authorized to manipulate and maintain the website. The administrator can decide which article or section that a particular group of user can access in the website. The monitoring section of the system can be viewed solely by specific user or it can be made public.

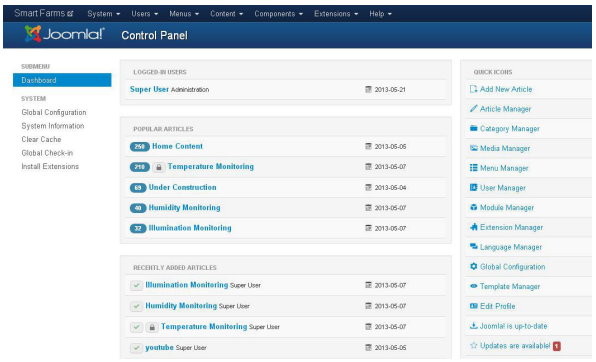


Fig. 7 Administrator Backend

The environmental data stored in the database was used to present in the content management system report in a form of a graph. Also, the mean, high and low measured value for the current day monitoring was computed and included in the report. Stored data in the database was retrieved using SQL command. Fig. 8-10 shows the environmental parameters being monitored.

The data from the WSN has been presented in the CMS platform, fig.8 shows the temperature monitoring with time interval for each data of five minutes.

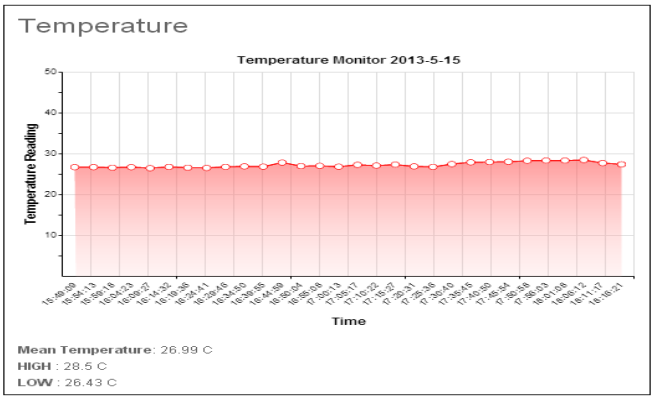


Fig. 8 Temperature Monitoring

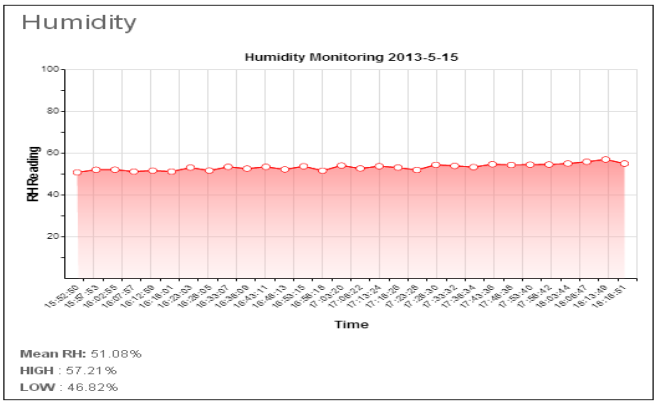


Fig. 9 Humidity Monitoring

The graphs demonstrated in fig. 8-10 show the last thirty reading for the particular environmental parameter. The Mean, High, and Low measured value was based on the current running data sent in the database and it is not just been based on the last thirty reading.

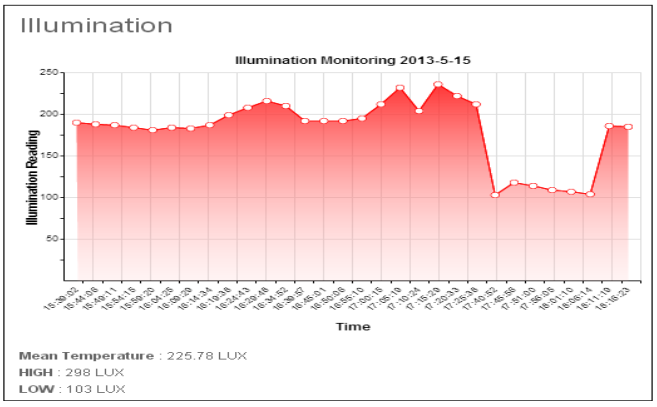


Fig. 10 Light Intensity Monitoring

E. Dai-Yun Organic Farm

Dai-Yun Organic farm is at the junction of Longtan and Guanxi, Taiwan (R.O.C.) which has an altitude about 250 meters. It had been verified and certified by Taiwan's organic production association in June 03, 2008. Currently, farm cultivating area is 8.8 hectares, (1.6 hectares for greenhouse, 7.2 hectares for open field). There are approximately 100 types of crops cultivated each year. Approximately 2000 bags is being harvest per day. Dai-Yun Organic Farm target

market are organic shops, organic supermarket and fine dining restaurants. There are approximately 150 locations in the province selling their products.

Dai-Yun Farm provides high quality and high throughput of vegetable crops annually; with this regard Dai-Yun accumulates vital information in the field of organic farming and the WSN plays a vital role in the monitoring, storage and analysis of the data from the greenhouse farm. The main goal is the repeatability of the process wherein high quality and high yield in the farm will be maintain or can be optimized. Results and finding will be made available through the content management system, and farmers, students, and researchers can be benefited in information that we can provide.

Fig. 11 Dai-Yun Organic Farm

III. FUTURE WORK

The wireless sensor network is composed of three sensors which are the temperature, relative humidity and light intensity sensors. We are in the process of integrating other sensors; carbon dioxide, soil moisture and pH sensors. Soil Moisture sensor will be integrated in the system for irrigation scheduling purposes, while the pH sensor will be used in determining the alkalinity and acidity of the irrigation water; various crops have different pH of water or soil requirement. Carbon dioxide sensor is one of the important factors that affect plant growth, monitoring of the carbon dioxide will be vital information in the Dai-Yun organic farm.

Advancement in the field of smartphone and cloud computing technologies paves a way for a new approach of providing information and services. With this regard, we are taking into consideration of integrating smartphone application that will maximize the capability of our Environmental Monitoring System, and offer cloud computing services.

The Content Management System can be extended not only in the greenhouse environmental monitoring but we can also extend its capability in developing “Smart Home” Monitoring and automation.

ACKNOWLEDGMENT

This work would like to extend its gratitude to Chip Implementation Center (CIC) for the technical support given.

REFERENCES

[1] G M. H. Jensen, “Controlled Environment agriculture in deserts, tropics and temperate regions - A World Review,” Acta Hort. (ISHS), vol. 578, pp. 19-25, 2002.
[2] G. A. Giacomelli, R. L. Patterson, and P. D. Sadler, “Telepresence technologies and practices for enabling remote semi-autonomous CEA Food Production,” ISHS Acta Horticulturae, vol. 761, pp. 21-31, 2007.
[3] Su Baishun; Han Yuhui, "Design of Greenhouse Wireless Monitoring System Based on ZigBee," E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on , vol., no., pp.1,4, 7-9 Nov. 2010
[4] Zhou Yiming; Yang Xianglong; Guo Xishan; Zhou Mingang; Wang Liren, "A Design of Greenhouse Monitoring & Control System Based on ZigBee Wireless Sensor Network," Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on , vol., no., pp.2563,2567, 21-25 Sept. 2007
[5] IEEE, Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-speed physical layer extension in the 2.4 GHz band. IEEE standard 802.11b. In IEEE. The Institute of Electrical and Electronics Engineers Inc: New York, NY, USA, 1999.
[6] Anastasi, G.; Farruggia, O.; Lo Re, G.; Ortolani, M. Monitoring high-quality wine production using wireless sensor networks. In 42st Hawaii International International Conference on Systems Science (HICSS-42 2009): Waikoloa, Big Island, HI, USA, 2009
[7] Content Management Bible, Bob Boiko. John Wiley & Sons, 2005.
[8] Texas Insteruments, “A True System-on-Chip Solution for 2.4-GHz IEEE 802.15.4 and ZigBee Applications”, Product datasheet, Feb 2011
[9] Sensirion company, " SHT1x (SHT10, SHT11, SHT15) Humidity and Temperature Sensor ", product datasheet, ver4.3, May 2010.
[10] Intersil Americas Inc, “Integrated Digital Ambient Light Sensor with Interrupt Function”, product datasheet, March 3 2009.
[11] What is Joomla?.(2013, May 15). Retrieved from http://www.joomla.org/about-joomla.html

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 79–82

Wireless Sensor Network Based Soil Moisture Monitoring System Design

Wen-Yaw Chung[1], Jocelyn F. Villaverde[1,2], Janine Tan[1]
[1]Department of Electronics Engineering
Chung Yuan Christian University
Zhongli, Taiwan R.O.C.
[2] School of Electrical, Electronics and Computer Engineering
Mapua Institute of Technology
Manila, Philippines
Email: eldanny@cycu.edu.tw, jfvillaverde@mapua.edu.ph, tan.janine.s@gmail.com

Abstract—**Soil moisture is the amount of water in the soil. It plays a crucial role for efficient photosynthesis, respiration, transpiration and transportation of minerals and other nutrients through the plant. Proper irrigation schedule is very critical to plant growth. This paper presents a soil moisture monitoring system which comprises of commercially available soil moisture sensor Decagon EC-5, low power nRF24L01 wireless transceiver, and MPC82G516A microcontroller. The sensed soil moisture content from the EC-5 sensor node will be transmitted to the coordinator via 2.5GHZ wireless signal. The coordinator will transmit the data to the computer using RS232 interface. The key components used in the proposed system are low cost and more flexible. The locally developed MPC82G516A microcontroller consumes low power due to its 1-T instruction time. In addition, the EC-5 sensing electrode also has advantages of easy installation and replacement in the farm. The developed system provides a better data transmitted and processed wirelessly and it can serve as a basis for efficient irrigation scheduling.**

I. Introduction

WIRELESS sensor network (WSN) is now used in broad area of applications such as industrial monitoring, healthcare application, home automation and traffic control. The implementation of wireless sensor network does not only concentrate on the applications mentioned. Farmers can take advantage of the advancement of technology. Intelligent farming is now implemented to monitor the status of the field environment. The parameters that are monitored in the greenhouse are temperature, relative humidity, light intensity and others that have effect on the quality of produce.

One factor that affects the quality of crops is the content of water in the soil, also called soil moisture. It is a major component of the soil relative to plant growth. If the soil moisture is optimum for plant growth, plants can readily absorb water. Irrigation schedule is needed to meet the increasing demand of food. Soil moisture can be categorized based on the volumetric water content. This will determine the saturation level of the soil [1]. Modern intelligent agriculture uses wireless transmission to monitor other parameters in the greenhouse such as temperature, humidity and other environmental information [2].

One primary concern on wireless transmission is the power consumption [3]. The problem with other wireless transceiver like Zigbee and Bluetooth is the power consumption. As the range of the transmission becomes longer, the module needs more power to transmit the data. Another consideration also is the cost of the system. In this paper, our group implements a low power 2.4 GHz wireless transceiver for soil moisture monitoring and irrigation scheduling in agricultural greenhouse. The wireless sensor network module was developed in our lab was low cost because our group used locally commercial components and design our own sensor and coordinator node circuit board.

II. System Overview

The WSN soil moisture monitoring system is composed of sensor node and coordinator. Fig. 1 shows the system's block diagram. The sensor node is comprises of the EC-5 sensor, MPC82G81GA microcontroller, and nRF24L01 transceiver. The EC-5 sensing electrode gives voltage input to the MPC82G516A microcontroller. The analog to digital converter (ADC) of the microcontroller will process the voltage input into digital forms and send it to nRF24L01. The sensor node transmits the data to the receiver node. The receiver node is connected to the computer via RS232 as shown in right side of Fig. 1. The transmitted data will be stored in the database to categorize the soil moisture content. This will also be the basis for water irrigation scheduling

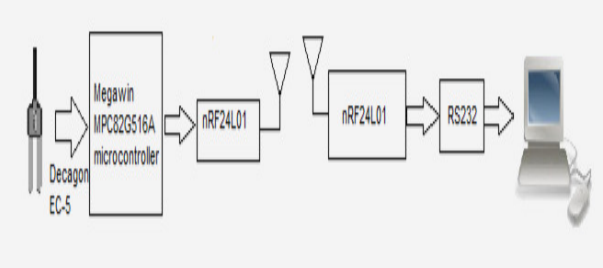


Fig. 1 WSN Soil Moisture System's Block Diagram

This work was supported by National Science Council, Taiwan (NSC 102-2221-E-033-066).

III. HARDWARE DESIGN

A. Decagon EC-5 Soil Moisture Sensor

The EC-5 is a commercially available soil moisture sensor provided by Decagon Devices. It measures the dielectric constant of the soil in a frequency domain technology. It uses 70 MHz frequency to make it insensitive to soil texture [4]. It measure volumetric water content from 0 to 100% in determining the soil moisture. The volumetric water content of saturated soil range from 40% to 60%. The range varies on the different type of soil. Fig. 2 shows the EC-5 sensing electrode and enclosed circuitry.

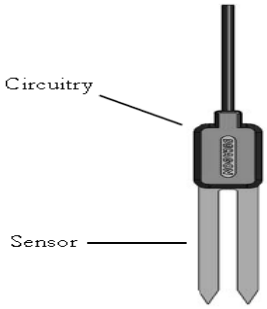


Fig. 2 EC-5 Physical Diagram [4]

The circuitry of the sensor readout comprises of three components namely: the phase detector, oscillator and the buffer [5]. The phase detector is connected to the transmission line of the sensor and the output of the buffer circuit. A low pass filter is included to give an output voltage that is constant and proportional to the difference in phase between the inputs of the phase detector. Oscillator circuit provides an output with the desired frequency. It also includes an astable multivibrator the gives continuous state to the oscillator.

The EC-5 sensor was designed to install easily in the soil. There is no need to dig the soil. To place the EC-5 in the soil, just push it until the desired depth as shown in Fig 3.

Fig. 3 EC-5 Sample Installation [4]

B. Megawin MPC82G516A

The MPC82G516A is compatible with the instruction set of 8051 microcontroller. Idle mode and Power-Down Mode are the two power saving modes of MPC82G516A that reduce the power consumption. Another feature of this microcontroller to lessen the power consumption is that it can operate in a lower speed by using the 8-bit system clock [6]. There is no need for external analog to digital converter (ADC) circuit because it has its own 10-bit ADC.

C. nRF24L01

The nRF24L01 is a 2.4GHz transceiver for low power wireless application. It has power mode and standby mode as part of the power management feature of the chip [7]. Because it has a built in state machine, it can switch to four modes of operations. In the power down mode the chip consume less current because the nRF24L01 is disabled but all the registers from the Serial Peripheral Interface (SPI) are present and can be activated anytime. In the Rx mode, the nRF24L01 act as a receiver. This is the active mode of the chip where it consumes power. The Tx mode transmit the packet. This is in the active mode when the packet is ready to transmit. It remains in the Tx mode until it finished to transmit all the packet. The standby mode minimizes the current consumption but the crystal oscillator is still active. It will enter in the standby mode if there the Rx mode or Tx mode is not active. The mode of operation is based on the single direction relaying transmission method [8] as shown in Fig 4.

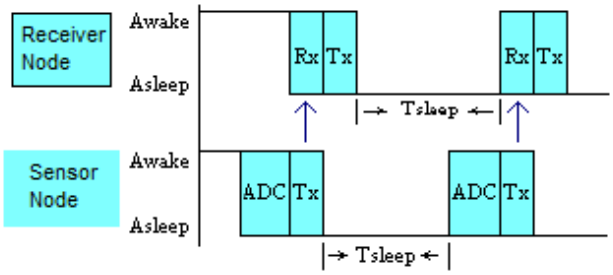


Fig. 4 Relaying Transmission

IV. METHODOLOGY

The sensed soil moisture values in this paper used volumetric water content. Volumetric water content is the ratio of the volume of water and the total sample volume of dry soil. The soil sample used to get the calibration curve of the EC-5 sensor came from Dai-Yun Organic Farm [9]. In getting the calibration curve of the EC-5 soil moisture sensor our group used the standard direct method of determining the soil moisture [10].

The analog data obtained from the characteristic curve served as the voltage input of the analog to digital converter of the MPC82G516A microcontroller. The microcontroller is programmed using Keil uVision4 IDE based on the

obtained calibration curve of the EC-5 sensor. The wireless transceiver nRF24L01 is connected to the MPC82G51A microcontroller to transmit the data to the receiver node. Using Tera Term terminal emulator, the transmitted data can be viewed to the computer monitor and store in the database. The stored values will be used in making the irrigation schedule of the greenhouse. These values determined the volumetric water content of the soil. The volumetric water content will be the basis of the condition of the soil. There are four classifications that will be considered in the irrigation scheduling namely: dry soil, slightly moist soil, moist soil and wet soil. With this regards, the farmer can decide how frequent water is needed in the farm.

A low cost wireless sensor network node is illustrated in Fig 5. The network module can be integrated into different sensors. The components are locally available making it less costly with the other sensor nodes available in the market.

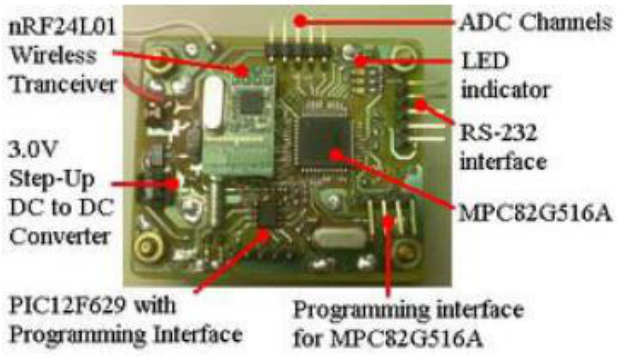


Fig. 5 Wireless Receiver Node Board [8]

A repeater node is also integrated in this wireless sensor network to extend the range of transmission distance and to resolve the line of sight problem occurring in the WSN. In order to maintain the low power consumption of the system a sleep mode and active mode is established in the repeater node. The repeater node will enter the sleep mode after the last packet is transmitted to the main coordinator.

V. Results

The result of the calibration curve is depicted in Fig 6. The accuracy of the output values were compared to the output reading of the ECH20 utility software connected to the Em5R data logger. The output reading of the two devices showed that the output voltage is proportional to the water content of the soil. The saturation level of the soil starts from 35% and 40%. These values obtained from the sensor node will serve as the basis for irrigation scheduling. Fig 7 illustrated the sample output window of the ECH20 utility software.

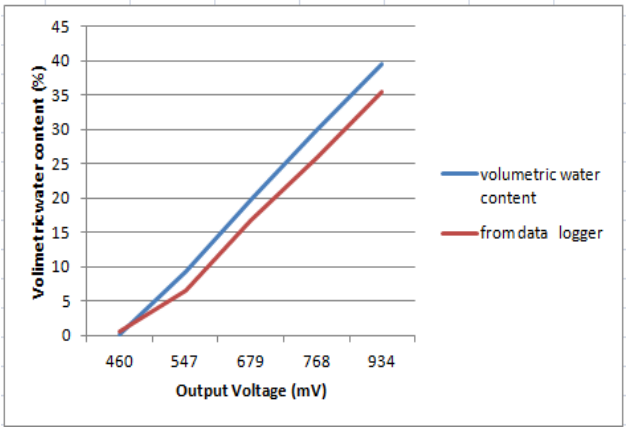


Fig. 6 Characteristic or calibration curve

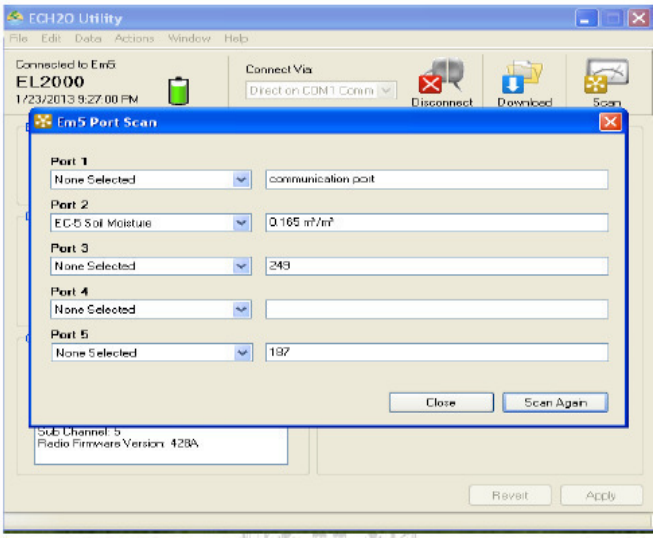


Fig. 7 Data logger Sample Output Window

The output of the wireless sensor network is shown in Fig 8. The transmitted data were displayed in the computer monitor. These values are stored in the database for developing a user interface design for irrigation schedule.

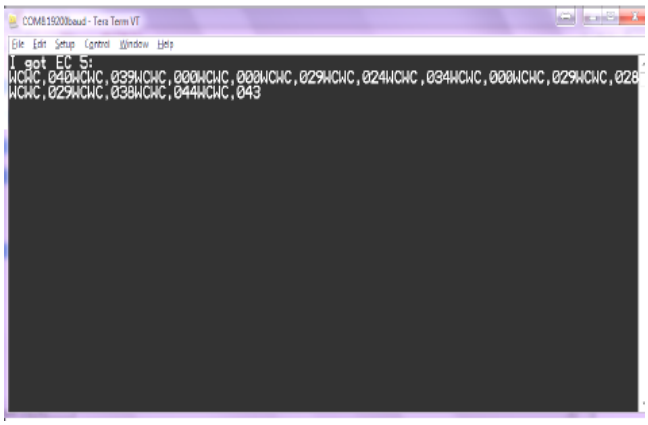


Fig. 8 Tera Term Wireless Transmission Output

VI. Conclusion

This paper implemented a low cost wireless sensor network system for soil moisture monitoring. The flexibility of the sensor node and transceiver node to integrate to other type of sensor was tested and implemented on this work. The characteristic curve proved the accuracy of the sensor in determining the soil moisture content. It also showed that volumetric water content can be used to monitor and schedule the irrigation in the agricultural greenhouses more efficiently.

VII. Future work

The volumetric water content of the soil moisture can be classified into four categories. Dry soil, slightly moist, moist, and wet are the possible classification of the soil moisture to better understand the values. Fuzzy logic approach will be use to determine the soil moisture characteristic. The design of the graphical user interface is also being considered. This system can also be integrate in a web design application or include in the parameters of the precision agriculture.

Acknowledgment

This work would like to extend its gratitude to Chip Implementation Center (CIC) for the technical support given.

References

[1] Rafael Muñoz-Carpena, Sanjay Shukla, Kelly Morgan, “Field Devices for Monitoring Soil Water Content”, BUL343, June 2004.

[2] Zheng Ma, Xing Pan, “Agricultural Environment Information Collection System Based on Wireless Sensor Network”, IEEE Global High Tech Congress on Electronics”, 2012, pp. 24–28.

[3] Zhu Yao-lin, Zhang Gao-qiang, Zhu Lei, Xu Jin, “Design of Wireless Multi-point Temperature Transmission System Nased on nRF24L01”, IEEE, 2011, pp. 780-783.

[4] Decagon Devices, “EC-5 Soil Moisture Sensor: User’s Manual Version, 2” 2012 [online]. Available http://www.decagon.com/assets/Uploads/EC-5-Manual.pdf.

[5] Gaylon S., Campbell, Pullman, WA (US), Warren C. Greenway, Cambridge, ID(US), “Moisture Detection Apparatus and Method”, United States, US6 904 789 B2, June 14, 2005.

[6] Megawin, “MPC82G516A 8-bit Microcontroller: Datasheet version A4”, 2008 [online]. Available http://www.silaresearch.com/data/82g516a.pdf.

[7] Nordic Semiconductor, “nRF24L01 Single Chip 2.4GHz Transceiver: Product Specifications V2.0,” July 2007. [Online]. Available http://www.nordicsemi.com.

[8] Wen-Yaw Chung, Chien-Lin Chen, Jyen-bin Chen, “Design and Implementation of Low Power Wireless Sensor System for Water Quality Monitoring”, IEEE, 2011.

[9] Wen-Yaw Chung, Janine Tan, Meo Vincent Caya, Arnold Paglinawan, “Wireless-Based Soil Moisture Monitoring System for Irrigation Scheduling in Greenhouse”, IECEP Conference Proceeding, 2012.

[10] C.A. Black, D.D. Evans, J.L. White, L.E. Ensminger, F.E. Clark, “Methods of Soil Analysis, Part 1: Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling, 1965.

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 83–88

Time synchronization in Wireless Sensor Networks

Krzysztof Daniluk
Warsaw University of Technology, Institute of Control and Computation Engineering
Email: K.Daniluk@stud.elka.pw.edu.pl

***Abstract*—In this paper a survey of selected topics concerning development of wireless sensor network systems, from the point of view of time synchronization algorithms, is presented and discussed. The focus is on currently used time synchronization techniques. Finally is presented a novel concept for time synchronization scheme.**

I. Introduction

An accurate and consistent sense of time is essential in sensor networks, especially in wireless sensor networks. Distributed wireless sensor networks need time synchronization to coordinate the communication, computation, sensing and actuation of distributed nodes. Used techniques for time synchronization, as well as the newest solution in this field are discussed. The remainder of this paper is organized as follows. The reasons for using time synchronization are described in **chapter II,** clock and frequency issues, as well as the fundamental technique for clock synchronization are presented in **chapter III**, review of time synchronization algorithms for wireless sensor networks are discussed in **chapter IV**, packet delay model is described in **chapter V**, novel concept in time synchronization for wireless sensor networks is presented in **chapter VI.** The paper concludes in **chapter VII.**

II. Reasons for time synchronization

There can be distinguished different reasons to use time synchronization, the most crucial are presented below.

A. Sleep scheduling

One of the most significant sources of energy savings is turning off the radios of sensor devices in the situation when they are not active. It means, that without proper synchronization such technique cannot exist and work correctly and efficiently.

B. Medium-access

TDMA-based medium-access schemes require that nodes are synchronized. There is a need to assign distinct slots for collision-free communication.

C. Coordinated actuation

Distributed actuators require synchronization in order to coordinate the actuators through distributed control algorithms. Time synchronization is again a crucial element in this process.

D. Coordinated signal processing

Time stamps are needed to determine which information from different sources can be fused/aggregated within the network.

E. Multi-node cooperative communication

Multi-node cooperative communication techniques involve transmitting in-phase signals to a given receiver. Such techniques [1] have the potential to provide significant energy saving and robustness, but again, there is required synchronization, as key element of the communication process.

III. Clock and frequency

The clock at each node consists of timer circuitry, often based on quartz crystal oscillators. After each K ticks/interrupts of the timer the clock is incremented.

In real-life conditions, timer circuits, especially in low-end devices, are unstable and error prone.

Let assume following notation:
fo is the ideal frequency,
deltaF – the frequency offset,
Df – the drift in the frequency,
Rf(t) an additional random error process.
It means, that instantaneous oscillator frequency *Fi(t)* of oscillator *i* at time *t* can be modeled as follows:

$$Fi(t) = fo + deltaF + Dft + Rf(t) \quad [2]$$

Frequency drift and the random error term may be neglected, so can have a simpler linear model for clock non-ideality, which is following.

$$Ci(t) = (Alfa\_i + Beta\_i)(t)$$

Where *Alfa_i* is the clock offset at the reference time t=0 and *Beta_i* the clock drift (rate of change with respect to the ideal clock).

It means, that the more stable and accurate the clock, the closer *Alfa_i* is to 0, and the closer *Beta_i* is to 1.

A clock is said to be fast if *Beta i* is greater than 1, and slow otherwise.

It is good to underline, that typical sensor nodes have drift rate of +- 40microseconds per second.

This work was partially supported by National Science Centre grant NN514 672940

Very often are specified manufactured clocks with a maximum drift rate parameter *p*, such that *1-p <= Beta_i <= 1+p*.

Motes, so typical sensor nodes, have *p* values on the order of 40 ppm (parts per million), which corresponds to a drift rate of +-40 microseconds/second.

That means, that any two clocks that are synchronized once may drift from each other at a rate at most *2p*.

On the other hand, to keep their relative offset bounded by gamma seconds at all times, the interval *Tsync* corresponding to successive synchronization events between these clocks must be kept bounded using following formula

Tsync <= gamma/2p

where relative offset is bounded by *gamma* seconds at all times

A. Fundamental technique – Cristian's algorithm

Clock synchronization is the synchronization of both clock drift and clock offset.

The two clocks are synchronized if they are running on the same frequency and showing similar time.

Clock Drift: it is the difference in frequency of the clocks at which they are ticking

Clock Offset: it is the difference of time between two clocks

Accuracy: Accuracy of a clock is how well its time compares with global time

Efficiency: The time and energy needed to achieve synchronization

One of the most significant sources of energy savings is well known mechanism - turning off the radio.

A fundamental technique for two-node clock synchronization is known as Cristian's algorithm [3].

Let assume following, node A sends a request to node B (which has the reference clock) and receives back the value of B's clock, *Tb*. Node A records locally both the transmission time T1 and the reception time T2.

In Cristian's time-synchronization algorithm, there are many sources of uncertainty and delay, which have big influence on its accuracy.

An important issue in synchronization processes is message latency.

Message latency can be decomposed into four components:

- Send time: includes processing time, as well as time taken to assemble and move the message to the link layer
- Access time: includes random delays while the message is buffered at the link layer caused by contention and collisions
- Propagation time: time taken for point-to-point message travel.
- Receive time: time taken to process the message and record its arrival

Approximation of the message propagation time can be done with formula *(T2-T1)/2*.

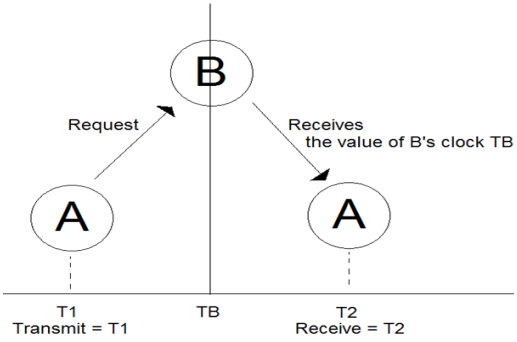


Fig 1. Cristian's synchronization algorithm

If the processing delay is known to be *I*, then it is better to estimate as *(T2-T1-I)/2*.

There can be found more complex approaches, where are taken several round-trip delay samples and are used minimum or mean delays.

B. Network time protocol used on the Internet for time synchronization, based on Cristian's algorithm

The network time protocol (NTP) [4] is used widely on the Internet for time synchronization. The key feature is, that it uses a hierarchy of reference time servers providing synchronization to querying clients, essentially using Cristian's algorithm.

IV. Review of Algorithms for time synchronization in wireless sensor networks

In following chapter are presented and discussed time synchronization algorithms used in Wireless Sensor Networks.

A. RBS – Reference Broadcast Synchronization algorithm

Let assume, that three nodes A, B, C are in the same broadcast domain.

Knowing, that B is a beacon node, i.e. broadcasts the reference signal, that is received by both A and C simultaneously, the two receivers record the local time when the reference signal is received.

Receivers, i.e. nodes A and C then exchange this local time stamp through separate messages.

It is enough for the two receivers to determine their relative offsets at the time of reference message reception.

Presented scheme can be extended to use in greater numbers of receivers.

The key feature of RBS is that it eliminates completely sender-side uncertainty.

There can be seen improved synchronization in scenarios, where are significant sender delays, especially when time

stamping has to be performed at the application layer instead of the link layer.

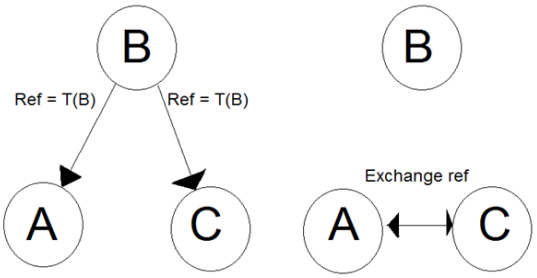


Fig. 2. Time synchronization in RBS algorithm

B. Pair-wise sender-receiver synchronization – timing-sync protocol for sensor networks (TPSN)

The timing-sync protocol for sensor networks (TPSN) [5] is classical sender-receiver synchronization, similar to Cristian's algorithm.

Node A transmits a message, which is locally stamped as T1.

Next, it is received by node B, which stamps the reception time as its local time T2. Then, node B sends the packet back to node A, marking the transmission time locally at B as T3.

At the end the message is received at node A, which marks the reception time as T4.

Assumptions:
Offset between nodes A and B is defined as *DELTA*.
Propagation delay between them is ***d.***

Then,
$T2 = T1 + DELTA + \boldsymbol{d}$
$T4 = T3 - DELTA + \boldsymbol{d}$

Result looks following:
$DELTA = [(T2 - T4) - (T1 - T3)]/2$
$\boldsymbol{d} = [(T2+T4) - (T1+T3)]/2$

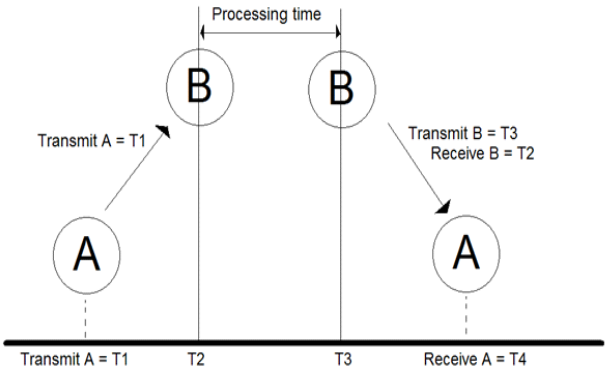


Fig. 3. Time synchronization in TPSN algorithm

The key fact in network-wide time synchronization in TPSN is that time synchronization is obtained level-by-level on a tree structure. For example, nodes at level 1 first synchronize with the root node, then nodes at level 2 synchronize with one node at level 1, it happens until the moment, when all nodes in the network are synchronized with respect to the root.

C. Flooding time synchronization protocol (FTSP)

The flooding time synchronization protocol (FTSP) [6] has big advantage over already presented algorithms, it reduces following sources of uncertainties, which exist in RBS and TPSN algorithms:

- Interrupt handling time
- Modulation/encoding time

Interrupt handling time

Delay in waiting for the processor to complete its current instruction before transferring the message in parts to the radio.

Modulation/encoding time

Time taken by the radio to perform modulation and encoding at the transmitter, as well as the corresponding demodulation and decoding at the receiver.

FTSP uses a broadcast signal from a single sender to synchronize multiple receivers.

Different than in RBS, the sender actually broadcasts a time measurement, and the receivers do not exchange messages among themselves.

Each broadcast provides a synchronization point (a global-local time pair) to each receiver.

In FTSP there are distinguished two main components:

- Multiple time measurements
- Flooded messaging

Multiple time measurements

The sender takes several time stamp measurements during transmission, one after a set of SYNC bytes used for byte alignment. The measurements are normalized by subtracting an appropriate multiple of the byte transmission time, and only the minimum of these multiple measurements is embedded into the message. On the receiver side, multiple time measurements are taken and the minimum of those is used as the receiver time. This is very important, because serves to reduce the jitter in interrupt handling and the (de)coding and (de)modulation time. With few time stamps, improvement in precision can be obtained - in a Mica Mote platform - from the order of tens of microseconds to the order of about one microsecond.

Flooded messaging

To propagate the synchronization information, a flooding approach is used. An important fact, is that a single node in the network provides the global clock. The receiver gets a broadcast message, i.e. a reference synchronization point.

When a receiver has several reference points, then it becomes synchronized itself.

Nodes can collect reference points either from the global reference node. Reference point can be also from other nodes that are already synchronized. The frequency of the flooding provides a tradeoff between synchronization accuracy and overhead.

D. Predictive time synchronization

Predictive time synchronization is only reasonable for very short time intervals. In the real-life conditions, clock drift can vary over time quite a lot because of environmental temperature and humidity changes.

It means, that clock drifts must by continually reassessed.

An simple approach to this measurement is to resynchronize nodes periodically at the same interval.

It may cause unnecessarily high overhead in many cases.

This problem is addressed by the predictive synchronization mechanism [7], where the frequency of inter-node time sampling is adaptively adjusted.

Environments are characterized by a time constant *T* over which drift rates are highly correlated, which can be determined through a learning phase.

Assuming that the time sampling period is *S*, a window of T/S sample measurements is used in this technique not only to predict the clock drift (through linear regression), but also to estimate the error in the prediction.

A technique is used to adapt the sampling period: if the prediction error is above a desirable threshold, the sampling period S is reduced multiplicatively; and if it is below threshold, the sampling period is increased accordingly. Such adaptive scheme provides long-term synchronization in a self-configuring manner.

E. A wireless sensor network system for structural-response data acquisition

A wireless sensor network system for structural-response data acquisition [8] presents an interesting lightweight alternative to clock-synchronization approaches.

The approach is to collect and record latency measurements within each packet.

Only the base station is obligatory to have an accurate reference clock.

Since the radio propagation delays are insignificant, what is measured at each hop is actually the time that the packet spends at each node – which can be of the order of milliseconds due to queuing and processing delays.

Presented approach assumes that time stamps can be added as close to the packet transmission and reception as possible at the link layer. It means, that it is robust to many of the sources of latency uncertainty that cause errors in other synchronization techniques.

An important fact is that it is vulnerable to varying clock drifts at the intermediate nodes.

F. Time Synchronization Protocol for Wireless Sensor Networks using Clustering

Time synchronization means bringing all the sensor nodes to a common notion of time [10]. It is very important to know the order of events that has been sensed by the nodes for data fusion.

In presented algorithm, the root node (RN) initiates the synchronization process by multicasting the Syn_start packet to level-1 cluster heads (CHs), fig. 4.

The Syn_start packet contains packet sending time-stamp t1.

After receiving Syn_start packet by all the level-1 CHs, the CH with CH_ID=1 responds back by sending the Syn_ack packet which contain t1, t2 and t3, where t1 is RN send time, t2 is the CH packet receive time and t3 is CH packet send time.

Now RN calculates the propagation delay (d) and multicasts Syn_pkt that contains (d, t) the delay *d* and global time *t*.

When the level-1 CHs receive Syn_pkt with d and t then each CH will compute its drift and set the local clock according to the global time.

When all the level-1 CHs are synchronized, then these CHs become RN for level-2 CHs and the same process will be repeated in the next step.

The purpose of this algorithm is to set the logical clock of the CHs and cluster nodes with global time. The delay (d) and offset can be defined as

$d=[(t2-t1) + (t3-t4)]/2$

$Offset = t + d - LocalTime$

Where *t* is global time, *d* is the delay which will be constant for single hop communication, Offset is the time deviation of the two nodes (i.e. offset), LocalTime is CHs/nodes local time.

In fig.4 is presented message delay estimation process...

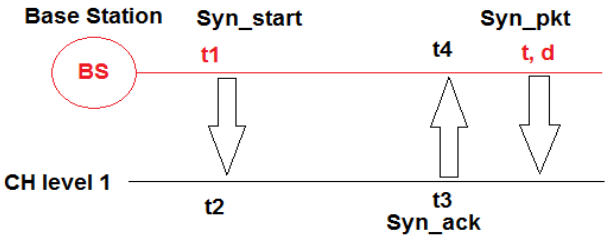


Fig. 4. Message Delay Estimation

… and in fig. 5 multilevel clustering scheme.

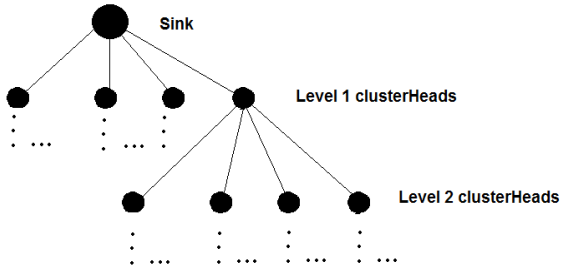


Fig. 5 Multilevel clustering

V. Packet delay model

Packet delay may be decomposed into the following six components.

A. Send Time

The time required by the sender to construct the packet and deliver to the MAC layer. Due to processor load it is non-deterministic.

B. Access Time

The waiting time for packets to get access to the wireless channel, also non-deterministic.

C. Transmission Time

The time that the sender takes to transmit the packet bit by bit at the physical layer, depends on the length of the packet and transmission baud rate.

D. Propagation Time

The time taken by the packet from the sender to receiver on a wireless link. It is deterministic and depends on the distance between sender and receiver.

E. Reception Time

The delay it takes for the receiver to receive the packet. It depends on the packet length and the transmission baud rate.

F. Receive Time

It is the time taken by the receiver to process the incoming packet

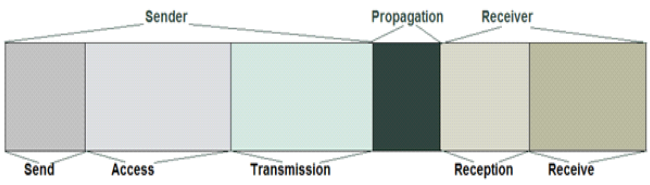


Fig. 6 Packet delay components

VI. Concept of time synchronization in wireless sensor network for health monitoring

The concept for time synchronization in wireless sensor network for health monitoring is presented. The focus is on efficient, reliable synchronization technique.

Network requirements, like traffic loads and latency may be different at different time. Increasing traffic loads cause unacceptable latency, because of data overload. These facts may be dangerous for patients, who e.g. are not yet fully diagnosed.

It is very important, that real-time transmission of life-critical data is guaranteed and latencies are minimum.

An interesting approach seems to be switching between normal state and emergency state, which means dynamically changing of data rate in order to minimize the latencies.

Novel synchronization scheme is proposed to decrease the overhead.

In fig. 7 all data from implantable sensors are within range of synchronization point, which has global clock. All data between implantable sensors go through synchronization point, which is also a kind of base station, covering all implantable sensors on the patient body. When implantable sensor wants to send data to external device, like doctor's computer, these data also go through synchronization point.
Suggested synchronization point is following: synchronization with global clock situated in synchronization point is done, when implantable sensors send its readings to synchronization point (which is also a base station collecting all data from sensors and forwarding data further). Synchronization process has to be simple, because implantable sensors have limited resources (limited battery), so it seems to have no way to enforce them to calculate the clock drifts and delays.
All calculations, needed for time synchronization are done on synchronization point side, i.e. the most suitable way seems to be a scenario, when implantable sensor sends its data to synchronization point together with its radio signal strength and timestamp when packet was sent on the sensor side. Synchronization point having these information, can calculate estimated delay between implantable sensor's clock and global clock located in synchronization point. This solution is probably not perfect, but is limiting energy usage on the implantable sensor's side, because only sensing data together with value of radio signal strength of such sensor with time-stamp of sent time on the sensor side are sent and nothing more. The estimated clock difference is sent back to implantable sensor. It is good to underline, that this process occurs only during first communication process between sensor and synchronization point and later only randomly,

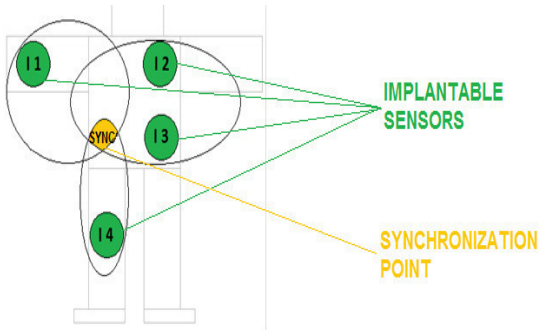


Fig. 7 Synchronization in health monitoring

not during each sending/receiving process. Synchronization time is important for health monitoring, because order of sensed data has to be the same as order of sent and received data between interested endpoints. This is crucial for patient's health.

VII. Conclusion

Synchronization belongs to core configuration problems in WSNs.

It is a fundamental service building block useful for many network functions, including time stamping of sensor measurements, coherent distributed signal processing, cooperative communication, medium-access, and sleep scheduling. Synchronization is necessitated by the random clock drifts that vary depending on hardware and environmental conditions.

Two approaches can be distinguished for time synchronization:

- Receiver-receiver synchronization technique of RBS
- Sender-receiver approach of TPSN

RBS has the advantage that multiple receivers can be synchronized with fewer messages.

It has been shown that these can provide synchronization of the order of tens of micro-seconds.

The flooding-time synchronization protocol improves performance by another order of magnitude by reducing uncertainties due to jitter in interrupt handling and coding/modulation.

An interesting direction for further research seems to be analyzing presented concept for time synchronization in wireless sensor networks in the scenario of use in health monitoring process. In such scenario are used implantable medical devices, which have to synchronize their clocks with one chosen point called synchronization point. There is presented an idea, where each sensor node sends together with sensed data – only during first communication process and later in a random mode - the radio signal strength and time-stamp of sent data on the sensor's side. The synchronization point can then estimate the delay, based on radio signal strength and time-stamp when packet was sent.

References

[1] E.Khandani, J. Abounadi, E. Modiano, L. Zheng "Cooperative Routing in Wireless Networks" Proceedings of Allerton Conference, March 2004

[2] A. Hu, S.D. Servetto "AsymptoticallyOptimal Time Synchronization in Dense Sensor Networks" Proceedings of ACM WSNA, September 2003

[3] F. Cristian "A Probabilistic Approach to Distributed Clock Synchronization" Distributed Computing, 3, 1989, 146-158

[4] D. L. Mills, "Internet Time Synchronization: The Network Time Protocol" IEEE Transactions on Communications, 39, 10, October 1991, 1482-1493

[5] S. Ganeriwal, R. Kumar, M. B. Srivastava „Timing-Sync Protocol for Sensor networks" Proceedings of ACM SenSys'03, November 2003

[6] M. Maroti, B. Kusky, G. Simon, A. Ledeczi, "The Flooding Time Synchronization Protocol", Proceedings of ACM SenSys, November 2004.

[7] S. Ganeriwal, D. Ganesan, M. Hansen, M. B. Srivastava, D. Estrin "Predictive Time Synchronization for Long-Lived Sensor Networks", University of Massachusetts, Amherst, Technical Report TR-04-97, 2004

[8] N. Xu, S. Rangwala, K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, D. Estrin "A Wireless Sensor Network for Structural Monitoring" Proceedings of ACM Conference on Embedded Networked Sensor Systems (SenSys), November 2004

[9] J. Elson, L. Girod, and D. Estring, "Fine-grained network time synchronization using reference broadcasts", Proceedings of Fifth Symposium on Operating Systems Design and Implementation (OSDI 2002), Vol. 36, pp. 147-163, 2002

[10] G. C. Gautam, T.P. Sharma, V. Katiyar, A. Kumar „Time Synchronization Protocol for Wireless Sensor Networks using Clustering", IEEE International Conference on Recent Trends in Information Technology ICRTIT 2011, pp. 417-422, 2011

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 89–94

Improving the Usability of Wireless Sensor Network Operating Systems

Atis Elsts*† and Leo Selavo*†
*Faculty of Computer Science
University of Latvia, Riga, Latvia
†Institute of Electronics and Computer Science
Riga, Latvia
Email:{atis.elsts,selavo}@edi.lv

***Abstract*—Wireless sensor network research community has constructed a number of operating systems that enable development of sensor network applications using novel and appropriate software abstractions. Unfortunately, the abstractions are not always easily usable by inexperienced users, because the learning curves of these existing operating systems are quite steep.**

In this article we present selected usability aspects of *MansOS* wireless sensor network operating system. We believe that MansOS has the potential to make sensor network programming accessible for a broader range of programmers. The evaluation of MansOS suggests improved usability for non-expert users over TinyOS and other operating systems without compromising efficiency.

I. Introduction

ISO standards [1] formally define usability as: "The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions" (ISO/IEC 9126-1, 2000). In this paper, usability is generally understood as continuous rather than binary property. The usability of a particular software artifact is inversely proportional to the cognitive effort required to use it.

The need for better WSN software usability has long been recognized in the research community. A wireless sensor network operating system should provide clean and flexible services that allow the developer to create highly-concurrent low duty-cycle applications naturally.

TinyOS [2] was the first to address this challenge. It is small, efficient and has been highly influential. Unfortunately, the system has gained notoriety as being difficult to learn [3]. Sensor network operating systems developed after TinyOS tried to increase usability and learnability in several ways: by allowing to write applications in plain C [4], using preemptive multithreading [5], or supporting familiar, UNIX-like abstractions [6].

We believe that while TinyOS did great by introducing many new abstractions suitable for WSN, more work could have been devoted making these abstractions accessible for a broad range of programmers. The ideas in TinyOS could have more real-world impact and be better understood if they were made available for plain C programmers, including those without WSN experience.

We have identified several usability benefits present in MansOS. It provides:

- Abstractions that are either already familiar or easy to grasp (Section III-C, Section III-D).
- Interchargeable components that can be enabled or disabled without changing application source code (Section III-B).
- Improved portability (Section III-A).
- Reduced source code size and prototyping time for simple sense-and-send applications (Section IV-A, Section IV-B).

The hard challenge was to implement usability while not giving up on WSN-specific design goals, in particular, efficiency.

II. Related Work

TinyOS is a seminal sensor network OS that has very high impact in the research community. A primary goal of TinyOS was to enable and accelerate WSN research [7].

TinyOS is written (and requires applications to be written) in a custom programming language: *nesC*. The need to learn a new language makes using the OS more complicated for non-expert users. Additionally, there are more factors that contribute to the steepness of TinyOS learning curve: novel software abstractions (for example, static virtualization, parametrized components), the fact that concurrency is fully exposed to the user, and high generality and modularity of the OS itself [3].

In contrast to many early attempts to apply TinyOS to real-world deployments, in more recent years TinyOS powered sensor networks often turn out to be a complete success. Nevertheless, finding new contributors to the core code is challenging [3]. The authors of TinyOS admit that over-generalization and too finely grained components add to the difficulty of modifying the OS itself. For example, the code for CC2420 radio driver is distributed across 40 source files, making it very hard to build a mental model of it.

Contiki is a lightweight operating system created at the Swedish Institute of Computer Science [4]. Contiki applications are written in plain C, support dynamic loading and unloading of components, and run on top of many popular WSN hardware platforms (including TelosB).

Contiki is recognized for its unique (in WSN context) execution model: on top of event-based kernel lightweight cooperative threading primitives are implemented, called *protothreads*

[8]. Although not without limitations, protothreads is a simple and elegant alternative both to event-driven application code and to preemptive multithreading. Most of stateful system services (networking etc.) are implemented as protothreads in Contiki.

MANTIS OS is a WSN OS created at University of Colorado at Boulder [5].

Multithreading is a key design feature of Mantis OS, added to increase the usability of the OS. Multithreading allows to naturally interleave processing-intensive tasks (such as data encryption or compression) with time-sensitive tasks (such as network communication). The authors stress that manual partitioning of complex tasks in smaller time slices is not trivial and sometimes require knowledge about the semantics of the algorithm.

Mantis is designed to work on multiple hardware platforms. In particular, application code can also be compiled to run natively on x86 architecture.

LiteOS [6] takes the ideas already seen in Contiki and Mantis to the logical extreme, making the OS as much UNIX-like as possible. Being easy-to-use is declared as one of LiteOS primary design goals, approached through features like (1) a distributed file system as an abstraction for the sensor network, (2) dynamic loading of separate "applications" (i.e. threads), (3) advanced event logging, dynamic memory support and other features. By reusing concepts from UNIX operating system, LiteOS tries to reduce the steepness of the learning curve, because the existing knowledge of the system programmer is leveraged. However, the suitability of these concepts to WSN is not always clearly shown.

LiteOS runs on a single hardware architecture (Atmel); the portability of the system is low, because inline assembly is frequently used.

Other operating systems and software libraries like **Arduino** [9] can be used to build sensor networks, but they are not WSN-specific and not well-optimized for the task.

III. MansOS Aspects and Abstractions

A. Code Organization

At the moment, MansOS supports ten hardware platforms, but only three hardware (MCU) architectures: Texas Instruments MSP430, Atmel AVR, and x86. Therefore, a distinction is made between platform-specific and architecture-specific code. This simple invention allows to greatly reduce the proportion of platform specific code.

MansOS code organization has both similarities and differences to TinyOS and Contiki. One remarkable difference is the size of *platform*-specific code, which is much lower in MansOS (because *architecture*-specific code is introduced). Architecture-specific code roughly corresponds to to `cpu` directory in Contiki, but the unification of shared code is not a complete in Contiki as it is in MansOS. For example, only in MansOS the periodic timer interrupt handler code (the "heartbeat" of the system) is unified and shared by all platforms.

For example, the official Contiki code for XM-1000 sensor device provided by AdvancticSYS [10] has 1920 lines of code (excluding BSL script, application examples, and file `checkpoint-arch.c` because MansOS has no equivalent functionality). Their TinyOS code has 1228 lines of code (excluding BSL script, MAC protocol, and CC2420X driver). MansOS, in turn, has only 629 lines of XM-1000 specific code. The rest of the code is reused from TelosB platform without copying it. Some of the new code (for MSP430 Series-2 MCU support) is actually reused by other platforms (e.g. Zolertia Z1). Excluding that code, there are only 251 lines, in other words, 7.6 times less than in Contiki, and 4.9 times less than in TinyOS. Clearly, this leads to faster portability, and, more importantly, better maintainability.

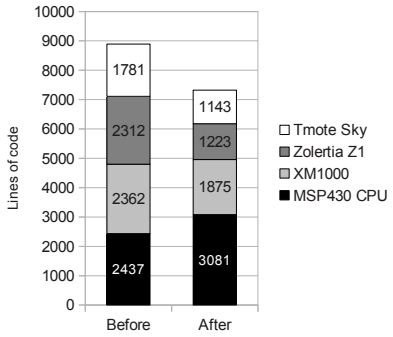


Fig. 1: The result of applying MansOS code organization to three Contiki hardware platforms and MSP430 MCU architecture

In order to show that the gains are present specifically because of MansOS code organization and not due to other unrelated reasons, we also modified the source code of Contiki itself. We selected just three hardware platforms (Tmote Sky, Zolertia Z1, and AdvancticSYS XM1000) that share a common MCU architecture: MSP430. We simply moved all files that correspond to `arch` code in MansOS from `platform` folders into `cpu/msp430` folder in Contiki. Creating a new `arch` folder in Contiki would have lead to better code organization, but was not neccessary to prove the point of this optimization. The result (excluding platform-specific example application code) is shown in Fig. 1. The Contiki source code size was reduced by 1570 lines of code, which is 17.7 % of code in directories of these three platforms (`platform/sky`, `platform/z1`, `platform/xm1000`) and `cpu/msp430` directory, in sum. Applying the same technique to more platforms with the same MCU architecture would lead to increased gains.

B. The Component Selection Mechanism

TinyOS allows the application developer to specify which components to use in application's configuration file. This leads to two benefits: firstly, optimal binary code size, as only components that are actually used are included in the final application; secondly, to extensibility and flexibility, as it now

becomes simple to integrate platform-specific or application-specific components in the OS.

In contrast, C-based operating systems like Contiki and Mantis have have limited means to achieve something similar. Contiki adopts "everything is included by default" strategy, which leads to large code size (Fig. 4). Mantis allows to select components with large granularity – high-level components are compiled as a separate libraries, which may or may not be linked against the application. Unfortunately, not only this becomes unreasonable if smaller component granularity is needed, but also tends to create unresolvable circular dependencies between the libraries.

The solution that MansOS adopts is to use an additional file next to application's source code: a *configuration file*. The configuration file allows to select or exclude components to use in the specific application. The resulting component granularity is larger than in TinyOS, but this approach is also useful for setting system-wide policies, e.g. the radio channel to use, the MAC protocol to use, its queue size and so on. These policies cannot be set in C source code in equally efficient way, because setting them either in a function call or by changing a global variable adds to run-time overhead.

MansOS build system ensures that each platform provides a reasonable set of components that are enabled by default. The dependencies between components are also resolved automatically. The novice user may successfully start using MansOS without being aware that the component selection mechanism exists. On the other hand, the experienced user may optimize or extend his applications by excluding and including components manually.

There are three benefits brought by such a mechanism:

- Modularity: components that implement the same interface can be used interchangeably, often without changing application C source code. For example, two versions of the kernel (event-based and thread-based) can be replaced with each other by changing just one line in configuration file.
- Extensibility: application-specific or platform-specific components can be easily added and used. For example, if a platform has a specific actuator and a driver of that actuator is added to MansOS, the actuator can be enabled for that platform by adding only a line in platform's configuration file. If the hardware actuator is also added for a specific application on another platform, it can be enabled by adding a single line in application's configuration file.
- Efficiency: unused components are not included in the binary image. This allows to achieve more efficient binary code and RAM usage, which in turn reduces prototyping and debugging time (Section IV-B).

The MansOS component selection mechanisms allows to achieve both function-granularity and file-granularity garbage collection at the link stage. The latter is needed to exluded components that are referenced by the kernel at initializion, but are not used neither by application code nor other components.

In order to show that MansOS approach effectively scales to other sensor network operating systems, we partially implemented the component selection algorithm for Contiki as well. The implementation consists of a patch that is applied to the development version of the Contiki source code repository, changes 37 source files and includes 272 code insertions and 41 deletions. In contrast to the project-specific configuration file option already present in Contiki, this new mechanism is much more comprehensive, generic, allows to exclude whole files from build, and is enabled by default.

C. Timing

When studying the source code state-of-art operating systems that run on TelosB platform, we made an interesting conclusion: none of the three systems analyzed (TinyOS, Mantis, and Contiki) provide high-accuracy millisecond-precision timers. Furthermore, out of the three only the latter provides high-accuracy long-term time accounting in SI units.

The cause of the problem lies in the fact that hardware timer ticks used for time accounting do not have high-accuracy mapping to milliseconds. The high-precision (20 ppm error), but low-frequency crystal has "binary", 32 768 Hz frequency.

Mantis is triggering the interrupt in *approximately* the needed time interval and adding the timer ticks passed since last interrupt to a system time counter. Unfortunately, the timing error accumulates with time.

TinyOS puts the responsibility for time correction on the user. The accounted milliseconds are called "binary milliseconds" in TinyOS; every second contains 1024 binary milliseconds. For novice users, this approach is confusing, as nothing in the name of `TMilli` suggests that it does not refer to the ordinary (SI time unit) milliseconds. The confusion is evidenced by the number of questions and help requests at TinyOS user mailing list[1].

The timing accuracy in the C code generated by TinyOS nesC compiler could be easily fixed: in fact, only one line must be changed, replacing a binary shift operation with multiplication. However, the TinyOS timers API [11] is so generic and multifaceted that introducing such a fix in it is much harder.

The Contiki solution is to not to provide millisecond timer abstraction at all, Instead, `CLOCK_SECOND` macro is defined for mapping from hardware timer ticks to SI units. Contiki also provides a highly accurate global time counter, but only with second granularity.

Timing accuracy is important in WSN, as we discovered ourselves in a precision agriculture deployment [12]. The users of the system wanted to record timestamped microclimate data from multiple locations. Although time-synchronization protocols for WSN exist (even with *micro*second precision, for example, [13]), and MansOS includes a simple version of such a protocol, it was of limited help in this situation: the network often become partitioned because some of intermediate nodes died. The motes we were using had neither a Real Time Clock chip nor a GNSS receiver, therefore software-only solution was required.

Millisecond abstraction improves usability, because is allows to use familiar rather than unfamiliar measurement unit.

[1] https://www.millennium.berkeley.edu/pipermail/tinyos-help/

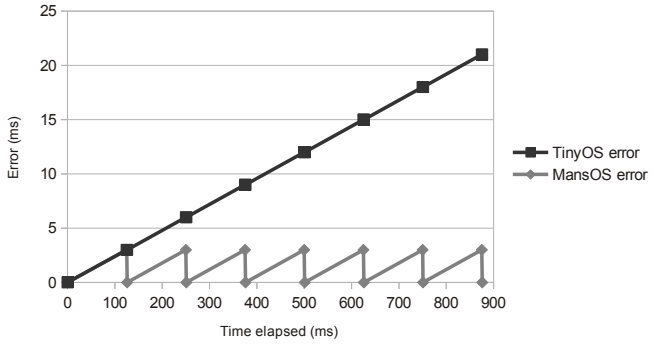


Fig. 2: Timing accuracy comparison

The idea adopted in MansOS is to use two hardware timers: a "counter" timer running with *approximately* 1000 Hz frequency, and a "correction" timer running with *precisely* 8 Hz frequency. Every time the counter timer fires, the global millisecond time counter is increased by one. Every time the "correction" fires, the counter is decreased by three. The accuracy error, which is cumulative in TinyOS, is kept bounded to 3 milliseconds in MansOS (Fig. 2).

Losing short-time accuracy is the trade-off of MansOS solution. 3 millisecond error is not noticeable by a human, but might make a difference when, for example, time slots for high-contention TDMA MAC protocol need to be allocated with high precision. In such a case, the user is better off by using hardware timers directly.

D. Memory allocation

Initially the memory for system-level objects (sockets, files, queued packets) in MansOS was allocated statically, using array variables at file-scope. The size of the arrays was either guessed by the system developer, or was determined in configuration files of applications. It was hard to extrapolate the correct size that a "typical" application is going to need, therefore the memory was often either wasted or allocated insufficiently.

A solution was discovered: make the object's user responsible for memory allocation for that object. Now all respective object initialization functions take a pointer to the object as an argument. This pattern is used for opening both files and sockets. Similarly, allocation of queued packets is now done at MAC protocol level. Each protocol driver can reasonably well estimate how large packet queue it is going to need. For most MAC protocols, a buffer of just one packet is sufficient.

All threads are also allocated statically, avoiding the potential reliability problems inherent in systems like Mantis, which allow dynamic memory allocation for new threads at run-time, and making the code more amenable to static analysis and checking.

Levis [3] praises parametrized components implemented in TinyOS as one of its most worthy features. However, we show that nesC is not required to implement an abstraction similar to parametrized hardware components. While in TinyOS the functions for separate hardware components are generated by the nesC compiler, in MansOS they are written by hand. To take a concrete example, consider an application that prints characters using USART #1 serial interface on a TelosB node. Such an application requires that the USART module is initialized in serial protocol mode. TinyOS generates code specifically for USART #1; since USART #0 is not used, the resulting application code is smaller than if a generic configuration function was used. MansOS achieves the same result by providing two different functions, one for each USART, and a high-level, inline wrapper function *serialInit()*. If GCC linker optimizations are enabled, the resulting code is as small as in TinyOS, because the USART #0 function code is optimized away.

IV. Evaluation

In this section, MansOS is experimentally evaluated and compared to TinyOS, Contiki, and Mantis (latest publicly available versions on 18th May, 2013). LiteOS is not included in the comparison as it lacks TelosB support.

Four test applications were analyzed: *Active* – busy looping application, *RadioRx* – radio packet reception, *RadioTx* – radio packet transmission, *Combined* – senses internal voltage and temperature, sends values to radio and writes them to the external flash memory[2].

A. Source Code Measurements

In this test, application source code length is compared. Lines of code are counted excluding comments and empty lines, but including configuration files. The source of nesC and Contiki applications were formatted by using TinyOS and Contiki example applications as basis, respectively, while C applications were formatted according to MansOS coding guidelines[3].

C code is approximately at the same level of abstraction as nesC code. Therefore, by comparing the size of the source code, approximate information about application complexity written in these languages can be deduced.

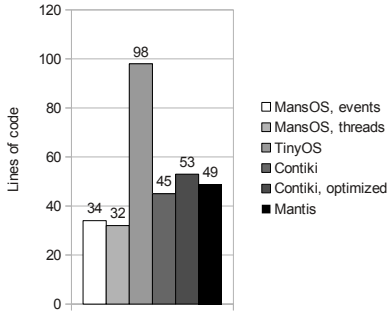


Fig. 3: Source code size comparison of the *combined* application

The results are provided in Fig. 3. Lower source code size in MansOS suggests better usability compared to other OS, especially to TinyOS. (The "optimized" version of Contiki required additional lines in application configuration file.)

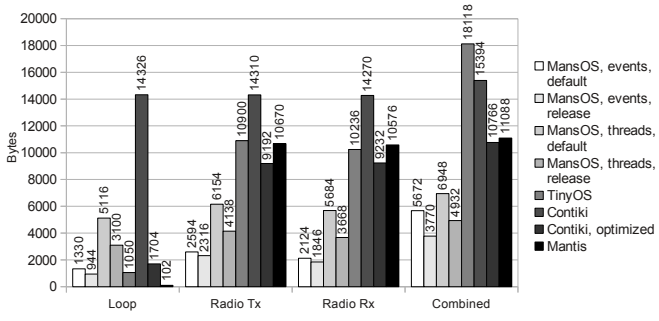
[2]Implementations available at http://mansos.edi.lv/dissert/testapps.tgz
[3]Available at http://mansos.edi.lv/wiki

B. Binary Code Measurements

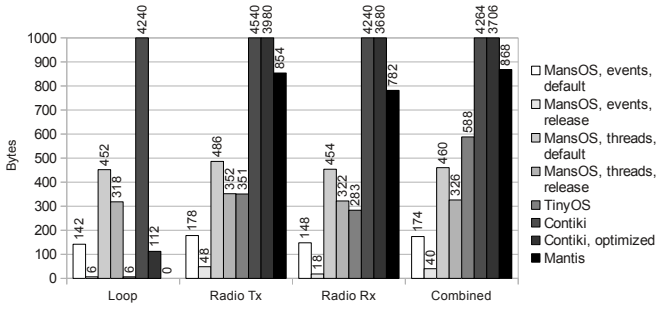
The code sizes were obtained by compiling with msp430-gcc version 4.5.3, using `-Os` optimization level. MansOS and Contiki builds also enabled linker optimizations, achieving garbage collection of unused functions.

Since WSN applications are often prototyped on more powerful devices than are used for the final deployment, we also consider a size-optimized ("release") versions of MansOS applications. To build such a release version, only one change was required: addition of `USE_PRINT=n` configuration option. As the motes in the deployed WSN usually cannot be monitored using serial interface, such a change is natural at the end of software prototyping, when it is prepared for the release. Certain parts of debugging code (e.g. the *ASSERT* macro) remained even in the "release" versions.

Two versions of each application were also developed for Contiki. The first set was used the default, limited component selection mechanism present in this operating system. `nullmac_driver`, `nullrdc_driver`, and `framer_nullmac` options were selected for all of the applications, and `CONTIKI_NO_NET` define was also selected in `Makefile` of the *loop* application. In order to show the improvements possible in this default configuration mechanism, a second, "optimized" set of application versions was developed on top of the optimized version of Contiki that partially incorporated the MansOS component selection mechanism.



(a) Flash memory usage



(b) RAM usage

Fig. 4: Application flash memory and RAM usage comparison

The results are given in Fig. 4 (Contiki RAM usage not shown in full). We note that the universality of the comparison is limited, as more advanced algorithms are used in some operating systems. For example, the external flash driver in TinyOS automatically adds CRC to data records, while in MansOS and other operating systems raw memory access takes place. However, the goal of this comparison is not as much to determine which OS can implement the same functionality in shorter code, but rather to determine the usability for out-of-the-box OS, including how big typical application prototypes are. Therefore the default coding style (platform-independent wherever possible) and build options are used.

There are various reasons why the other operating systems use more resources than MansOS. TinyOS uses resources for hardware component virtualization. Contiki includes most of functionality available on the platform by default. Mantis includes a lot of debugging code, for example, descriptive ASCII strings and code that operates LEDs. However, the core functionality of the system, for example the scheduler, also is implemented with larger overhead than in MansOS [14].

The optimized version of Contiki demonstrates good results for this set of applications: at least 30.0 % reduction in flash usage and at least 12.3 % reduction in RAM usage, showing that the MansOS component selection mechanism is a viable way how to reduce superfluous resource usage in other operating systems as well.

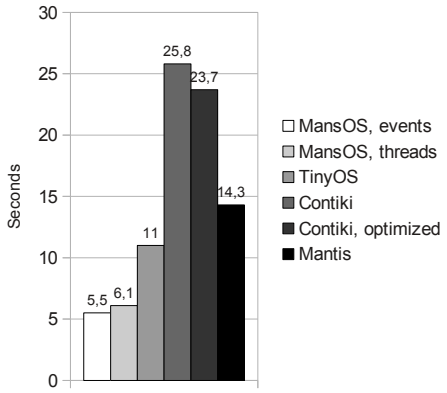


Fig. 5: Time of compilation and upload of the *combined* application

Binary code size (at least indirectly) has significant impact on developer productivity. If 20 second shorter binary code upload takes place, for a network of ten nodes the difference is 3 min 20 sec in total, but for 100 nodes: more than half an hour. Even though not many novice developers are going to work with 100-node networks, many of them are going to create and upload a lot of versions of their software while trying to debug a problem. Shorter code sizes make this process much faster (Fig. 5).

C. Execution Flow Analysis

When tracing the execution, one discovers that both operating systems do similar tasks:

- initialize the watchdog;
- calibrate the digitally-configurable oscillator;
- initialize hardware timers;
- initialize LEDs;
- put the external flash chip in deep sleep mode;

- initialize the list (MansOS) or array (TinyOS) of software timers;
- run the core function / scheduler / loop.

In MansOS event-based version, `appMain()` is used for user initialization only and for entering an infinite loop which calls `sleep()`. The real work is done by a timer callback function, called repeatedly by the system alarms processor.

In MansOS multithreaded version, array of two threads is initialized, and the user thread is started. (System thread is executed in the main execution context.) The user thread's start function is `appMain()`, which never returns, but performs all the work and calls `sleep()` in a loop.

In, TinyOS a task loop runs all tasks that are posted and ready to run. If no task is posted, the scheduler puts system in a low power mode. Each interrupt in turn causes the associated event handler to run.

The MansOS code usually is smaller not because it is missing a critical functionality, but because TinyOS offers much more options to the user (in ADC control and radio control most prominently for this example). Other that that, TinyOS also includes this extra functionality:

- code for 8 hardware timers by default, even though not all of them are used by the application;
- code for CSMA access of the CC2420 radio, while MansOS uses it at the PHY layer directly;
- code for Active Message creation and management, which MansOS send out data in an untyped C structure;
- code for resource arbitration (has to be done partly manually in MansOS).

The extra RAM usage in TinyOS is mostly because of arbitration code: more variables are required to keep in track the state of the system. Also, some components (such as CC2420 radio driver) use run-time variables where MansOS allows only compile-time changes (for example, whether to ACK automatically, whether to do address recognition, etc.). In several cases the state is stored in parallel to hardware, which also stores the same state (for example, for radio channel). There is also a buffer for radio packet reception, even though it is never required by the application logic.

From all of this, only manual calls to `sleep()` and manual resource arbitration (avoided in chip drivers that were implemented later) decrease usability. None of the problems make MansOS impractical to use.

V. Conclusions

We have presented aspects of MansOS and compared its usability with several existing wireless sensor network operating systems.

For non-expert users, MansOS improves usability over other C-based WSN OS by implementing semi-automatic high-level component selection mechanism that allows to easily extend, modularize, and optimize applications and the OS itself. Compared to the other WSN OS analyzed, it is the only one that offers *accurate* millisecond-precision software timers and time accounting on TelosB platform.

Some of the results (component selection mechanism and four-layer code architecture) have been adapted to Contiki with good results (more than 10 % improvements in resource usage and source code size, respectively).

MansOS adopts several features from TinyOS, for example, static memory allocation, extensive compile-time and link-time optimizations to remove unused code, and some aspects of parametrized hardware interfaces. It improves the usability of these features by implementing them in plain C.

Acknowledgements

We want to thank to all who have contributed to MansOS in these years since 2008. Girts Strazdins is thanked for his help in the user study, and Andrey Vihrov for valuable MansOS improvement ideas.

References

[1] A. Abran, A. Khelifi, W. Suryn, and A. Seffah, "Usability meanings and interpretations in ISO standards," *Software Quality Journal*, vol. 11, no. 4, pp. 325–338, 2003.

[2] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, "System architecture directions for networked sensors," *ACM Sigplan Notices*, vol. 35, no. 11, pp. 93–104, 2000.

[3] P. Levis, "Experiences from a Decade of TinyOS Development," in *Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI'12)*, 2012.

[4] A. Dunkels, B. Grönvall, and T. Voigt, "Contiki - a lightweight and flexible operating system for tiny networked sensors," in *Proceedings of the Annual IEEE Conference on Local Computer Networks*. Los Alamitos, CA, USA: IEEE Computer Society, 2004, pp. 455–462.

[5] S. Bhatti, J. Carlson, H. Dai, J. Deng, J. Rose, A. Sheth, B. Shucker, C. Gruenwald, A. Torgerson, and R. Han, "MANTIS OS: An Embedded Multithreaded Operating System for Wireless Micro Sensor Platforms," *Mobile Networks and Applications*, vol. 10, no. 4, pp. 563–579, 2005.

[6] Q. Cao, T. Abdelzaher, J. Stankovic, and T. He, "The LiteOS Operating System: Towards Unix-Like Abstractions for Wireless Sensor Networks," in *IPSN '08: Proceedings of the 7th international conference on Information processing in sensor networks*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 233–244.

[7] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, "Tinyos: An operating system for sensor networks," in *Ambient Intelligence*. Springer Verlag, 2004.

[8] A. Dunkels, O. Schmidt, T. Voigt, and M. Ali, "Protothreads: simplifying event-driven programming of memory-constrained embedded systems," in *Proceedings of the 4th international conference on Embedded networked sensor systems*, ser. SenSys '06. New York, NY, USA: ACM, 2006, pp. 29–42.

[9] J. Brock, R. Bruce, and S. Reiser, "Using Arduino for introductory programming courses," *Journal of Computing Sciences in Colleges*, vol. 25, no. 2, pp. 129–130, 2009.

[10] AdvanticSYS, "XM1000," http://www.advanticsys.com/wiki/index.php?title=XM1000, 2013, [Online: accessed 26.04.2013.].

[11] C. Sharp, M. Turon, and D. Gay, "TinyOS TEP 102: Timers," http://www.tinyos.net/tinyos-2.x/doc/html/tep102.html, 2007, [Online: accessed 10.02.2013.].

[12] A. Elsts, R. Balass, J. Judvaitis, and L. Selavo, "SAD: Wireless Sensor Network System for Microclimate Monitoring in Precision Agriculture," in *Proceedings of the 5-th international scientific conference Applied information and communication technologies (AICT 2012)*, 2012, pp. 271–281.

[13] M. Maróti, B. Kusy, G. Simon, and Á. Lédeczi, "The flooding time synchronization protocol," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*. ACM, 2004, pp. 39–49.

[14] A. Elsts, G. Strazdins, A. Vihrov, and L. Selavo, "Design and Implementation of MansOS: a Wireless Sensor Network Operating System," *Computer Science and Information Technologies, Scientific Papers University of Latvia*, vol. 787, pp. 79–105, 2012.

Information Technology for Management, Business & Society

IT4MBS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the disciplines of information technology and information systems. The IT4BMS area emphasizes the issues relevant to information technology and necessary for practical, everyday needs of business, other organizations and society at large. This area takes a sociotechnical view on information systems and relates also to ethical, social and political issues raised by information systems.

Events that constitute IT4BMS are:

- **ABICT'13 –** 4th International Workshop on Advances in Business ICT
- **Agent Day'13**
- **AITM'13 –** 11th Conference on Advanced Information Technologies for Management
- **IT4L'13** – 2nd Workshop on Information Technologies for Logistics
- **KAM'13 –** 19th Conference on Knowledge Acquisition and Management
- **MMT'13 –** 3rd Workshop on Modeling Multi-commodity Trade: Towards Smart Systems
- **TAMoCo'13** – Techniques and Applications for Mobile Commerce

11th Conference on Advanced Information Technologies for Management

We are pleased to invite you to participate in the 10th edition of Conference on "Advanced Information Technologies for Management AITM'2012". The main purpose of the conference is to provide a forum for researchers and practitioners to present and discuss the current issues of IT in business applications. There will be also the opportunity to demonstrate by the software houses and firms their solutions as well as achievements in management information systems.

Topics

The topics of interest include but are not limited to:

- Concepts and methods of business informatics
- Business Process Management and Management Systems (BPM and BPMS)
- Management Information Systems (MIS)
- Enterprise information systems (ERP, CRM, SCM, etc.)
- Business Intelligence methods and tools
- Strategies and methodologies of IT implementation
- IT projects & IT projects management
- IT governance, efficiency and effectiveness
- Decision Support Systems and data mining
- Intelligence and mobile IT
- Cloud computing, SOA, Web services
- Agent-based systems
- Business-oriented ontologies, topic maps
- Knowledge-based and intelligent systems in management

Event Chairs

Dudycz, Helena, Wrocław University of Economics, Poland

Dyczkowski, Mirosław, Wrocław University of Economics, Poland

Korczak, Jerzy, Wroclaw University of Economics, Poland

Program Committee

Abramowicz, Witold, Poznan University of Economics, Poland

Alt, Rainer, University of Leipzig, Germany

Andres, Frederic, National Institute of Informatics, Tokyo, Japan

Chmielarz, Witold, University of Warsaw, Poland

Cypryjanski, Jacek, University of Szczecin, Poland

Czarnacka-Chrobot, Beata, Warsaw School of Economics, Poland

Damiani, Ernesto, Università degli Studi di Milano, Italy

De, Suparna, University of Surrey, Guildford, United Kingdom

Dufourd, Jean-François, University of Strasbourg, France

Kannan, Rajkumar, Bishop Heber College (Autonomous), Trichy, India

Kersten, Gregory, Concordia University, Montreal, Canada

Kowalczyk, Ryszard, Swinburne University of Technology, Melbourne, Victoria, Australia

Ligęza, Antoni, AGH University of Science and Technology, Poland

Magoni, Damien, University of Bordeaux – LaBRI, France

Owoc, Mieczyslaw, Wroclaw University of Economics, Poland

Pankowska, Malgorzata, University of Economics in Katowice, Poland

Sikorski, Marcin, Gdansk University of Technology, Poland

Stanek, Stanislaw, General Tadeusz Kosciuszko Military Academy of Land Forces in Wroclaw, Poland

Teufel, Stephanie, University of Fribourg, Switzerland

Ziemba, Ewa, University of Economics in Katowice, Poland

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 99–104

An examination of students' attitudes and preparedness for the introduction of ICT-enabled learning at university

Dr inż. Colin F. Hales
Uniwersytet Rzeszowski
Wydział Ekonomii
Zakład Metod Ilościowych
ul. Ćwiklińskiej 2, Rzeszów, Polska
E-mail: colhales@univ.rzeszow.pl

Dr Barbara Fura
Uniwersytet Rzeszowski
Wydział Ekonomii
Zakład Metod Ilościowych
ul. Ćwiklińskiej 2, Rzeszów, Polska
Email: bfura@univ.rzeszow.pl

***Abstract*—The article presents research results relating to the attitudes of students' towards a wider use of ICT in the education process (e-learning/blended learning). The analyses were conducted against the backdrop of the students' subjective assessment of their general computer aptitude as well as their access to and use of ICT. Results prove that students' opinions on the wider use of ICT in the learning process was influenced by a subjective assessment of their general computer skills. That level was in turn determined by factors including study specialization as well as socio-demographic characteristics of the respondents, such as gender, place of residence and age. Access to a computer/Internet had some influence upon familiarity with given applications and online activities. Research results confirm the presumption that students with a higher appraisal of their competence were more positively inclined towards a wider use of ICT than those with lower levels of competence.**

I. Introduction

THE traditional context of learning is experiencing a radical change. Teaching and learning are no longer restricted to traditional classrooms [1]. Teaching with the use of computers and computer applications is becoming ever more popular. Electronic learning (e-learning), referring to the use of electronic devices for learning, including the delivery of content via electronic media such as Internet, audio or video, satellite broadcast, interactive TV, CD-ROM, and so on, has become one of the most significant developments in the information systems industry [2]. The rapid expansion of the Internet as a delivery platform, combined with the trends towards location-independent education and individualization, has motivated universities to invest their resources on developing online programs [3].

Blended learning on the other hand, refers to an integrated environment, which combines the advantages of e-learning and traditional classroom teaching [4]. Generally, blended learning means any combination of learning delivery methods, including most often face-to-face instruction with asynchronous and/or synchronous computer technologies [5]. Blended learning – in its simplest form, the thoughtful integration of online and face-to-face-instruction [6] is being used with increased frequency around the world [7]. Thousands of corporate training programs, institutions of higher education schools participate in blended learning [8].

Although *e-learning* environments are popular, there is minimal research on instructors' and learners' attitudes toward these kinds of learning environments [2]. Usefulness and ease of use have proved to be key determinants of the acceptance and usage of *e-learning*. On the contrary, little is known about students' perceptions in a blended learning setting [9]. During the recent years, a significant volume of research on the effective use and integration of Information and Communication Technologies (ICT) in education practices is observed. The main feature that differentiates the *e-learning* systems from the traditional learning environments is the degree of technology usage and the gradual shift of control and responsibility of the learning process to the learners, giving them the opportunity to learn anytime, anywhere. This shift of control seems to positively influence the learning effectiveness of learners [10]. Other findings [11] suggest that computer self-efficacy, among others, significantly affect performance expectations. Interaction has a significant effect on learning climate.

Many researchers have tried to evaluate the role of technology in the learning process with much emphasis being placed on the components of the learning systems [12]. Davis and Wong [13] pointed out that important factors having paramount influence on participation and engagement with a learning structure could be associated with their acceptance and affective responses towards the system. Others have used a reasoning quite similar to the technology acceptance model approach while investigating this phenomenon. Shu-Sheng Liaw et al. states that behavioral intention to use e-learning is influenced by perceived usefulness and self-efficacy [14] and further found that trainers were more positively inclined to use *e-learning* systems.

II. Data and Methods

The main objective of the study was an investigation into the attitudes of students towards a wider use of Information and Communication Technologies (ICT) in the process of delivering education (e-learning and blended learning) at a higher education institution. In accordance with the belief that the attitude of a student towards ICT enabled learning, in some way is influenced by their general aptitude in the area of using computer tools, the results were presented on the backdrop of their subjective judgment of their general computer aptitude. In addition observations were made into the possible influence/role that Internet access that students possess (and to computers as such) on the student's approach towards the integration of ICT's in their education process.

This article presents the results of surveys conducted among students of selected faculties of the University of Rzeszow. The studies were carried out in the 2011/12 academic year, among a random sample of 420 respondents.

Opinions of the respondents were presented against the background of socio-demographic variables, such as age, place of residence (size of population) and gender. Since in the researchers' opinion, study profile could also have an impact on the attitude of students to the matter under consideration, we decided to profile students according to the study specialization. These specializations were later grouped according to their common characteristics (by faculties). Empirical data was collected through direct surveys. The sample of students taken into consideration for this study was obtained through cluster sampling.

The following research hypotheses were tested:

- The level of general computer competence as well as the familiarity and ability to use certain computer applications by the students was influenced by a computer/Internet access
- The attitude of students towards the wider use of ICT was dependent upon the level of competence in using a computer
- The general level of computer skills was determined by factors such as, study specialization as well as socio-demographic characteristics of the respondents.

I.In order to verify those statements the Spearman's rank correlation analysis and the chi-square test of independence were applied. The strength of the relationships demonstrated with the use of test of independence was measured using Cramer's V coefficient of association. The Student's t-Test was employed to determine the incidence of statistical differences in the self-accessed competence levels of women and men.

III. Characteristics of the Research Sample

The sample population totaled 420 young people of whom 65.5% (275) were females while 34.5% (145) were males. 33.3% of the respondents were rural inhabitants while 66.7% resided in urban areas. A significant percentage of the respondents were third year (38.%) and second year (31.7%) students. Students of the first, fourth and fifth years constituted the remaining 30% of the study sample. The respondents were students in the fields of Administration, English Studies, Landscape Architecture, Polish Studies, Physiotherapy, Mathematics, Environmental Protection, Agriculture, Tourism and Recreation, Physical Education as well as Food Technology. The Administration and English Language Studies specializations each accounted for 14.3% i.e. per 60 students, while the remaining fields of studies each represented 7.1% of the sample i.e. per 30 students. More than half (55.2%) of the respondents were aged 21-22, while the over 23 and 19-20 age groups accounted for 29% and 15.7% respectively.

IV. Evaluating respondents' ICT usage

Assessing the use of ICT in learning and communication was undertaken through the identification of major sources of information, the main tools of communications used by responding students as well as place of Internet (computer) access. The most commonly used source of information by the respondents were the Internet (76%), television (11.7%), and the radio (7.9%). A meager 2.4% of respondents identified the print media as the most frequently used source of information. The results point to the dominant role of the Internet in sourcing for information by the respondents. This serves as a premise in favor of using this medium for knowledge acquisition by students.

Besides serving as a prominent source of information for the respondents, the Internet is also a means of personal communication. 6.7% of respondents use the electronic mail as the most common medium of communication. It is the third most frequently used communication tool after the mobile phone (72.4%) and personal contact (18.3%).

Other types of communication tools such as landline phones, traditional mail and others constituted a negligible proportion (2.6%) of responses. Over three-quarters of respondents accessed the Internet from their homes, while 9.3% at place of work and only 4.8% at the university. About 7.4% of respondents accessed the Internet from other places. Worthy of note, is the low number of students citing the university environment as a place where they can/do access the Internet. This might partly be due to the fact that at the time of the study, the wireless access system was under development and computer labs were not easily accessible to students for out of class activities.

V. assessment of Computer Skills

The application of modern information and communications technologies in the learning process requires adequate levels of computer skills from both students and teachers. In order to establish the students' potential in the event of implementing online learning an evaluation of their subjectively perceived levels of computer skills were carried out using a measurement ranking scale of 1-10, where 1 indicates very weak level and 10 indicates very strong competence. Fig. 1 shows the distribution of responses regarding students' self-evaluation of their computer skills.

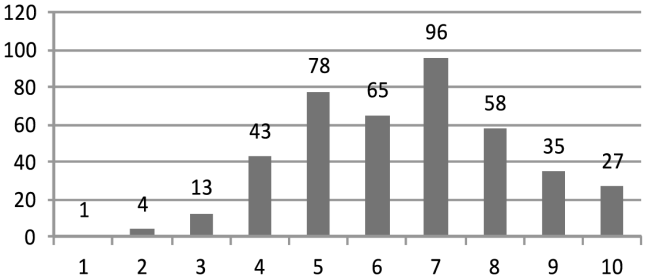


Fig.1 Overall rating of students computer skills.

The most often recorded value of self-assessment was at the level of 7 which was declared by almost 23% of the re-

spondents. More than 50% of the sample judged their competence at a level of 7 and above which gives a clear indication that the average level of competence among the students surveyed was quite satisfactory. The general computer level of the students was examined with respect to the students' gender (Table 1).

TABLE I.
DIFFERENCES IN THE LEVEL OF COMPUTER COMPETENCE BETWEEN MEN AND WOMEN - T-STUDENT TEST RESULTS

Var.	N (M)	N (W)	Avg. (M)	Avg. (W)	t	df	p
Skill Level	145	275	6.99	6.20	4.19	418	0.00

Through the use of t-Student's test it was proven that the average general competence level between men and women differed (0,00<0,05). The average competence level among males was higher than that for females. Many researches have tried to examine the relationship between the gender and the attitudes towards computers and their research have yielded varied results. Abouserie and Moss [15] or Kesici et al. [16] could not find a significant relationship between college students' attitudes towards computer-assisted learning and their gender while Whitley did record gender differences in computer-related attitudes and behavior [17]. The research results herein determine that there is significant difference between college students' computer attitudes based on gender in parallel with the findings from other studies which found a relationship between gender and computer use [18], [19]. A comprehensive review of the literature of gender differences in computer-related behavior reveals a myriad of conflicting results [20].

Besides, the subjective assessment of the respondents' overall level of computer competence, an analysis was made of the level of competence with respect to each of the following applications, i.e., Word, Excel, Power Point, E-mail software as well as the use of search engines fig. 2.

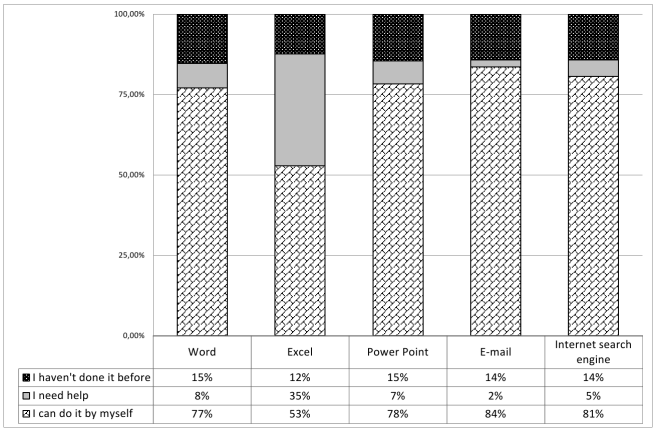


Fig. 2. Overall rating of students computer skills

The highest level of competence was observed in their ability to send e-mails with attachments . More than 83% respondents confirmed their ability to use electronic mails. A similar proportion of respondents (80.7%) confirmed they can independently use Internet search engines. This fact indicates that the responding students are competent in using the Internet to seek information and also to communicate. From amongst the programs that constitute the Microsoft Office package, the highest skills level was observed in respect of Power Point. The level of independent use of Word was a little less while the lowest was with using Excel.

In order to ascertain the degree of use of modern technologies by students, the frequency of use of chosen tasks or programs was evaluated. Among the tasks and programs listed, students most often made use of theme forums, and e-mails in contacts with their teachers and lecturers. They also gained experience having been co-participants in video-conferencing. E-learning platforms and participation in online courses were, however, less popular. More than 47% of respondents never had work-related experience in e-learning environments, while over 12% had never heard of it. Almost 47% of respondents had never participated in Internet run courses and 6.2% had never heard of such. The data presented point to the relatively low popularity of e-learning tools in the learning-teaching process. This not only poses potential obstacles to the introduction of aspects of e-learning but also indicates tools that needs be popularized prior to the introduction of online teaching.

VI. LEVEL OF COMPUTER SKILLS AND AVAILABILITY OF INTERNET/COMPUTERS

In order to verify the hypothesis that overall level of computer competence is statistically significantly correlated with Internet (computer) availability the Spearman rank correlation analysis (with both variables presented as ranked scales) was applied. Results of this analysis confirmed that better Internet (computer) availability was associated with higher level of overall computer competence. This conclusion was reached based on the positively ranked correlation coefficient (0.08). However, the lack of statistical significance of the indicated correlation (0.10>0.05) makes its generalization beyond the tested sample impossible. That fact that the overall level of competence of students of Rzeszow University was associated with Internet (computer) accessibility could not be demonstrated. While there weren't grounds to reject the hypothesis of equivalence of the correlation coefficient being zero, it is the same correlation analysis that enables the identification of a statistically significant relationship between the level of skills with selected programs/tasks and the availability of the Internet (computer).

TABLE II
SPEARMAN CORRELATION FOR VARIABLES: PLACE OF INTERNET ACCESS AND LEVEL OF COMPUTER SKILLS WITH SELECTED PROGRAMS

Variables: place of Internet (comp.) access & skills level with programs	Spearman's rho	p
Word	0.2182*	0.000
Excel	0.0988*	0.043
Power Point	0.1996*	0.000
E-mail	0.2381*	0.000
Internet search engines	0.2721*	0.000
Word	0.2182*	0.000

* at level of 0.05.

Results translate to mean that as Internet/computer availability increased so did the level of skills in the use of the programs/tasks indicated. The strongest dependency was noted between Internet/computer availability and the use of Internet search engines (0.27). Computer availability, on the other hand, translates into competence in use of e-mail, Word, Power Point programs but with least impact on knowledge of Excel program. These results may suggest that students continue to use Internet/computer more for entertainment than for learning. The Spearman correlation analysis was also applied to test dependencies between frequency of use of selected programs/tasks and Internet/computer availability. The results are presented in table 3.

TABLE III.
SPEARMAN RANKED CORRELATION FOR VARIABLES: PLACE OF INTERNET ACCESS AND FREQUENCY OF USE OF SELECTED PROGRAMS

Variables: place of Internet (computer) access & frequency of use of selected programs	Spearman's rho	p
Internet-based course	0.1720*	0.0004
Discussions on theme forum	0.2923*	0.0000
Video-conference	0.2047*	0.0000
e-Learning platform	0.1442*	0.0031
Contact with teachers via e-mail	0.1051*	0.0314

* at 0.05 level

All the dependencies were positive and statistically significant. Internet/computer availability translated primarily into university students participating in discussions conducted on Internet forums, video-conferencing, Internet-based courses, working in e-learning environment as well as contact with tutors/lecturers via electronic mail.

VII. STUDENTS' OPINION ON THE USE OF ICT IN THE TEACHING PROCESS

One of the objectives of the study was to assess the attitude of students at the University, towards teaching achieved through the use of modern technologies. The implementation of such teaching could therefore be justified by the fact, that students perceive the benefits accruing to use of modern technology. While almost 54% of students wholly agree with the statement that ICT offers good access to necessary study materials, 41% admit that the Internet facilitates sharing of information and observations with people of similar interests, another 34.8% respondents concluded that audio-visual materials can improve the teaching process. While a wider use of ICT is supported by the fact that more than 20% of the respondents rejected the statement that e-learning is time-consuming, more than 18% claim that it is difficult finding specific information on websites and yet another over 17% of them claim that learning with the use of ICT has failed to meet expectations.

There were, amongst students, opinions which favored traditional teaching to teaching with the use of ICT. Many students still prefer to read printed texts – an opinion which over 44% of respondents agreed with and another 30% more or less are in support of. Almost 49% of respondents prefer learning using traditional methods while for almost 46% finding specific information on website constitutes a difficulty. Half of the respondents have opinions that the implementation of such teaching forms requires the acquisition of upgraded computer skills. In addition, about 34% of respondents are of the opinion that teaching via the Internet is not as effective as teaching using traditional methods.

There were, among the positive correlations, statistically significant relationships between: the level of computer skills and the willingness to ask teachers/lecturers questions via ICT as well as between the first variable and the assessment of possible improvements to the learning process using ICT. It was, on this basis, concluded that the higher the level of the University students' computer skills, not only the greater their willingness to contact lecturers using modern tools but also the greater their conviction of the existence of potential possibilities for improvements using ICT.

Based on the negative, statistically significant correlation between levels of computer competence and the variable relating to preferences in reading printed texts as well as between the first and the opinion about difficulties in finding specific information on websites, it was concluded that the said difficulties were observed more by the University students, who rated their level of computer competence as weak.

VIII. OVERALL ASSESSMENT OF COMPUTER COMPETENCE AND CHARACTERISTICS OF RESPONDENTS

Another hypothesis that was verified was of the existence of relationship between overall level of computer competence and the characteristic features of respondents. The characteristic features tested were gender, place of abode, field of study and age of the respondents. The level of competence, initially measured as a ten-point scale was presented as three-point scale distinguishing three levels of computer competence i.e., very weak/weak, average, and good/very good. The procedure made it possible to examine the existence of relationships between the variables by using the Pearson's chi-square test of independence.

Since the level of probability test p turned out less than the assumed level of significance (0.05), the hypothesis of independence of variables was rejected. It was thus shown that the level was determined separately for each of the listed characteristic features of the respondents. The strongest correlation coefficient measured using the Cramer's V coefficient of association was noted between level of competence and field of study of respondents (0.24). Other dependencies displayed similar strength.

IX. ATTITUDES OF STUDENTS TOWARDS THE INTRODUCTION OF E-LEARNING AND THE USE OF COMPUTERS IN TEACHING

The last stage of the analysis was to test the attitudes of students towards the introduction of online teaching into the University's curriculum.

Over three-quarters of the respondents identified the opportunities offered by the implementation of e-learning.

Students confirmed that this form of teaching makes education acquisition more available for all including disabled persons or those from regions distant from the University. More than 60% of respondents saw the opportunity of taking courses within and outside the country through teaching via the Internet and conceded that such form of knowledge acquisition is attractive since it allows easier contact with teachers and thus making the learning process convenient due to the large availability of varied sources of information. The respondents, on the other hand, doubted if this form of teaching offers real opportunities for speedy response to teaching effects and if it actually ensures more effective and systematic teaching curriculum.

The introduction of elements of e-learning inherently involves the use of computers in the teaching-learning process. The tendency of students to use the computer in the learning process is illustrated in fig. 3.

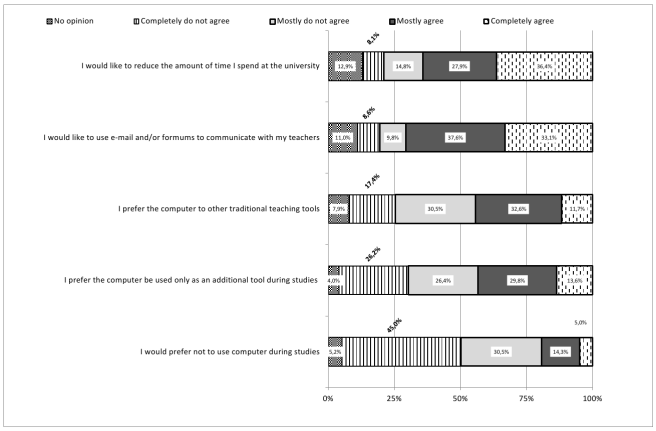


Fig. 3. Attitude of students towards the use of computer in teaching-learning process

More than 44% of respondents expressed their willingness to replace traditional teaching tools with computer aided solutions. Furthermore, over 26% of respondents disagreed with the statement that the computer is only an additional element in their course of study, thus to a certain extent, emphasizing the importance of this tool for students in their studies. Simultaneously, as many as 45% of them would not like to be deprived of the possibilities of carrying out part of their studies with the use of the computer. Students appear to acknowledge the time saving opportunity and other benefits offered through the introduction of Internet-based teaching and also confirm that this approach would facilitate easier contact with their teacher.

The obtained results concur with the literature [21] providing empirical evidence supporting the usefulness of e-mail as a promising aid to promote student cognitive growth pertaining to computer knowledge and skills. However, as Lee [22] succinctly warned, "Richness or leanness is not an inherent property of the electronic mail medium, but an emergent property of the interaction of the electronic mail with its organizational context." For e-mail to make a beneficial impact on education he states that to cultivate an online two-way communication, instant feedback provided by recipients would be an important task all participants need to attend to. However Ruberg and Miller [23] pointed out that access to essential hardware and software need to be in place to ensure online communication among participants.

Among the hypotheses which were tested was the belief that students with better computer competence levels would be more inclined to embrace e-learning as an important aspect of their curriculum activities. This hypothesis was tested using the rank correlation analysis between the level of competency and students' opinions on the use of ICT in the education process. These results are presented in Table 4.

On the basis of the probability of the test values and the sign (negative) of coefficient of correlation it can be stated that there was a significant relationship between the judgment of computer competence and:

- the propensity to use a computer during studies (-0.098), as well as,
- the inclination to use a computer only as an additional tool in the learning process (-0.160).

Significantly disparate positive correlations were discovered between the competency level and:

- the tendency to rank computer assisted tuition above traditional methods (0.130) and
- the readiness to use e-mail to communicate with instructors (0.084).

The results herein presented allow for the legitimate assumption of validity of the aforementioned hypotheses. Students declaring lower levels of competence were less inclined to use digital tools in the course of study as well as they seemed to underrate the value of ICT in their study activities.

TABLE IV.
SPEARMAN'S RANK CORRELATION FOR VARIABLES: ATTITUDE TOWARDS THE USE OF COMPUTER IN LEARNING PROCESS AND OVERALL LEVEL OF COMPUTER SKILLS

Variables: overall level of computer skills& opinion on using computer in teaching-learning process	**Spearman's rho**	**p**
I would not to use computer during studies	-0.0976*	0.045
I prefer the computer be used only as an additional tool during studies	-0.1597*	0.001
I prefer the computer to other traditional teaching tools	0.1301*	0.007
I would like to use e-mail and/or forums to communicate with my teachers	0.0835**	0.087
I would like to reduce the amount of time I spend at the university	-0.0452	0.355

* at 0.05 level; ** at 0.1 level.

X. CONCLUSIONS

For the vast majority of students the most prominent source of information appears to be the Internet while they predominantly use mobile phones for personal communication. The survey indicates that the majority of university students accessed the Internet from their homes. This gives the foundation to assume that a move towards the wider integration of information and communication technologies into the

curriculum should be a most natural succession for such a body of students Consequently there's reason to assume that it is highly probable that the introduction of an e-learning/blended learning solution may yield favorable results. Furthermore most of the students declared manageable proficiency levels in their use of ICT thus further supporting the belief that they would be well equipped to employ such technologies in the course of study.

This corresponds with opinions from the literature [24] which reports that learner's attitudes towards technology and blended learning may have significant effects on the success of the ICT application in higher education.

Research results did not allow for the assumption that access to a computer influenced the level of students' competency but did show a positive effect on their familiarity and knowledge of certain computer applications like Word, Excel, PowerPoint. Students in possession of computers also seemed to be more acquainted with tasks like using search-engines as well as e-mail programs. A positive relationship was also noted between those who had access to a computer and their knowledge and use of e-learning and blended learning tools i.e. online courses, forums and e-mail.

Students' general competency was influenced by their study specializations, as well as gender, age and place of abode. Students were more willing to incorporate ICT into the repository of learning tools as subjective assessment of their computer competence improved.

REFERENCES

[1] Y. S. Wang, H. Y. Wang, and D. Y. Shee, "Measuring e-learning systems success in an organizational context: Scale development and validation," Computers in Human Behavior, vol. 23(1), pp. 1792–1808, 2007.

[2] S. S. Liaw, H. M. Huang, and & G.D. Chen, "Surveying instructor and learner attitudes toward e-learning," Computers & Education, vol. 49 (4), pp. 1285–1296, 2007.

[3] S. Ozkan, and R. Koseler, "Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation," Computers & Education, no. 53, p. 1285, 2009.

[4] Handbook of blended learning: Global perspectives, local design, C. J. Bonk, and C. R. Graham, Ed., San Francisco: Jossey-Bass/Pfeiffer, 2005.

[5] H. J. So, T.A. Brush, "Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors," *Computers & Education*, no. 51, p. 321, 2008.

[6] D. R. Garrison, and H. Kanuka, "Blended Learning: Uncovering its Transformative Potential in Higher Education," *The Internet and Higher Education*, vol. 7(2), pp. 95–105, 2004.

[7] J. S. Drysdale, C. R. Graham, L. R. Halverson, and K. J. Spring, "An analysis of research trends in dissertations and theses studying blended learning," *Internet and Higher Education*, vol. 17, 90, 2013.

[8] A. G. Picciano, J. Seaman, P. Shea, and K. Swan, "Examining the extent and nature of online learning in American K-12 Education: The research initiatives of the Alfred P. Sloan foundation," *Internet and Higher Education*, vol. 15(2), pp. 127–135, 2012.

[9] N. Tselios, S. Daskalakis, and M. Papadopoulou, "Assessing the Acceptance of a Blended Learning University Course," *Educational Technology & Society*, vol. 14 (2), pp. 224–235, 2011.

[10] S. W. Chou, and C. H. Liu, "Learning effectiveness in a Web-based virtual learning environment: a learner control perspective," *Journal of Computer Assisted Learning*, no. 21(1), pp. 65–76, 2005.

[11] H. Wu, R. D. Tennyson, and T. L. Hsia, "A Study of Student Satisfaction in a Blended E-Learning System Environment," Computers *& Education*, vol. 55(1), pp. 155–164, 2010.

[12] H. Ishizuka, "Language Learning System Realizing SLA Theories in a 3D Virtual World," in T. Bastiaens, and M. Ebner, Ed., *Proceedings of World Conference on Educational* Multimedia*, Hypermedia and* Telecommunications, pp. 3651–3659, Chesapeake, VA: AACE, 2011.

[13] R. Davis, and D. Wong, "Conceptualizing and Measuring the Optimal Experience of the eLearning Environment," *Decision Sciences Journal of Innovative Education*, vol. 5/1, pp. 97–126, 2007.

[14] S. S. Liaw, H. M. Huang, and & G.D. Chen, "Surveying instructor and learner attitudes toward e-learning," *Computers & Education*, vol. 49 (4), pp. 1285–1296, 2007.

[15] R. Abouserie, and D. Moss, "Cognitive style, gender, attitude toward computer assisted learning and academic achievement," *Educational Studies*, 18(2), pp. 51–160, 1992.

[16] S. Kesici, I. Sahin, and A. O. Akturk, "Analysis of cognitive learning strategies and computer attitudes, according to college students' gender and locus of control," *Computers in Human Behaviour*, no. 25, p. 532, 2009.

[17] B. E. Jr. Whitley, "Gender differences in computer-related attitudes and behavior: A meta-analysis," *Computers in Human Behavior*, vol. 13(1), pp. 1–22, January 1997.

[18] I. Varank, "Effectiveness of quantitative skills, qualitative skills, and gender in determining computer skills and attitudes: A causal analysis," *The Clearing House*, vol. 81(2), pp. 71–80, 2007.

[19] A. M. Solvberg, "Gender differences in computer-related control beliefs and home computer use," *Scandinavian Journal of Educational Research*, vol. 46(4), pp. 409–427, 2002.

[20] R. H. Kay, "An analysis of methods used to examine gender differences in computer-related behavior," *Journal of Educational Computing Research*, vol. 8(3), pp. 277–290, 1992.

[21] F. Y. Yu, H. J. Jessy Yu, "Incorporating e-mail into the learning process: its impact on student academic achievement and attitudes," *Computers & Education*, no. 38, pp. 117–126, 2002.

[22] A. S. Lee, "Electronic mail as a medium for rich communication: an empirical investigation using hermeneutic interpretation" *MIS Quarterly*, pp.143–157, 1994.

[23] L. F. Ruberg, and M. G. Miller, "Moving from U.S. mail to e-mail," ERIC No: ED363312, 1993.

[24] J. Webster, and P. Hackley, "Teaching effectiveness in technology-mediated distance learning," *The Academy of Management* Journal, vol. 40(6), 1282–1309, 1997.

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 105–110

IaaS Cloud Model for e-Ordering and e-Invoicing

Kiril Kiroski, Marjan Gusev and Sasko Ristov
Ss. Cyril and Methodius University,
Faculty of Computer Sciences and Engineering,
Rugjer Boshkovikj 16, PO Box 393, 1000 Skopje, Macedonia
Email: {kiril.kjiroski, marjan.gushev, sashko.ristov}@finki.ukim.mk

***Abstract*—This paper addresses the modeling of an IaaS cloud e-Ordering and e-Invoicing solution. We have faced several challenges in the process of migration of the existing web application on the cloud, such as the reorganization of the proposed non-cloud model and the definition of the software modules that can communicate in a dynamic environment with various number of virtual machines. In the design process we have classified two kinds of software modules: static and dynamic. The classification is based on the level of required elasticity and scalability, as well as the activity that is associated with the corresponding module (tool). To finalize implementation of this idea, we have developed indicators to determine the required conditions for activation of additional cloud resources, or deactivation of involved resources. The result of this design process is a new cost effective solution with increased availability and performance compared to the non-cloud model. We have also recognized the advantages and disadvantages of such cloud solution, and discussed the characteristics that make our approach unique.**

***Index Terms*—Cloud Architecture; Cloud Computing; e-Invoicing; e-Ordering.**

I. Introduction

PROCUREMENT is the fundamental business process of each company, organization or government institution, since its basic functionality is to use or offer various products, services and/or information. Electronic procurement provides basic means to fulfill these processes, since it enables a faster, more meaningful and more integrated way to carry out corresponding purchases or sales. While electronic procurement represents a wide area of interest, authors in this paper concentrate on electronic orders and electronic invoices, as constituting parts of the post-award phase of electronic procurement.

E-Ordering and e-Invoicing are two basic processes in the e-Business area, since the essence of the business activity of a given company is the exchange of goods that another company provides. Therefore, each business process ends with a kind of an offer, order and invoice that defines the parameters of the exchange and its value.

Cloud computing, on the other hand, is a relatively new technology, which has been in use in some kind of form for decades [1]. For example, virtualization is a technique for masking the physical characteristics of computers and computer related resources, thus enabling a uniform way for other systems, applications or end-users to interact with the above mentioned resources [2]. It is arguably the biggest technology driver behind cloud computing present for almost 40 years, having a long history, starting in the mainframe environment and arising from the need to provide isolation among users [3], i.e. to enable a multi-tenant environment.

Modern computers are sufficiently powerful to use virtualization to present the illusion of many smaller virtual machines (VMs), each running a separate operating system instance [4]. Cloud computing introduces security, scalability and reliability into a group of centrally managed and virtualized physical machines. It offers scalability and elasticity on a large scale, and makes an illusion of limitless resources.

The overall research and development problem in this paper is the description of the design process to develop an IaaS (Infrastructure as a Service) cloud e-Ordering and e-Invoicing solution, although the ideas can be also exploited on other service models, like PaaS (Platform as a Service) and SaaS (Software as a Service).

In this paper, we propose a solution that can efficiently utilize the cloud resources to achieve scalable and elastic system in a multi-tenant environment for e-Ordering and e-Invoicing. The paper is organized as follows. Section II presents the related work that we found in the literature. In Section III, we lay the background for e-Procurement, with accent on e-Ordering and e-Invoicing. Section IV describes the required transformations to migrate the existing web application solution to a cloud based web services. We elaborate the pros and cons of the newly proposed model and its architecture in Section V. Section VI discusses how the model mitigates the virtualization negative impact to the performance and achieves maximum performance with the minimal utilized resources. Finally, Section VII presents conclusions and plan for future work.

II. Related Work

In this section we provide an overview of recent research in this field, and give an overview of e-Ordering and e-Invoicing solutions available on the market for general use, most frequently in the form of e-Procurement or e-Purchasing.

Today's global competitive market marshals new and innovative technologies into everyday processes. According to Boss et al. [5], procurement processes that take longer time, can be streamlined with cloud computing. The vast number of various researches has proven that public procurement conducted through electronic means can be vital to create new jobs and generate sustainable and long-term development of businesses, providing tax payers and public services' users

with best value for money, as stated by Carayannis and Popescu [6].

Research of e-Procurement in Europe, conducted by Bof and Previtali [7], provides investigative insight into various national models and their respective economic impact on both government and citizens's welfare. Trkman and McCormack [8] analyze e-Procurement benefits through related technological and organizational changes, while also estimating advantages and disadvantages of public procurement through electronic means.

Kiroski et al. [9] evaluate the development and trends for e-Government services, especially e-Procurement and its sub phases, i.e. e-Ordering and e-Invoicing. They show that the most optimal path for modeling and developing functional e-Ordering and e-Invoicing solution is to utilize all benefits of cloud computing, and particularly Software as a Service (SaaS), being today's most utilized and popular form of cloud computing. The authors have previously established the major characteristics of the existing software solutions, measured their success in implementation of the major software goals, and evaluated the level of completion of the objectives, the system should possess. The main idea is to follow international standards regarding ordering and invoicing software, and thus ensure interoperability with most European platforms [10]. In this paper we present means to enable cloud functionalities of the improved model for e-Ordering and e-Invoicing solution [11].

The ultimate goal of e-Purchasing through electronic means is to minimize the cost of producing, transporting and storing this kind of documents, to minimize cost and expenses of the procurement process. Estimates of decreased prices through electronic purchases go as high as fantastical 20 percent [12]. According to Pan-European Public Procurement OnLine (PEPPOL), if all public procurement can be accomplished through such platform, annual savings will surpass 50 billion Euros [13].

III. Background

Every company, as it begins working, regards computers and printers as basic costs of work. Then, they need to be interconnected within a company, after which servers for data storage and centralized resource management should be acquired. These steps, of course, require the involvement of trained and educated staff which will take care of installing, maintaining and servicing of the software.

While the company grows, the need for using new technologies, platforms, software and new employees who will maintain the information system arises. These additional expenses impose an additional burden on companies, regarding both the initial and system maintenance costs.

Cloud computing can help in this kind of situation, since it enables lease and use of new software only when needed, and then cancels the services or replaces them with other ones, without the obligation to change the business process structure, information technology or IT staff in the company. This is all due to the fact that everything is "on the Cloud" [14].

Advantages of using such a system based on Web service, and particularly through cloud implementation, instead of Web application are numerous, and include:

- Elasticity and scalability - resources are provided and freed according to current needs.
- Simplicity - information provided through services are very easily adaptable to a suitable form for our needs, and sometimes it is even not necessary to adapt them.
- Cost-effectiveness - using cloud services does not require additional technical staff, or additional hardware resources.
- Business agility - cloud computing enables companies to quickly start using a service, scale used resources according to actual needs, forge links to other cloud services, or stop using them altogether.
- Resiliency - clouds are built from a large number of inexpensive hardware resources, and applications are hosted in virtualized environments, thus lowering the chances to lost connectivity, functionality or even data in cases of failures in the cloud.
- Interoperability - if need arises to exchange information between different systems, web services (in case of accordingly correctly formatted attributes) can facilitate this process to great extent.

The main goal of this paper is to propose a cloud enabled model, which enables the above stated benefits, and at the same time minimizes the problems that are encountered while working on Internet.

Implementations of an e-Ordering and e-Invoicing solution through traditional provision of web services cannot offer sustainable performance since the predicted system may require huge amount of unpredictable resources only in appropriate small periods during the peak period (such as end of one year and beginning of another), while most of the time, this system requires much less predictable resources. Using server-cluster architecture to create scalable and highly available solutions [15] will only partially solve a performance peak problem since it is hard to be managed and administered. On the other hand, the software solution has to be scalable. The solution presented in previous sections can be hosted on a classic, third generation system architecture, as depicted in Figure 1. The third generation system architecture consists of three subsystems, i.e. Application Layer, Middleware and Application Domains. We can notice that there are no mechanisms for providing support in case of increased usage, hardware or network problems, and so on.

Cloud computing, as a new technology trend, provides computing resources in addition to software in form of a service [17]. Cloud providers offer on-demand illusion of infinite elastic resources using virtualization [18]. Cloud computing addresses a set of services providing scalable, QoS guaranteed, normally personalized, inexpensive computing platforms on demand [19]. Cloud billing model is similar to basic utilities billing models, i.e. it is proportional to the amount of use. Therefore, cloud computing is admissible for small

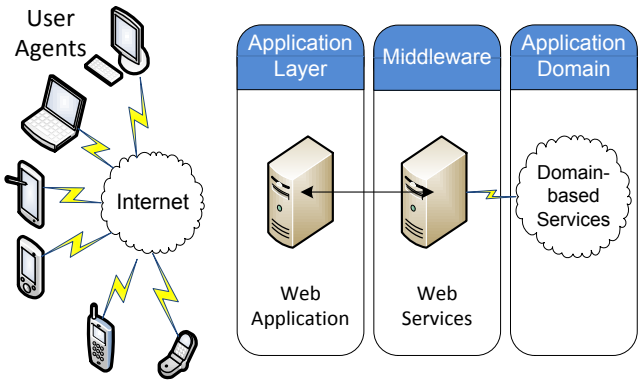


Fig. 1. Third generation architecture, using Web services [16]

and medium business enterprises, especially in the present economic crisis.

Small and medium enterprises can benefit if they migrate their e-Ordering and e-Invoicing systems onto the cloud since the cloud is capable of offering excellent flexibility and scalability of resources, storage space, computational power and network access, and most importantly lower cost. Cloud lowers the disk space requirements, and enables newest software versions and monitoring of the installation progress in each of the cloud service models, i.e. IaaS, PaaS and SaaS.

IV. Modeling the Cloud Solution

In this section we describe the required transformations to migrate an existing solution from a web application to an IaaS cloud-based solution. The cloud-based modular approach is used to realize e-Ordering and e-Invoicing software modules and tools. A new idea is introduced to enable elasticity and scalability based upon used resources, i.e., virtual machines (VMs), in a multi-tenant cloud environment.

We have faced several challenges to move the existing solution onto a cloud. The main problem concerned how to organize different software tools in modules and than schedule them on a particular VM in an efficient cost-effective architecture model. The first step was to find out which software tools need extensive processing and therefore to classify them according to the needed resources for their execution. Finally, this process will determine the scheduling of software tools into modules and map them onto various VMs.

Based on previous research and experimenting, and model developed for e-Ordering and e-Invoicing, inclusion of model's software tools into modules resulted with the following classification. The classification criteria is based on determination of processing demand and possible scheduling to a VM instance. We propose an organization of system modules into two groups: Static and Dynamic. The former is composed of the software modules (tools) that do not require increased computing power, as these modules (tools) are not utilized on a regular basis, and will not be accessed by every user. The latter refers to e-Orders and e-Invoices creation tools, and related common activities. Software modules (tools) to be

TABLE I
Proposed organization of software modules (tools).

Activity	Modules	Related Tools
Static	Management	User and web site administration tool
		Backup and recovery tool
	Usability	Creating and managing catalogues
		Customization tool
		Workflow establishing tool
		Reporting tool
Dynamic	Basic Ops.	Order creation and editing tool
		Invoice creation and editing tool
		Search and view tool
	Advanced Ops.	Invoice dispute tool
		Resolve dispute tool
		Invoice revoke tool
		Order revoke request tool
		Resolve revokes tool

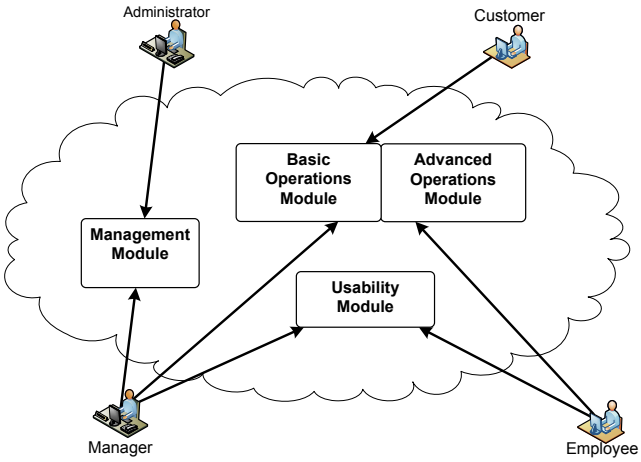


Fig. 2. Cloud-based modular approach to e-Ordering and e-Invoicing tools and modules.

realized can be clustered in relation to these requirements. The results of software module analysis are presented in Table I.

In addition to the proposed classification of software tools into Static and Dynamic groups, we have also applied a clustering method. Our deeper analysis showed that classified software modules can be implemented as four different clusters - software modules, created according to static or dynamic nature, in addition to the basic criteria for clustering as a logical group. This process will allow improved functionality, security, and access control. An example of such a implementation of modular solution can be found in [16].

The new design resulting from both the classification and clustering of software modules of e-Ordering and e-Invoicing is shown in Figure 2.

Static software modules for the proposed solution are:

- Management module - intended for Administrators and partially Managers, since main activities involve occasionally used tools for user administration, web site administration, and backup and recovery tool.
- Usability module - used by Employees to create and maintain Catalogues, and tool for forms customization. It also contains Workflow establishing tool and Reporting

tool, used by Managers and Employees to automate document creation and for reporting activities.

Dynamic software modules, which require more resources at peak times, and provide critical functionalities of the solution, are:

- Basic operations module - used by Managers, Employees and Customers alike, in order to create, edit and delete orders and invoices. It also provides Search and view tool for created orders and invoices.
- Advanced operations module - this module involves advanced activities with invoices and orders, such as disputing invoices and revoking invoices and orders, in most part by Employees and Customers. It also enables Managers to resolve disputes, and decide about revoke requests.

Figure 2 presents four different types of users. Each of those users access one or more of the indicated modules to accomplish authorized tasks. Breakdown of users and respective tasks are shown in Table II. Functionalities provided through system modules were mapped for appropriate user groups:

- The administrator can manage users, by giving certain roles to different types of users. He/she also customizes the look and functionalities on the web site. Backup and recovery are also the responsibility of the administrator.
- The manager can manage employees and customers, and customize certain global fields (tax rates, discounts...). He/she can determine the level of automation and workflow for different employees, and is able to make decisive conclusions regarding orders and invoices. The manager resolves order revocation requests. The manager can also use Search and view tool and Reporting tool.
- The employee can create order or invoice when required, and can edit them when approved by a manager. He/she can resolve invoice disputes in case he/she is authorized to. The employee is responsible for creating catalogues and linking them to orders. He/she can also use Search and view tool and Reporting tool, and track customers' payment history.
- The customer can view the catalogue and use it to create orders. He/she can also edit still unprocessed orders and require manager attention when items in question surpassing employee authority, or dispute about order/invoice content. The customer can request an order to be revoked entirely.

TABLE II
USER GROUPS FOR E-ORDERING AND E-INVOICING SOLUTION.

User Groups	Roles
Administrator	Designs the Home page
	Edits the home page fields
	Manages user accounts for managers, employees and customers
	Manages backup and recovery facility
Manager	Creates list of employees
	Manages customer accounts
	Views/analyses reports about incoming orders
	Views/analyses status reports for invoices
	Edits invoices
	Edits global invoice settings
	Edits global order settings
	Approval of invoices/orders
	Revoke invoices
	Approve revoking of orders
Employee	Creates invoices
	Edits common invoice fields
	Views status reports for invoices
	Contacts slow paying customers
	Maintain Catalogues
	View orders
	Resolve invoice disputes
	Revoke invoices
Customer	Creates orders
	View order status
	View/pay invoices
	Dispute invoices
	Request revoking of orders

V. PROS AND CONS

This section addresses a discussion of possible cloud solution's advantages and disadvantages. A great deal of both positive and negative effects is due to the nature of the cloud approach, while the solution's scope and area of application also provides for some of the possible benefits and shortcomings.

A. Advantages

This section discusses the benefits of the new proposed model and architecture. The proposed model is capable of utilizing features and characteristics offered as a solution for each cloud service layer from IaaS to SaaS. This model for e-Ordering and e-Invoicing system can be hosted onto the cloud and can use seemingly infinite cloud resources dynamically to reduce the costs and to provide better performance.

1) Elasticity and Scalability: The underlying infrastructure in the proposed cloud platform is based on VMs, all sharing a common resource pool, enabling automated utilization of more resources in peak times, and less resources (which will be routed to other VMs hosted in the Cloud) when there is no more need for them. Such benefits could be enabled through the use of data clusters, web server load-balancing, etc.

2) Simplicity: Basic services can be provided as individual services or applications. Those are:

- User access management;
- Orders, Catalogues and Invoices;
- Automated purchase lifecycle; and
- Document export and import.

Above could be provided as a simple web service or an complete web application on discretionary of the software user.

3) Cost-effectiveness: Management and maintenance of the e-Ordering and e-Invoicing solution will be a responsibility of the provider, requiring no technical staff on behalf of the service consumer different software level agreements:

- Web based solution available for purchase;
- Paying for using the service on periodical bases; and
- Per-transaction fee.

4) Business agility: Software lifecycle is provided in a few quick steps, such as:

- Companies and individuals will be able to apply for using the service;
- Administrator will be able to register new clients in the roles of managers and employees;
- Manager will be able to register new customers, as well as grant or revoke special permissions to employees; and
- Revoke order shall also permit requests for nullifying the contract for using the service.

Authorized users that are assigned with a role of Administrator will be able to disable clients, i.e. the managers and employees. A backup and recovery tool enable the users to download the documents.

5) Resiliency: E-Ordering and e-Invoicing solution is intended to run on several VMs on a specific cloud platform. A number of cloud platforms are capable of running such a software.

B. Disadvantages

There is a wide application area of an e-Ordering and e-Invoicing system, especially when most of the European countries will support the electronic procurement. Electronic documents will not just replace the existing orders and invoices, but will enable efficient usage in company's inventory systems and bookkeeping software, or in realization of the ERP (Enterprise Resource Planning) at broader sense. Non-existence of such a solution shall inquire additional initial effort for its development. On the other hand, the necessity for interoperability with an existing platform on European or world level, shall necessarily define the way such a system should be designed. Although this model for e-Ordering and e-Invoicing cloud-based solution offers significant benefits both for business and individuals, its architecture implies additional effort on resolving following challenges.

1) Service model: In this paper, the accent is mostly on presenting an IaaS solution for e-Ordering and e-Invoicing, while our future efforts will include PaaS and SaaS counterparts.

2) Interoperability: Realization of the solution in order to integrate with existing platforms. While the software is distinctive and provides unique features, it must provide an interoperability with software from other vendors and/or countries.

3) Portability: Designing software capable for migration between platforms and service providers. Solution can be developed on one platform, but it must enable fast and streamlined porting on another platform.

4) Security: Implementing security mechanisms and best practices in order to guarantee data security, integrity, and confidentiality. Financial data is one of the most sensible data in today's world, so great deal of problems may source out from lack of safety and security measures.

5) Data privacy: Catalogs, orders and especially invoices, are sensitive data and the customers require access management. Different economic operators may not gain access into one another's personified data.

6) Law compliance: Validity and legality of electronic documents, especially invoices must be compliant with governing laws for both parties' countries. Since this software is intended for multinational use, in-compliance to national governing laws and regulations is not acceptable and may incur further implications. Additionally, the private data should not leave the EU countries, regardless if they are paper or electronic.

7) Total Costs: Although cloud solutions are cost-effective, total costs may incur over-expenditure in case of excessive use. Software economic model must take into account economic justifiability for using cloud solution on agreed SLA, and not simply buying and owning the platform.

VI. Discussion

The cloud virtrual environment behaves different for the same load in differently time depending on the number of active VMs on a particular cloud node and the total number of active VMs in the whole cloud [20]. Ristov et al. [21] determined that the virtualization in cloud computing degrades almost 30% of the SOA (Service Oriented Architecture) web service performance compared to traditional "bare metal" web service hosting infrastructure with the same hardware resources. Lloyd [22] found that the performance decreases enormously for input / output intensive applications (more than 110% overhead) and decreases lightly (about 10% overhead) for CPU intensive applications. Windows Azure can achieve even superlinear speedup (speedup greater than linear speed, i.e., greater than the number of used threads and CPU cores while scaling the resources) [23].

Therefore, the real challenge is how to mitigate this relatively huge performance degradation and performance discrepancy. Migrating the existing third generation system architectures on the cloud and thus using the cloud features of scalable and elastic theoretically unlimited resources can even increase the cost, rather than to decrease it.

The most common cloud providers, such as Microsoft (Windows Azure), Google (Compute), or Amazon (EC2) offer different types of VM instances. Although the prices and SLAs (Service Level Agreements) are different for each cloud provider for the same resources, their pricing models have the common issue, that is, the price increases linearly while scaling the resources [24], [25], [26]. But many authors show that the performance is not scaling the same as the resources (and the price). The maximum speedup (performance) is limited by Gustafson's Law [27].

Gusev et al. determined that Multi-VM environment (many VMs with a single CPU core) provides up to 10 times better performance compared to the Single-VM environment (one huge VM allocated with all cores) [28]. Therefore, scaling the number of VMs (and balancing the load among them) is the better solution rather than scaling the resources in a single VM instance. The former solution is more reliable and available, because the system will continue to work if some VMs fail. Introducing a resource broker in our solution reduces the number of active VMs, thus reducing the overall cost of the solution and making it a green cloud solution.

VII. Conclusion and Future Work

E-Ordering and e-Invoicing are the most important parts of an e-Procurement business process. In this paper, we presented the design of an IaaS cloud e-Ordering and e-Invoicing model. The final goal is to efficiently and effectively utilize the cloud using scalable, elastic and theoretically unlimited computing and storage resources.

We have realized a new IaaS cloud solution, by identifying and organizing static and dynamic software modules. Dynamic modules can activate various virtual machines, forming an efficient solution.

Future work will be based on experimental research of performance of this solution and compare several other solutions. Also we plan to develop PaaS and SaaS elastic and scalable solutions.

Our proposed cloud model focuses only on a part of e-Procurement, i.e. for e-Ordering and e-Invoicing. We will continue with our research to propose a comprehend cloud model for the whole e-Procurement business process.

While proposing cloud model for partial e-Procurement solution, we are confident that proposed model can be also beneficial in other application domains, such as asset management, content delivery networks, computing, etc.

References

[1] B. Williams and G. Reese, *The Economics of Cloud Computing*. Cisco Systems, 2012.

[2] IBM, "White paper: Virtualization in education." 2007. [Online]. Available: http://www-07.ibm.com/solutions/in/education/download/Virtualization\%20in\%20Education.pdf

[3] S. Crosby and D. Brown, "The Virtualization Reality: Are hypervisors the new foundation for system software?" *ACM queue – Computer Architecture*, vol. 4, no. 10, Dec. 2006. [Online]. Available: http://queue.acm.org/detail.cfm?id=1189289

[4] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," *ACM SIGOPS Operating Systems Review*, vol. 37, no. 5, pp. 164–177, 2003.

[5] G. Boss, P. Malladi, D. Quan, L. Legregni, and H. Hall, "Cloud computing," *IBM white paper*, vol. 1369, 2007.

[6] E. G. Carayannis and D. Popescu, "Profiling a methodology for economic growth and convergence: learning from the eu e-procurement experience for central and eastern european countries," *Technovation*, vol. 25, no. 1, pp. 1 – 14, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166497203000713

[7] F. Bof and P. Previtali, "National models of public (e)-procurement in europe," *Journal of E-Government Studies and Best Practices*, vol. 5, pp. 1–14, 2010.

[8] P. Trkman and K. McCormack, "Estimating the benefits and risks of implementing e-procurement," *Engineering Management, IEEE Transactions on*, vol. 57, no. 2, pp. 338–349, 2010.

[9] K. Kiroski, M. Kostoska, M. Gusev, and S. Ristov, "Growth rate analysis of e-government development," in *Proceedings of the Fifth Balkan Conference in Informatics*, ser. BCI '12. Novi Sad, Serbia: ACM, 2012, pp. 106–111.

[10] O. Association, "About OpenPEPPOL," Feb. 2013. [Online]. Available: http://www.peppol.eu

[11] K. Kiroski, M. Gusev, M. Kostoska, and S. Ristov, "Modifications and Improvements on CEN/BII Profiles," in *ICT Innovations 2011*, ser. Advances in Intelligent and Soft Computing, L. Kocarev, Ed. Springer Berlin / Heidelberg, 2012, vol. 150, pp. 395–404.

[12] M. Singer, G. Konstantinidis, E. Roubik, and E. Beffermann, "Does e-procurement Save the State Money," *Journal of Public Procurement*, vol. 9, no. 1, pp. 58–78, 2009.

[13] PEPPOL, "Future eProcurement," Feb. 2013. [Online]. Available: http://www.peppol.eu/peppol-project/eu-wide-interoperability/future-eprocurement

[14] GTSI Group, "Cloud Computing: Building a Framework for Successful Transition," 2009. [Online]. Available: http://www.gtsi.com/cms/documents/White-Papers/Cloud-Computing.pdf

[15] J.-H. Ho, M.-Y. Luo, and C.-S. Yang, "Building a Scalable Digital Learning System on Server Clusters," *J. Inf. Sci. Eng.*, vol. 23, no. 3, pp. 803–819, 2007.

[16] S. Ristov, M. Gusev, G. Armenski, K. Bozinoski, and G. Velkoski, "Architecture and organization of e-assessment cloud solution," in *Global Engineering Education Conference (EDUCON), 2013 IEEE*, March 2013, pp. 736–743, best paper award.

[17] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, "A View of Cloud Computing," *Comm. of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[18] P. M. Mell and T. Grance, "SP 800-145. The NIST Definition of Cloud Computing," National Institute of Standards & Technology, Gaithersburg, MD, United States, Tech. Rep., 2011.

[19] L. Wang, J. Tao, M. Kunze, A. Castellanos, D. Kramer, and W. Karl, "Scientific Cloud Computing: Early Definition and Experience," in *10th IEEE International Conference on High Performance Computing and Communications, 2008. HPCC'08.* IEEE, 2008, pp. 825–830.

[20] Y. Koh, R. Knauerhase, P. Brett, M. Bowman, Z. Wen, and C. Pu, "An analysis of performance interference effects in virtual environments," in *ISPASS 2007, IEEE Int. Symp. on*, april 2007, pp. 200 –209.

[21] S. Ristov, G. Velkoski, M. Gusev, and K. Kiroski, "Compute and Memory Intensive Web Service Performance in the Cloud," in *ICT Innovations 2012*, ser. Advances in Intelligent and Soft Computing, S. Markovski and M. Gusev, Eds. Springer Berlin / Heidelberg, 2013, vol. 257, pp. 215–224.

[22] W. Lloyd, S. Pallickara, O. David, J. Lyon, M. Arabi, and K. Rojas, "Migration of multi-tier applications to infrastructure-as-a-service clouds: An investigation using kernel-based virtual machines," in *Grid Computing (GRID), 2011 12th IEEE/ACM International Conference on*, 2011, pp. 137–144.

[23] M. Gusev and S. Ristov, "Superlinear speedup in windows azure cloud," in *2012 IEEE 1st Int. Conf. on Cloud Networking (CLOUDNET) (IEEE CloudNet'12)*, Paris, France, Nov 2012, pp. 173–175.

[24] Amazon, "Ec2," May 2013. [Online]. Available: http://aws.amazon.com/ec2/

[25] Microsoft, "Windows azure," May 2013. [Online]. Available: http://www.windowsazure.com/pricing/

[26] Google, "Compute engine," May 2013. [Online]. Available: http://cloud.google.com/pricing/

[27] J. L. Gustafson, "Reevaluating Amdahl's law," *Communication of ACM*, vol. 31, no. 5, pp. 532–533, May 1988.

[28] M. Gusev, S. Ristov, G. Velkoski, and M. Simjanoska, "Optimal resource allocation to host web services in cloud," in *Cloud Computing (CLOUD), 2013 IEEE 6th International Conference on*, June 2013, pp. 948–949.

Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 111–116

A Declarative Approach to Cost Estimation in Multi-Project Environment

Marcin Relich
University of Zielona Gora,
ul. Licealna 9, 65-216 Zielona Gora, Poland
Email: m.relich@wez.uz.zgora.pl

***Abstract*—The paper aims to present a cost evaluation for multi-project environment, taking into account imprecision in activity duration and cost. Data specification in the form of discrete α-cuts enables the connection of distinct and imprecise data, and the implementation of a constraints satisfaction problem with the use of constraint programming. Moreover, using α-cuts, optimistic, pessimistic, and several intermediate scenarios concerning the project scheduling and cost can be obtained and considered in terms of different risk levels. Each scenario can be assessed according to criteria such as time, cost, and risk level. A declarative form of the description of a multi-criteria decision problem allows its implementation in constraint programming languages and facilitates the development of a decision support system. The proposed methodology can be easily incorporated into available fuzzy project scheduling software to provide a better perception of risk that is usually obscured in the conventional approach.**

I. Introduction

THE traditional approach to project scheduling is the well-known CPM (Critical Path Method) and PERT (Program Evaluation and Review Technique). The hypothesis made in CPM that activity durations are deterministic and known is rarely satisfied in real life where tasks are often uncertain and variable [1]. The inherent uncertainty and imprecision in project scheduling has motivated the proposal of several fuzzy set theory based extensions of activity network scheduling techniques. Among these extensions can be found, for instance, resource-constrained fuzzy project-scheduling problem [2], criticality analysis of activity networks with uncertainty in task duration [3], fuzzy repetitive scheduling method [4], and fuzzy dependency structure matrix for project scheduling [5]. Considerable research effort has been recently focused also on the application of constraint programming frameworks in the context of project scheduling [6], [7].

The Constraint Programming (CP) environment seems to be particularly well suited to modelling real-life and day-to-day decision-making processes at an enterprise. CP is qualitatively different from the other programming paradigms, in terms of declarative, object-oriented and concurrent programming. Compared to these paradigms, constraint programming is much closer to the ideal of declarative programming: to state what we want without stating how to achieve it [8]. CP is an emergent software technology for a declarative Constraints Satisfaction Problem (CSP) description and can be considered as a pertinent framework for the development of decision support system software aims.

Declarative programming languages base on the idea that programs should be as close as possible to the problem specification and domain [9]. In the field of constraint-based scheduling, two strengths emerge: natural and flexible modelling of scheduling problems as CSP and powerful propagation of temporal and resource constraints. Thus, the scheduling problem is modelled as CSP at hand in the required real-life detail and it enables to avoid the classical drawbacks of being forced to discard degrees of freedom and side constraints. Discarding degrees of freedom may result in the elimination of interesting solutions, regardless of the solution method used. Discarding side constraints gives a simplified problem and solving this simplified problem may result in impractical solutions for the original problem [10]. The limitations of imperative languages provide the motivation to develop a reference model of project management in an enterprise and to implement it in declarative languages. The advantage of working with such a model is that users are driven by the system to produce the required results, whilst the manner in which the results are produced is dependent on the preferences of the users [11].

The model formulated in terms of CSP determines a single knowledge base and it enables effective implementation in constraint programming languages, as well as the development of a task-oriented decision support system (DSS) for project portfolio planning. As a result, the problem specification is closer to the original problem, obtaining solutions that are unavailable with imperative programming. Moreover, the descriptive approach enables the specification of decision problems according to deductive reasoning (a query about the results of proposed decisions) and abductive reasoning (a query about decisions ensuring the expected results).

Cost planning is crucial for the assessment of cash flow during project implementation. Moreover, an accurate cash flow is required in conducting project cost-benefit analysis, the determination of project financing requirements and in performing earned value analysis [12]. Several researchers have applied different approaches to fuzzy set theory or probability theory in project flow generation and analysis (e.g. [13], [14], [15]). However, the main focus in the research concerning fuzzy project scheduling is principally on the calculation of early/late start and finish times and the determination of activity and path criticality, whereas issues related to uncertainty in cost with the use of a declarative approach have not yet been comprehensively addressed.

The goal of this research is to present the use of constraint programming to fuzzy project scheduling and cost evaluation in multi-project environment whose durations and costs are in the imprecise form. The model of project portfolio planning is specified in terms of fuzzy CSP, using constraint programming to seek a solution to the problem, and enabling cost analysis at different α-levels. An α-cut is a crisp set consisting of elements of fuzzy set A which belong to the fuzzy set at least to a degree of α. The proposed methodology is relatively similar to what practitioners are using to generate project cost and cash flows but is considerably more effective and realistic in modelling uncertainty. The proposed DSS for project portfolio planning allows a decision-maker to obtain a set of project scenarios and to perform analysis of cost uncertainty at different α-levels, which appears to be more intuitive than alternative methodologies that employ other fuzzy techniques.

The remaining sections of this paper are organised as follows: Section 2 presents a problem formulation in terms of fuzzy CSP for project portfolio scheduling. A method of fuzzy project scheduling and cost generation is shown in Section 3. An illustrative example of the approach, which presents the decision problem specification, is presented in Section 4. Finally, some concluding remarks are contained in Section 5.

II. Fuzzy constraint satisfaction problem for project portfolio scheduling

The specification of project portfolio scheduling encompasses technical parameters, expert's experiences and user expectations in the form of a knowledge base, i.e. as a set of variables, their domains, and a set of relations (constraints) that restrict and link variables. In this context, it seems natural to classify some decision problems as CSP. The problem formulation in terms of CSP enables a simplified description of actuality, i.e. a description encompasses the assumptions of object, implementing therein tasks, and a set of routine queries - the instances of decision problems [6].

In a classical form, the structure of the constraints satisfaction problem may be described as follows [10]: *CSP* = ((V, D), C), where: V - a set of variables, D - a set of discrete domains of variables, C - a set of constraints. In turn, for the imprecise description of variables, the Fuzzy Constraints Satisfaction Problem (FCSP) takes the following form:

$$FCSP = ((\tilde{V}, D), C) \tag{1}$$

where:

- $\tilde{V} = \tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_n$ - a finite set of n fuzzy variables that are described in the form of fuzzy number (a finite set of discrete α-cuts);
- $D = d_1, d_2, ..., d_n$ - a set of domains for n fuzzy variables;
- $C = c_1, c_2, ..., c_m$ - a finite set of m constraints limiting and linking decision variables.

Given a set of projects $P = \{P_1, P_2, ..., P_I\}$, where the project P_i consists of J activities: $P_i = \{A_{i,1}, ..., A_{i,j}, ..., A_{i,J}\}$. The *j*-th activity of *i*-th project is specified as follows: $A_{i,j} = \{s_{i,j}, z_{i,j}, t_{i,j}, dp_{i,j}\}$, where

$s_{i,j}$ - the starting time of the activity $A_{i,j}$, i.e., the time counted from the beginning of the time horizon H;
$z_{i,j}$ - the completion time of the activity $A_{i,j}$;
$t_{i,j}$ - the duration of the activity $A_{i,j}$, $s_{i,j} < z_{i,j}$;
$dp_{i,j}$ - the financial means allocated to the activity $A_{i,j}$.

The project P_i is described as an activity-on-node network, where nodes represent the activities and the arcs determine the precedence constraints between activities. According to this, the precedence constraints are as follows:

- the *k*-th activity follows the *i*-th one:

$$s_{i,j} + t_{i,j} \leq s_{i,k} \tag{2}$$

- the *k*-th activity follows other activities:

$$\begin{aligned} s_{i,j} + t_{i,j} &\leq s_{i,k}, \\ s_{i,j+1} + t_{i,j+1} &\leq s_{i,k}, \\ &..., \\ s_{i,j+n} + t_{i,j+n} &\leq s_{i,k} \end{aligned} \tag{3}$$

- the *k*-th activity is followed by other activities:

$$\begin{aligned} s_{i,k} + t_{i,k} &\leq s_{i,j}, \\ s_{i,k} + t_{i,k} &\leq s_{i,j+1}, \\ &..., \\ s_{i,k} + t_{i,k} &\leq s_{i,j+n} \end{aligned} \tag{4}$$

CSP can be considered as a knowledge base that is a platform for query formulation as well as for obtaining answers, and it comprises of facts and rules that are characteristic of the system's properties and the relations between its different parts. As a consequence, a single knowledge base facilitates the implementation of a decision support system.

The distinction of decision variables that are embedded in the knowledge base as an input-output variable permits to formulate the standard routine queries concerning project cost analysis such as:

- is there a cost scenario at given α-level and constraints (e.g. project deadline, budget, precedence constraints)?
- is there a cost scenario for a given cost limit, and if yes, what starting times of project portfolio activities $s_{i,j}$ ensure that the cost allocation $dp_{i,j}$ does not exceed the cost limit and other constraints?

The method of generation of admissible solutions for the above-described problem is presented in the next section.

III. Method of fuzzy project scheduling and cost generation

Imprecise variables determined by convex membership function $\mu(t)$ (e.g. a triangular fuzzy number $t =< a, b, c >$) can be specified as α-cuts. An α-cut is a crisp set consisting of elements belong to the fuzzy set at least to a degree of α ($0 < \alpha \leq 1$). An α-cut is a method of defuzzifying a fuzzy set to a crisp set at desired α-levels that correspond to

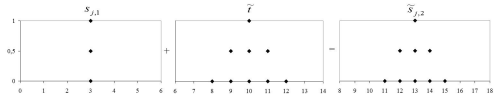


Fig. 1. Addition of fuzzy numbers in terms of discretized α-cuts

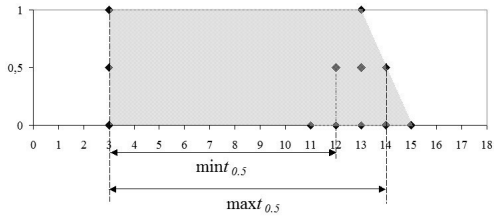


Fig. 2. Start and fuzzy completion time of activity

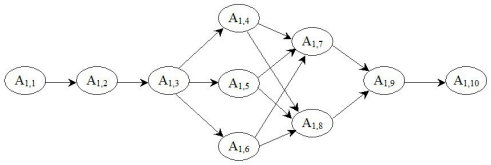


Fig. 3. Network diagram for project P_1

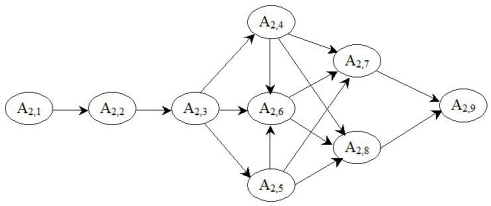


Fig. 4. Network diagram for project P_2

the perceived risk (α=1 meaning no risk, α=0– meaning the lowest risk, α=0+ meaning the highest risk). Additionally, the low (α=0–) and high (α=0+) values of every α-cut represent the optimistic and pessimistic outcomes of that risk level. The main objective of fuzzy project scheduling is to apply fuzzy set theory concepts to the scheduling of real world projects where task duration can be specified as fuzzy numbers instead of crisp numbers [12].

If in the fuzzy project scheduling algorithm, the start and completion times are in fuzzy form, then this usually leads to difficulties with the interpretation, since the fuzzy starting time of the activity can be greater than the fuzzy completion time. In the order to avoid this situation, it is assumed that the starting time of the activity is in a distinct form, whereas the completion time of the activity can be specified as a fuzzy number. The fuzzy completion time is the sum of the activity start with the fuzzy activity duration (see Fig. 1). It is noteworthy that using the presented methodology, the intersection of starting and completion time is impossible and the interpretation is unambiguous.

In order to calculate the required cost per unit of time, the cost of every activity needs to be divided by its duration. However, the duration varies for different possibility measures and for optimistic and pessimistic scenarios. In the absolute best case (mint_α), the activity starts as early as possible and lasts the minimum duration. In the absolute worst case (maxt_α), the activity starts as late as possible and lasts the maximum duration. An example of the interval of minimum (mint_α) and maximum (maxt_α) duration of the activity at the respective α-cut is presented in Fig. 2.

Fig. 2 shows an activity with a starting time of <3, 3, 3>, a duration of <8, 10, 12>, and a completion time of <11, 13, 15>. In this example, the duration intervals at α=0.5 are min$t_{0.5} = [3, 12]$ and max$t_{0.5} = [3, 14]$ and the activity cost is distributed in these intervals. In the best case, the activity begins as early as possible (3rd time unit) and lasts the minimum duration (9 time units), whereas in the worst case, it lasts the maximum duration (11 time units). Equivalently, minimum and maximum duration intervals representing optimistic and pessimistic scenarios for different possibility measures can be created for all α-levels (between 0 and 1). Consequently, the fuzzy start time and completion time mark the temporal start and completion boundaries of the activity within which the minimum and maximum duration intervals (mint_α, maxt_α) are defined for each α-cut.

The uncertainties of the duration and cost of an activity are positively correlated, so the minimum (mint_0) and maximum (maxt_0) cost distribution per unit of time h of the j-th activity at the level α depict the best and the worst scenario respectively. The presented approach can be expanded for several α-cuts (e.g. 0.1, 0.2, ..., 1) for mint_α and maxt_α, generating activity cost accordingly. A number of scenarios depend on a number of α-cuts. For instance, if a fuzzy number is described at 3 α-cuts, then there are 5 scenarios (mint_0, min$t_{0.5}$, t_1, max$t_{0.5}$, maxt_0). In turn, the use of 5 α-cuts results in 9 scenarios (mint_0, min$t_{0.25}$, min$t_{0.5}$, min$t_{0.75}$, t_1, max$t_{0.75}$, max$t_{0.5}$, max$t_{0.25}$, maxt_0). An example of the use of the presented methodology in constraint programming environment is presented in the next section.

IV. Illustrative Example

The example consists of three subsections: the description of the project portfolio, the analysis of the first admissible solution of the fuzzy scheduling problem, and the what-if analysis (the fuzzy scheduling problem for a given cost limitation). Both analyses contain the examination of fuzzy project Gantt charts and fuzzy project cost distribution.

A. Project portfolio description

It is assumed that the time horizon for the project portfolio ($P = \{P_1, P_2, P_3\}$) equals 34 months (H = 0, 1, ..., 34) and the budget of the project portfolio is fixed at 950 m.u. The network diagrams of the activities in the project portfolio are shown in Fig. 3-5.

The duration of some activities ($A_{1,7}$, $A_{1,10}$, $A_{2,4}$, $A_{2,7}$, $A_{2,9}$, $A_{3,4}$, $A_{3,5}$, $A_{3,6}$, $A_{3,7}$) is specified in the imprecise form. The sequences of activity duration for the considered projects can be described as follows: T_1 = (2, 1, 1, 6, 2, 2, "about 6",

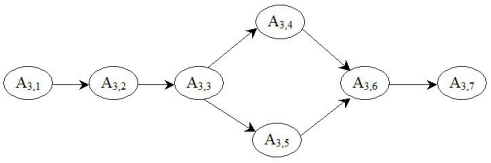


Fig. 5. Network diagram for project P_3

1, 4, "about 6"), T_2 = (2, 2, 1, "about 9", 6, 4, "about 6", 4, "about 4"), T_3 = (1, 1, 1, "about 6", "about 6", "about 4", "about 9"). For instance, the duration of the activity $A_{1,7}$ is "about 6", i.e. the activity can be completed within the time period of 4 to 8 units of time.

B. Fuzzy scheduling and cost distribution: first admissible solution

Fuzzy project scheduling and cost generation problem can be reduced to the following questions: is there a portfolio schedule (and if yes, what are its parameters) that follows from the given project constraints specified by the activity duration times, the deadline and budget of project portfolio? What risk levels are there for the different fuzzy project cost scenarios? The answer to the questions is connected with the determination of the starting $(s_{i,j})$ and completion $(z_{i,j})$ time of project portfolio activities and the allocation of financial means to the activities by different α-level $dp_{i,j,\alpha}$. For the considered project portfolio, and a-level equal to 1, the following sequences are sought: $S_1 = (s_{1,1,1}, ..., s_{1,10,1})$, $S_2 = (s_{2,1,1}, ..., s_{2,9,1})$, $S_3 = (s_{3,1,1}, ..., s_{3,7,1})$, $Z_1 = (z_{1,1,1}, ..., z_{1,10,1})$, $Z_2 = (z_{2,1,1}, ..., z_{2,9,1})$, $Z_3 = (z_{3,1,1}, ..., z_{3,7,1})$, $Dp_1 = (dp_{1,1,1}, ..., dp_{1,10,1})$, $Dp_2 = (dp_{2,1,1}, ..., dp_{2,9,1})$, $Dp_3 = (dp_{3,1,1}, ..., dp_{3,7,1})$.

Fig. 6 presents the first admissible solution (project portfolio schedule), in which the sequences of activity starting and completion time are as follows: S_1 = (0, 2, 3, 4, 4, 4, 10, 10, 16, 20), S_2 = (0, 2, 4, 5, 5, 14, 18, 18, 24), S_3 = (0, 1, 2, 3, 3, 9, 13), Z_1 = (2, 3, 4, 10, 6, 6, "about 16", 11, 20, "about 26"), Z_2 = (2, 4, 5, "about 14", 11, 18, "about 24", 22, "about 28"), Z_3 = (1, 2, 3, "about 9", "about 9", "about 13", "about 22"). The completion time of project P_1, P_2, P_3 equals "about 26", "about 28", and "about 22" months, respectively.

Fig. 7 presents five different cost scenarios for project portfolio (cumulative cost for project P_1, P_2, and P_3). At μ=1, the cost (dotted line) is equivalent to that generated from deterministic analysis. At μ=0.5, there is an optimistic scenario below and a pessimistic one above (dashed line). In turn at μ=0, the optimistic and pessimistic cost scenarios (solid line) have a wider spread indicating a higher degree of uncertainty. In the best case ($\min t_0$), the project portfolio will be completed in 22 months with the total cost of 632 m.u., whereas in the worst ($\max t_0$) in 34 months with the total cost of 920 m.u.

The above-presented S-curves are the basis for analyzing cost scenarios in project portfolio. However, S-curves are simply an edge of an S-surface that in practice is plotted by

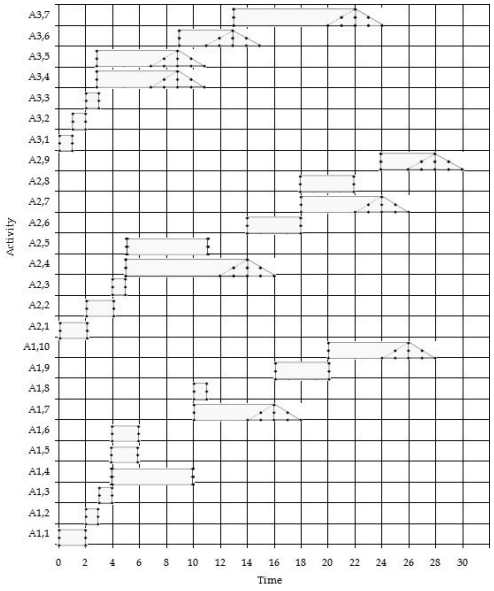


Fig. 6. Project portfolio schedule

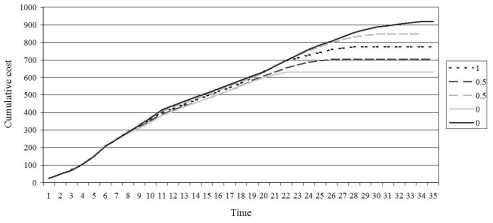


Fig. 7. Cumulative cost for project portfolio

connecting S-curves from some possibility levels from 0 to 1 at selected time periods. Fig. 8 shows the project S-surfaces for the best and worst scenarios. The selection of specific possibility levels and time intervals determines the size of the rectangular patches that form the S-surface and consequently the overall plot quality.

Compared with conventional 2 dimensional S-curves, the S-surface shows how both uncertainty levels and time affect the

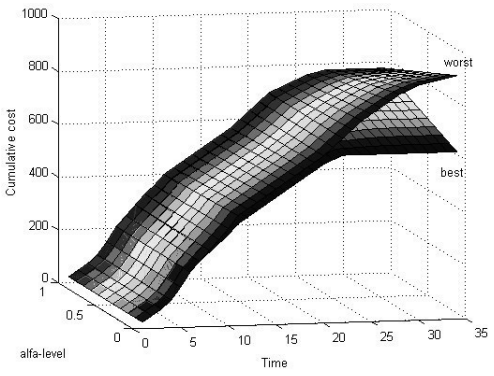


Fig. 8. Optimistic and pessimistic cumulative cost for project portfolio

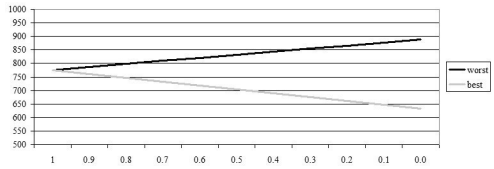


Fig. 9. S-surface cross section at 30th time unit

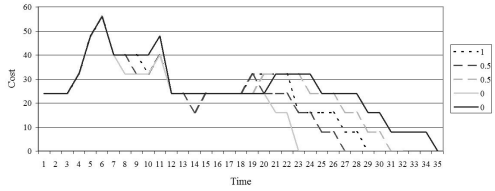


Fig. 10. Cost distribution for project portfolio

project cost. Thus, the surface steepness in terms of possibility and time provides additional insight about the project cost. At μ=1, the best and worst S-surface intersect each other. The presented approach also allows the decision-maker to examine surface cross section at specific times. Fig. 9 illustrates cross section at 30th time unit in which the cost variance of the best and worst case at the possibility level of μ=0 is 632 m.u. and 888 m.u., respectively.

The above-presented examples concern the cumulative cost for project portfolio. Further detailed analyses can include the cost distribution in the horizon of project portfolio (Fig. 10), as well as the analyses in the context of a single project, instead a set of projects.

Let us assume that the cost distribution should be not greater than 40 m.u. in each time unit. This constraint is not fulfilled in the 6th and 11th time units (see Fig. 10). A possibility of searching solution for the additional constraint in presented in the next subsection.

C. Fuzzy scheduling for a given cost limitation

Table 1 presents the results of solution seeking for the different strategies of variable distribution and the two cases: for a fuzzy number described at 3 and 5 α-cuts, respectively. The example was implemented in the Oz Mozart programming environment and tested on an AMD Turion(tm) II Ultra Dual-Core M600 2.40GHz, RAM 2 GB platform. The results show that the Naive and Split distribution strategy outperforms the First-fail ones.

The size of the instance equals 6,000; in turn the number of solutions equals 700 and 800 at 3 and 5 α-cuts, respectively. The sequences of activity starting and completion time are as follows: S_1 = (0, 2, 3, 4, 4, 6, 10, 11, 16, 20), S_2 = (0, 2, 4, 5, 8, 14, 18, 18, 24), S_3 = (0, 1, 2, 3, 3, 9, 9, 13), Z_1 = (2, 3, 4, 10, 6, 8, "about 16", 12, 20, "about 26"), Z_2 = (2, 4, 5, "about 14", 14, 18, "about 24", 22, "about 28"), Z_3 = (1, 2, 3, "about 9", "about 9", "about 13", "about 22"). The first admissible solution of project portfolio completion for the minimal total duration of project portfolio is presented in Fig. 11.

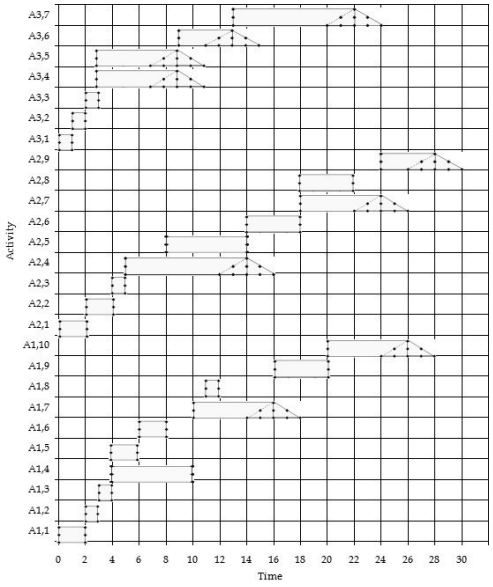


Fig. 11. Project portfolio schedule for limited cost distribution

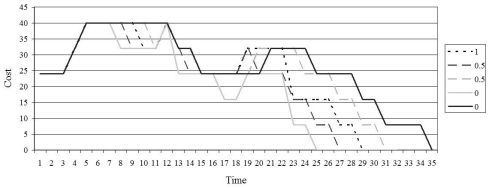


Fig. 12. Limited cost distribution

The re-scheduling implies the reallocation of financial means in project portfolio that is presented in Fig. 12. It is noteworthy that the cost distribution fulfils the constraint ($dp_{i,j} \leq 40$ m.u.). Moreover, the cost allocation is more even than for the case in subsection 4.2 (see Fig. 10). The presented approach allows the decision-maker to consider a wide range of further analyses. For instance, a risk level for cost scenario can be treated as an additional criterion for reducing a set of admissible solutions. The obtained schedules and cost scenarios provide a plan for project portfolio execution and are a basis for further adjustment aimed at fitting to real live execution.

V. Conclusions

The activity of a present enterprise comprises turbulent changes concerning technology, economics, and society [16]. Most projects are executed in the presence of uncertainty and are difficult to manage, due to comprising of many activities linked in a complex way. Hence, there is an increase in demand for new knowledge that enables the solution of problems encountered during complex project portfolio execution. In this case, knowledge concerning project management, especially fuzzy project scheduling, is particularly significant. In the current project implementation environment, a pure deterministic approach for the study of project cost is inadequate.

TABLE I
COMPARISON OF STRATEGIES FOR VARIABLE DISTRIBUTION

Case	Distribution strategy	Number of solutions	Depth	Time [sec]
3 α-cuts	Naive	700	25	7.51
3 α-cuts	First-fail	700	25	7.93
3 α-cuts	Split	700	16	7.57
5 α-cuts	Naive	800	25	10.00
5 α-cuts	First-fail	800	25	10.80
5 α-cuts	Split	800	16	10.41

The proposed approach takes into account several elements, such as the fuzzy activity cost and duration estimations, project S-surfaces, and cost distribution analysis. Data specification in the form of α-cuts enables the generation of a set of scenarios concerning the project scheduling and cost that can be assessed according to risk level. Moreover, the use of discrete α-cuts facilitates the merger of distinct and imprecise data, and implementation of a constraints satisfaction problem in the constraint programming environment that solves CSP with a significant reduction of the amount of search space. As a result, a task-oriented decision support system has been effectively developed. This system can support a decision-maker in obtaining answers to the following questions: is there a portfolio schedule (and if yes, what are its parameters, e.g. starting time of activities, risk level), and what starting times of project portfolio activities can ensure the specified level of project risk and the required cost allocation?

The limitations of existing commercially available tools (e.g. lack of possibility for data specification in an imprecise form, lack of abilities to solve problems defined in the multi-project environment) was the motivation to develop a design methodology for task oriented decision support systems aimed at project portfolio scheduling and fuzzy project cost generation. In that context, the presented approach can be considered as a new contribution to project management.

The number of α-levels can be modified according to the decision-maker's requirements. As a result, it can assist project managers to gain deeper insight into the sources and extents of uncertainty, which may in turn lead to the avoidance of troubles during project implementation. Also, as the methodology is useful in the assessment of financial requirements during project realization, it may prove practical in evaluating alternative project proposals during the feasibility stage. Moreover, it tends to achieve a balance between complexity of methodology and an intuitive, effective decision support system that is realistic in modelling uncertainty. Finally, its application in performing earned value analysis during project monitoring can also obtain useful results.

The subject of future research includes an extension of the proposed model to other fields of project management where the decision problems are semi-structured, such as project communications management and project team building. Moreover, further research will be aimed at developing a decision support system towards real-life verification. Future research also includes a determination of membership functions of fuzzy numbers by using e.g. fuzzy neural system, and their description in the discretized α-cuts.

REFERENCES

[1] F. A. Zammori, M. Braglia, and M. Frosolini, "A fuzzy multi-criteria approach for critical path definition," *International Journal of Project Management,* vol. 27, 2009, pp. 278–291.

[2] P. Bonnal, K. Gaurc, and G. Lacoste, "Where do we stand with fuzzy project scheduling?" *Journal of Construction Engineering and Management,* vol. 130(1), 2004, pp. 114–123.

[3] J. Fortin, P. Zielinski, D. Dubois, and F. Fargier, "Criticality analysis of activity networks under interval uncertainty," *Journal of Scheduling,* vol. 13, 2010, pp. 609–627.

[4] A. Maravas and J. P. Pantouvakis, "A fuzzy repetitive scheduling method for projects with repeating activities," *Journal of Construction Engineering and Management,* vol. 137, 2011, pp. 561–564.

[5] Q. Shi and T. Blomquist, "A new approach for project scheduling using fuzzy dependency structure matrix," *International Journal of Project Management,* vol. 30, 2012, pp. 503–510.

[6] Z. Banaszak, M. Zaremba, and W. Muszyñski, "Constraint programming for project-driven manufacturing," *Int. J. Production Economics,* vol. 120, 2009, pp. 463–475.

[7] G. Bocewicz, Z. Banaszak, and W. Muszynski, "Decision support tool for resource allocation subject to imprecise data constraints," IEEE International Conference on Control and Automation, New Zealand, 2009, pp. 1217–1222.

[8] P. Van Roy and S. Haridi, *Concepts, techniques and models of computer programming,* Massachusetts Institute of Technology, 2004.

[9] P. Hofstedt and P. Pepper, "Integration of declarative and constraint programming," *Theory And Practice of Logic Programming,* vol. 7, 2007, pp. 93–121.

[10] F. Rossi, P. van Beek, and T. Walsh, *Handbook of Constraint Programming,* Elsevier Science, 2006.

[11] M. Pesic, and W. M. P. van der Aalst, "A declarative approach for flexible business processes management," In: *Business Process Management Workshops, Lecture Notes in Computer Science,* Springer, vol. 4103, 2006, pp. 169–180.

[12] A. Maravas and J. P.Pantouvakis, "Project cash flow analysis in the presence of uncertainty in activity duration and cost," *Journal of Construction Engineering and Management,* vol. 30, 2012, pp. 374–384.

[13] G. A. Barraza, W. E. Back, and F. Mata, "Probabilistic forecasting of project performance using stochastic S curves," *Journal of Construction Engineering and Management,* vol. 130(1), 2004, pp. 25–32.

[14] J. S. Yao, M. S. Chen, and H. F. Lu, "A fuzzy stochastic single-period model for cash management," *European Journal of Operational Research,* vol. 170(1), 2006, pp. 72–90.

[15] M. Y. Cheng and A. F. V. Roy, "Evolutionary fuzzy decision model for cash flow prediction using time dependent support vector machines," *International Journal of Project Management,* vol. 29, 2011, pp. 56–65.

[16] M. Cambal, D. Caganova, and J. Sujanova, "The industrial enterprise performance increase through the competency model application," Proceedings of the 4th European Conferences on Intellectual Capital, Helsinki, Finland, 2012, pp. 118–126.

Software Systems Development & Applications

SSD&A is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the discipline of software engineering. The SSD&A area emphasizes the issues relevant to developing and maintaining software systems that behave reliably, efficiently and effectively. This area investigates both established traditional approaches and modern emerging approaches to large software production and evolution. Events that constitute SSD&A are:

- ATSE'13 – 4th International Workshop Automating Test Case Design, Selection and Evaluation
- IWCPS'13 – International Workshop on Cyber-Physical Systems
- PBDA'13 – Performance of Business Database Applications
- WAPL'13 – 4th Workshop on Advances in Programming Language

4th Workshop on Advances in Programming Languages

PROGRAMMING languages are programmers' most basic tools. With appropriate programming languages one can drastically reduce the cost of building new applications as well as maintaining existing ones. In the last decades there have been many advances in programming languages technology in traditional programming paradigms such as functional, logic, and object-oriented programming, as well as the development of new paradigms such as aspect-oriented programming. The main driving force was and will be to better express programmers' ideas. Therefore, research in programming languages is an endless activity and the core of computer science. New language features, new programming paradigms, and better compile-time and run-time mechanisms can be foreseen in the future.

The aims of this event is to provide a forum for exchange of ideas and experience in topics concerned with programming languages and systems. Original papers and implementation reports are invited in all areas of programming languages.

Topics

Major topics of interest include but are not limited to the following:

- Automata theory and applications
- Compiling techniques
- Domain-specific languages
- Formal semantics and syntax
- Generative and generic programming
- Grammarware and grammar based systems
- Knowledge engineering languages, integration of knowledge engineering and software engineering
- Languages and tools for trustworthy computing
- Language theory and applications
- Language concepts, design and implementation
- Markup languages (XML)
- Metamodeling and modeling languages
- Model-driven engineering languages and systems
- Practical experiences with programming languages
- Program analysis, optimization and verification
- Program generation and transformation
- Programming paradigms (aspect-oriented, functional, logic, object-oriented, etc.)
- Programming tools and environments
- Proof theory for programs
- Specification languages
- Type systems
- Virtual machines and just-in-time compilation
- Visual programming languages

Steering Committee

Lukovic, Ivan, University of Novi Sad, Serbia, Serbia
Mernik, Marjan, University of Maribor, Slovenia
Slivnik, Bostjan, University of Ljubljana, Slovenia

Event Chair

Janousek, Jan, Czech Technical University, Czech Republic

Program Committee

Aycock, John, University of Calgary, Canada
Chen, Haiming, Chinese Academy of Sciences, China
Henriques, Pedro Rangel, Universidade do Minho, Portugal
Horvath, Zoltan, Eotvos Lorand University, Hungary
Ivanovic, Mirjana, University of Novi Sad, Serbia
Kardas, Geylani, Ege University International Computer Institute, Turkey
Kollar, Jan, Technical University of Kosice, Slovakia
Kosar, Tomaz, University of Maribor, Slovenia
Liu, Shih-Hsi "Alex", California State University, United States
Lukovic, Ivan, University of Novi Sad, Serbia, Serbia
Mandreoli, Federica, University of Modena, Italy
Martínez López, Pablo E. "Fidel", Universidad Nacional de Quilmes, Argentina
Mernik, Marjan, University of Maribor, Slovenia
Milasinovic, Boris, University of Zagreb Faculty of Electrical Engineering and Computing, Croatia
Moessenboeck, Hanspeter, Johannes Kepler Universitat Linz, Austria
Papaspyrou, Nikolaos, National Technical University of Athens, Greece
Pereira, Maria Joao Varanda, Instituto Politecnico de Braganca, Portugal
Poruban, Jaroslav, Technical University of Kosice, Slovakia
Rodriguez, Jose Luis Sierra, Universidad Complutense de Madrid, Spain
Slivnik, Bostjan, University of Ljubljana, Slovenia
Splawski, Zdzislaw, Wroclaw University of Technology, Poland
Watson, Bruce, Stellenbosch University, South Africa

Position Papers Position Papers of the 2013 Federated Conference on
Computer Science and Information Systems pp. 121–126

Modelling Information Systems by Document Flow Description

Vladimir Ivančević, Marko Knežević, Ivan Luković
University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
Email: {dragoman, marko.knezevic, ivan}@uns.ac.rs

Verislav Djukić
Djukic — Software Solutions,
Gärtnerstr. 17, 90408 Nürnberg, Germany
Email: info@djukic-soft.com

***Abstract*—In this position paper, we argue in favour of three points related to document-centric modelling of information systems: (i) information systems of some organizations may be understood in terms of documents, actions, actors, and document flow; (ii) modelling document flow may be a central step in modelling an information system; and (iii) as an approach to information system modelling, document flow modelling may be coupled with the form-based approach to information system development that is featured in IIS*Case, a model-driven development tool, because the form concept is semantically close to the document concept. With respect to these claims, we formulate a document-centric and model-driven approach to information system development, explain its particularities, and present a plan of activities that should lead to its implementation. By relying on domain-specific languages, the proposed approach should allow generation of information systems supporting document manipulation within a document flow and process mining.**

I. Introduction

DOCUMENTS are central to information systems (ISs). They may represent a plan of future actions or evidence of past activities that are relevant to the functioning of an information system. Therefore, in this position paper, we argue in favour of three points regarding information system modelling with respect to documents. First, we argue that some information systems may be described by specifying instances of concepts belonging to one of the following categories: documents, actions regarding document manipulation, actors, and document flow. Each piece of information that is needed for the operation of these information systems belongs to an instance from one of the aforementioned categories. Second, we argue that modelling document flow may be considered a central part of information system modelling since it integrates instances from all of the four categories. Furthermore, specifying document flow is generally platform independent and detached from the technology specific terminology, which makes it accessible to business analysts and allows their involvement with the IS modelling process. Third, we argue that document flow modelling may be integrated with the form-based approach to IS development, which is utilized in the IIS*Case software development tool [1], owing to the semantic similarity between forms and documents. The integration of the two approaches would provide the means to develop process-aware information systems (PAISs), which in turn would also support the use of process mining in generated instances of ISs. Therefore, in this paper, we present the rationale behind our claims, as well as sources that are related to the aforementioned issues. We also elaborate on the proposed document-centric approach to IS development and give a plan of activities that should lead to the implementation of document flow modelling capabilities and process mining by relying partially on the features of IIS*Case.

The proposed IS development approach, which in addition to being document-centric is also model-driven, would allow rapid automatic generation of prototype ISs. However, its principal advantage over traditional approaches is its suitability for the development of ISs in organizations whose activities are tightly related to constant manipulation of high volumes of official documents. Generated ISs would have built-in features supporting document tracking within the flow, document flow analysis, and use of attributes that are typical of documents (such as references between documents and validity dates). Furthermore, domain-specific languages (DSLs) have a key role in the proposed approach, because each step in the modelling process requires creation of a different specification or model. For that reason, there is a strong need to have separate DSLs for specifying document structure, modelling document flow, and defining form types. It is also necessary to define transformations between these DSLs since specifications of document structure and flow models need to be integrated first and then translated to specifications of form types, which are further used in the generation of IS prototypes. As the proposed approach combines domain-specific modelling languages (DSMLs), transformation engines, and generators, it represents a fine example of Model-Driven Engineering [2].

In the following sections, we argue in favour of the three claims that are stated herein. In Section II, we illustrate how capturing document flow may help uncover activities that are relevant for the functioning of the information system in an organization. Section III points out how document flow modelling may serve as a basis for information system modelling and implementation. In Section IV, ideas from the two aforementioned sections are evaluated with respect to the IIS*Case tool, which allows form-based modelling and development of information systems.

The research presented in this paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant III-44010

II. Relevance of documents and document flow within organizations

According to Briet [3], whose work has significantly influenced modern perspective on documentation, document is defined as "any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental" [4]. Documents, whether they are paper-based or digital, are essential in any organization. This relevance may be further observed in the context of information retrieval, where a document is characterized as "a unit of retrieval. It might be a paragraph, a section, a chapter, a Web page, an article, or a whole book." [5]. Therefore, by being central to information retrieval, documents are also central to information systems and, consequently, could serve as a basic unit in IS modelling.

Although there are records that are required by state regulations, record keeping is primarily needed for purely practical reasons because information within documents about past, current, or planned activities is vital to the accomplishment of the organization's mission. For instance, in addition to keeping usual information about employees or finances, functioning of a faculty requires storing information about its central activities, such as teaching and research. Such records contain data about student enrolment, instructor assignment, individual assessment results, final grades, and many other aspects of faculty operation. These records are typically organized in the form of documents that have a precisely defined structure and life cycle. Furthermore, there are also strict regulations about who may create, have access, modify, or destroy a document. We illustrate the stated viewpoint on the example of electing an instructor for a vacant position at a faculty. The following description is based on a segment of the actual election process at the Faculty of Technical Sciences in Novi Sad. The dean of the faculty may make an official decision about announcing a call for a vacant position. Such decision represents an official document that needs to be signed and archived at the faculty. In the similar manner, the elective council of the department to which the vacant position is associated needs to make a decision to appoint committee for applicant evaluation, which is yet another document. Once these two documents come into effect, a third document, the decision on the appointment of the committee members for the evaluation of applicants, may be created. Next, applicants are required to submit a set of prescribed documents. Once the application period is closed, the committee members write a report on the applicants (another document). After another decision of the department's elective council and consent from the faculty's educational-scientific council, the dean may make a decision to elect the candidate that is recommended by the elective committee. As illustrated, instructor election may be a complex process involving creation of numerous documents and intervention of various faculty members and non-faculty personnel. It includes storing several decisions of faculty bodies (faculty documents) and candidate applications (also documents). Each relevant step of the process is marked by an official document.

For each workflow at a faculty or some other organization, there may be a model that contains: instances of required documents, description of action flow between these instances, actors that are involved in the workflow, and specification of actions regarding document manipulation. From the perspective of data, such a document-centric model could capture all the relevant aspects of activities within the modelled system. Document content determines which data need to be stored in the information system of an organization, while action flow indicates how the state of a document may change. Actors correspond to actual user groups of the information system while actions define which document data may be manipulated by which user group. Therefore, the four principal elements of such model include instances of: documents, actions regarding documents, actors, and document flow between actions.

The use of documents and document flows within a system generally requires formalisms for document description. For instance, according to the Pentaformat model [6], five components may be extracted from a document: content, structure, presentation, behaviour, and metadata. For that reason, it is necessary to define languages that could be utilized to describe all of the aforementioned dimensions. These languages could be further used to model various aspects of organizational systems and, consequently, ISs, such as communication between actors and the system where documents are means of implementing this communication.

The document-centric approach may not be the most convenient solution for all types of organization, particularly not for those where document flows are unstable or where structured documents are not much used. Nonetheless, in organizations where documents are explicitly acknowledged, precisely defined and constantly manipulated, such approach should be preferred because its main concepts are semantically closer to actual reality and functioning of those organizations.

III. Developing information systems with respect to document flow

In IS modelling, there are many well-tried approaches. As evidenced by the current popularity of UML (Unified Modeling Language) [7], BPMN (Business Process Model and Notation) [8], WS-BPEL (Web Services Business Process Execution Language) [9], and YAWL (Yet Another Workflow Language) [10], process and workflow modelling are often used as means of capturing the essence of the system for which an IS is built. Modelling document flow is generally similar to modelling a workflow since the nature of flow varies little between these modelling scenarios. The main difference between the two is that, in document flow modelling, an activity generally results in document creation or modification, while in workflow modelling a result is not necessarily tied to any document. If the main focus of understanding an organization and its activities is set on understanding manipulated official documents, which in many organizations represent the only relevant records, then it may be possible to reduce workflow modelling as a basis of IS modelling to document flow modelling. An added benefit of this approach would

be the possibility to employ best practices from workflow modelling, such as workflow patterns [11], [12], [13]. In the rest of this section, we elaborate on the proposed approach.

Documents represent the only version of truth. If a piece of information may not be found within the archived documents, then it is considered to be missing or unavailable. All of the official actions within an organization should be based on information that may be found within the documents. A system is not understood in terms of various modelled entities, which is typical for many traditional IS development approaches, but solely as an environment in which documents are circulating along the predefined paths. In many instances there is little practical difference between entity models and documents (document types). For example, at a faculty, information about individual students or staff members may be contained within personal dossiers. However, the key difference between the proposed and the traditional approaches is in the view of an organization since the proposed approach is based on a paradigm that is document-centred. Such paradigm implies that arbitrary entity types are substituted by documents and all of their particularities. Therefore, modelling document flow in terms of the four concepts mentioned in the previous section (document, action, actor, and flow) could prove to be a feasible approach for IS development. Nevertheless, in the context of both IS modelling and implementation, additional information about these four concepts needs to be specified.

For each document, there should be a complete specification of its structure and dependencies (relations to other documents). The document concept denotes all document instances that conform to a shared specification. A document may be divided into sections and these further decomposed into subsections and fields. A section should feature a name. For each field, there should be a name, data type, and indication whether it is mandatory to be filled in.

There are various constraints that are generally related to document instances. Temporal aspects of a document instance may be relevant in certain cases, e.g., employment period is typically specified in employment contracts for faculty staff members. Therefore, a document instance should feature optional information regarding its validity period: a date when it comes into effect and expiry date. There are some documents that should have only one active (valid) instance. Such cases should be explicitly specified for each document. In this manner, the satisfaction of the aforementioned constraints could be automatically checked. Moreover, the case when a document is expected to be signed (verified) by a group of authorized persons may also be explicitly indicated in the document specification.

Relations between documents (or their parts) should be explicitly defined for three principal reasons: (i) to avoid information redundancies; (ii) to support detection of inconsistencies in the model; and (iii) to support analysis of document flows and, consequently, of workflows in the organization.

These relations may belong to one of the two types: (i) referring relations, where a relation only marks other document (or its part) that is semantically relevant for the current document but whose contents are not included in the current document, similarly to the notion of a hyperlink in a web page; and (ii) inclusive relations, where a relation also indicates insertion of selected contents from other document into the current document.

In document flow, each step in the life cycle of a document represents a different document state. Each state is described by a set of actions that associated actors need to perform on a document instance or set of its parts so the instance could advance to the next state in the flow. A document flow is not restricted to a single document, as it should enclose all documents that are relevant to a well-defined procedure within an organization. Allowed actions include creation, reading, modification, signing, copying, and removal. An action may refer to a whole document or a set of its parts, with the exceptions of creation and removal, which may refer only to whole documents. Similarly to other flow modelling languages, there should be support for at least the following relations between states: sequences, parallelism, loops, exceptions, and cancelation. A single state may have multiple preceding or multiple succeeding states. Conditions for state transitions should include relational and logical operators, while satisfaction of these conditions would allow actions in one or more succeeding states depending on the actual relation between the current and the succeeding states. Each flow has at least one initial and at least one final state. Starting from an initial state in the flow, a final state may be reached by satisfying conditions for intermediary states that connect the initial and final state. Two flows may be associated by connecting a final state of the preceding flow to an initial state of the succeeding flow using one of the aforementioned state relations. In this manner, long flows may be decomposed and the complexity of flow models significantly reduced. Initial and final states are further classified as: (i) independent, which cannot be associated to any state from other flows; or (ii) dependent, which have to be related to a state in another flow.

As opposed to decomposition, it should be also possible to automatically form a complete model of document flows that features each individual flow, whether it is related to other flows or not. Such universal model could indicate all of the actions that initiate different workflows in an organization, as specified by independent initial states. Moreover, it could provide analysts with a valuable perspective on all document-related processes within an organization.

Documents and actors are also related through actor permissions regarding document manipulation. The permission concept is derived from the basic concepts, which are presented in Section II. Each actor permission refers to a selected document state, action and actor. An actor permission applies to all of the subsequent states of that document inside its flow. In the case of conflicting actor permissions for a single state, precedence is given to the permission whose assignment is closest to the state, starting from the state in conflict and going upwards in the document flow. As in the case of actions, actor permissions refer to document creation, reading, modification, signing, copying, and removal. Each type of permission may

be granted or denied. If there is no explicit declaration of actor permission for a document in some state, then such permission is considered to be denied. Furthermore, permissions regarding reading, modification, signing, and copying, may also refer to individual document parts.

The proposed approach is model-driven because the information that is necessary for generation of IS prototypes may be found within the models of document flow and the transformed flow information could be used as an input to the process of form-based IS development in IIS*Case or some other tool. Moreover, the application of the approach should result in generation of a PAIS, i.e., "a software system that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models" [14]. Generated ISs should allow their users to work in terms of document flow by featuring automatized notification for new work tasks that are related to manipulation of a single document. These tasks would correspond to states in flow models and, in a single document flow, an IS user should be able to perform only actions that are associated to the current state in the flow. However, another idea worth considering is the implementation of an option to bypass a part of a regular procedure because every model is a simplification of the real process and does not encompass all process instances that may occur in an organization. Automatic storing of information about all performed user actions (document manipulation according to some document state) would allow process mining, i.e., analysis of document flows in terms of time, actors, and actual vs. expected paths of documents in the organization (flow-to-model and model-to-flow conformance).

IV. Integrating document-centric IS development into IIS*Case

The document-centric view of a system, as expressed in the previous section, shares some of its main points with the approach to IS generation that is used in IIS*Case. In this section, we elaborate on how concepts such as document, action, actor, and document flow may be merged with the standard concepts of IIS*Case.

IIS*Case is a software tool for IS development that has been constantly extended and updated over the past 20 years [1]. It relies on the form concept as the basic unit in the development of database schemata [15], [16] and ISs [17]. Such form [18], [19] is actually a screen form in an application that is used to enter, view, modify, or remove data about an entity which may be recorded in an IS. It generally corresponds to a business document that is used in the organization supported by the IS. Form specification is fully detached from the technology that is used in IS implementation, which could facilitate integration of the two approaches. All of this indicates that the form concept may be employed in the proposed document-centric approach since form specification generally covers description of the corresponding document's structure. Furthermore, document flow, which may encompass several documents, could be mapped to the application system concept, which is primarily used to group form types in IIS*Case.

Using IIS*Case, it is possible to generate program code of an IS by specifying forms (form models) and then running a transformation process. The concepts that may be used to create these platform independent models (PIMs) of an IS are formally presented in a meta-model [20] which was specified using the EMF Ecore [21] version of MOF [22]. For form-based IS design, IIS*CDesLang [23], a textual DSL, was specified by an attribute grammar using the VisualLisa programming environment [24]. For the similar purpose, another textual DSL is being developed by directly using the meta-model of IS concepts to generate concrete syntax of the language. In the integration of the two approaches, these two DSLs could serve as target languages in the translation of document specifications. This would require construction of a new DSL for creating document descriptions that are in accordance with the guidelines from the previous section. In case the new document specification DSL is derived from a meta-model, one of the available QVT [25] implementations could be used to map the concepts of this new DSL to the IS concepts in IIS*Case. Another possibility that also deserves consideration is the direct use of the two IS design DSLs instead of a more document-centric specification language.

On the other hand, some concepts from the proposed document-centric approach do not have a match in IIS*Case. In the latest version of IIS*Case, there is no explicit notion of document flow. Therefore, a DSML should be devised for this purpose and combined with the document specification DSL. Since document flow is modelled by human experts, a graphical language would be the most convenient solution. The main reason for this decision may be found in the results of an empirical study that compared understanding of business process descriptions with respect to the used type of notation [26] because the authors concluded that business analysts benefited from reading a graphical model. Regarding the semantics, the new language could be formally based on Petri Nets [27] or a graphical variant of π-calculus [28], as both have been extensively used to model various types of processes. Whatever the formal foundation of the language, document flow modelling would become a central part of the IS development process in IIS*Case. It would shift focus from data entry and grant a more prominent role to processes within the organization, which has been a general trend in IS development in recent years.

Furthermore, there is a need to introduce actors (users) and actions/permissions regarding document use. Actors and their permissions would have to be explicitly specified, possibly within the new document flow language or the document specification language. These concepts would be tightly related to forms in the generated application: actors would correspond to application users, while actor permissions would correspond to data manipulation privileges within the application. As a result, both the meta-model and IIS*CDesLang would need to be updated with these new concepts.

In IS development, explicit acknowledgement of processes within an organization, has many far reaching implications. In the latest generation of process-aware ISs, process mining [29]

allows advanced analysis of processes within an organization and, consequently, their improvement. In the proposed document-centric approach, similar benefits could be obtained by making ISs that are generated using IIS*Case process-aware. In order to support process mining, key events (actions) in the generated ISs need to be recorded in special logs. In this context, an event designates a transition of a document to some of the modelled states. For each event, there should be information about the corresponding document instance, reached document state, actor whose action triggered that transition, and timestamp. Such event logs could serve as input to various types of process-related analyses.

Support for analysing explicitly modelled document flows could be automatically added to ISs during the generation process in IIS*Case. One component of each generated IS should be focused on process mining and support viewing of document flow models and recorded events. Out of the typical process mining scenarios [30], application support should be added for the following two: (i) conformance, in which recorded document flows are compared to those expected according to flow models; and (ii) extension, in which flow models are enhanced with concrete information about flow path coverage, load distribution between actors, and temporal aspects of flow. The conformance check would help analysts identify how exactly a procedure deviates from the plan (model), i.e., which recorded events are not expected since the action sequence that they are part of is not supported by the corresponding flow model. A closer inspection of such events could uncover periods when these deviations most often occur and indicate which actors are responsible. On the other hand, the extension could provide an analyst with valuable information on which path segments require the longest time to complete or which actors are disproportionately burdened by workload. Such insight could help reduce delays in workflows or achieve better load balancing.

The form-based approach to IS development in IIS*Case may serve as foundation for the proposed document centric approach. However, as previously mentioned, there are several key activities that should be performed in the integration of these two approaches: (i) a DSL for document structure specification needs to be constructed, together with adding support for translation of document specifications to programs written in IIS*CDesLang or the DSL for IS design that is being developed from the meta-model of the IIS*Case PIM concepts; (ii) a graphical DSML for document flow modelling needs to be constructed and coupled with the document specification language; (iii) notions of actor and action/permission need to be introduced into IIS*Case; (iv) ISs that are generated by IIS*Case need to be also process-aware; and (v) ISs that are generated by IIS*Case need to feature a set of typical process mining capabilities related to conformance and extension. The details on the implementation of process mining support within IIS*Case also need to be specified in our future research. Once all of the aforementioned activities are completed, a detailed evaluation of the proposed approach could be made, together with a comparison to other approaches to IS development.

V. Related work

The importance of the data perspective in process modelling has increased recently, as evidenced by the growing number of data-oriented approaches to workflow and process modelling [31], which also include document-based approaches.

In [32], the authors propose an XML Document Centric Workflow Management System (XDoc-WFMS), which supports embedding document flows within organization workflows, document access control, and agents handling various tasks within the system. Although XDoC-WFMS and the approach proposed herein share similar goals and concepts, there are three key differences. First, XDoc-WFMS uses the XML format to store documents, whereas we do not restrict our approach to a single document storage format — documents are going to be stored in a database automatically designed in the system generation process within the IIS*Case tool, which may include different types of databases. Second, as opposed to our approach, XDoc-WFMS does not acknowledge the need for process mining, which may be understandable given the period when the approach was conceived. Third, the main objectives of the two approaches do not completely coincide. Whereas XDoc-WFMS should be considered only a part of a larger system in which document processing should be integrated with other business activities, our approach is aimed at the development of information systems for organizations where every activity has to be formally documented. XFlow [33] represents another example of a document-centric workflow framework that revolves around agents and XML documents. In addition to this, it heavily relies on XML-related technologies such as XSLT, XPath, XForms, and SVG.

In [34], the authors present a document-driven workflow system not focused on the control flow perspective. Unlike the previous two XML-based approaches, this one is implemented using a relational database management system (RDBMS) and database triggers, which are used in the enactment of the workflow system. Although such approach may be suitable for ad hoc workflows, which lie between the well-structured workflows and less-structured cooperation systems, processes cannot be visualised because there is no predefined control flow, which in turn narrows down the environments where this approach may be used. For these reasons, we may adopt only some of the ideas exhibited in that approach.

VI. Conclusion

In this position paper, we outline some of our ideas on IS modelling and development with respect to the role of documents in organizations. We argue that understanding documents, their flows, and performed actions of actors in an organization may serve as a basis for IS development. The proposed document-centric and model-driven approach, which includes explicit modelling of document flows and document structure specification, could be a step preceding IS generation. On the other hand, the approach to IS development in IIS*Case, which revolves around specifying forms for data manipulation and generating executable code using these specifications, could be used in the later phase of IS

development. Based on the comparison of concepts from the two approaches, it appears that these approaches could be integrated owing to the similarity of the form concept to the document concept. Document specification and flow modelling would represent the first phase in IS development, while the form-based approach, after a series of intermediate transformations, would provide generation of an IS. Each step in this development process would include using one of the custom DSLs to create necessary specifications. In the context of integrating the two approaches, primary research activities are identified in this paper. Because the unified approach would acknowledge organization processes in the form of document flows, it could be also classified as process-based. The idea to make generated ISs process-aware and add "out-of-box" support for process mining is motivated by the latest trends in IS development that are gaining commercial recognition.

In the proposed approach, there are certain issues whose resolution is a part of future research efforts. These include: (i) finding a way to support changes in document structure or flow without the need to fully replace the previous version of IS (change patterns [35] could be a starting point); and (ii) extend the approach to support the idea of "a paperless office" where documents are (almost) exclusively stored and manipulated in their digital form. After adding new concepts and document flow modelling capabilities to IIS*Case, the unified approach could be more thoroughly evaluated and compared to other IS development approaches. This addition to IIS*Case could also serve as a basis for automatic generation of SOA (service-oriented architecture) system specifications and more thorough analysis of the modelled system. Moreover, the application of the proposed approach could improve understanding between the two key groups of people involved with IS development – business users who state requirements and actual system users on one hand, and analysts and IS designers on the other.

REFERENCES

[1] I. Luković, V. Ivančević, M. Čeliković, and S. Aleksić, "DSLs in action with model based approaches to information system development," in *Formal and Practical Aspects of Domain-specific Languages: Recent Developments*, ed., M. Mernik. IGI Global, Hershey, PA, 2012, pp. 502–532.

[2] D. C. Schmidt, "Model-driven Engineering," *IEEE Computer*, vol. 39, no. 2, 2006, pp. 25–31.

[3] S. Briet, *Qu'est-ce que la documentation?*, Éditions documentaires, industrielles et techniques, Paris, 1951.

[4] M. Buckland, "Information as Thing," *Journal of the American Society of Information Science*, vol. 42, no. 5, 1991, pp. 351–360.

[5] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley Longman Publishing, Boston, MA, 1999.

[6] A. Di Iorio, "Pattern-Based Segmentation of Digital Documents: Model and Implementation," Doctoral thesis, University of Bologna, Italy, 2007.

[7] Unified Modeling Language (UML), http://www.omg.org/spec/UML/

[8] Business Process Model and Notation (BPMN), http://www.omg.org/spec/BPMN/index.htm

[9] Web Services Business Process Execution Language Version 2.0, http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.pdf

[10] YAWL: Yet Another Workflow Language, http://www.yawlfoundation.org/

[11] W. M. P. van der Aalst, A. P. Barros, A. H. M. ter Hofstede, and B. Kiepuszewski, "Advanced Workflow Patterns," In *Proceedings of the Fifth IFCIS International Conference on Cooperative Information Systems (CoopIS'2000)*, volume 1901 of Lecture Notes in Computer Science, Springer, Eilat, Israel, 2000, pp. 18–29.

[12] W. M. P. van der Aalst, A. H. M. ter Hofstede, B. Kiepuszewski, and A. P. Barros, "Workflow Patterns," *Distributed and Parallel Databases*, vol. 14, no. 1, 2003, pp. 5–51.

[13] N. Russell, A. H. M. ter Hofstede, W. M. van Der Aalst, and N. Mulyar, "Workflow Control-flow Patterns: A Revised View," 2006, http://www.workflowpatterns.com/documentation/documents/BPM-06-22.pdf

[14] M. Dumas, W. Van Der Aalst, and A. Ter Hofstede, *Process-aware Information Systems*, John Wiley & Sons, Hoboken, NJ, 2005.

[15] I. Luković, S. Ristić, P. Mogin, and J. Pavićević, "Database Schema Integration Process — A Methodology and Aspects of its Applying," *Novi Sad Journal of Mathematics*, vol. 36, no. 1, 2006, pp. 115–140.

[16] I. Luković, P. Mogin, J. Pavićević, and S. Ristić, "An Approach to Developing Complex Database Schemas using Form Types," *Software — Practice & Experience*, vol. 37, no. 15, 2007, pp. 1621–1656.

[17] S. Ristić, S. Aleksić, I. Luković, and J. Banović, "Form-driven Application Development," *Acta Electrotechnica et Informatica*, vol. 12, no. 1, 2012, pp. 9–16.

[18] P. Mogin, I. Luković, and Z. Karadžić, "Relational Database Schema Design and Application Generating using IIS*CASE tool," in *Proceedings of International Conference on Technical Informatics*, Timisoara, Romania, Vol. 5, 1994, pp. 49–58.

[19] P. Mogin and I. Luković, "A Prototyping CASE Tool," in *Proceedings for the Dedicated Conference on Rapid Prototyping in the Automotive Industries*, XXVIII International Symposium on Automotive Technology and Automation, Stuttgart, Germany, 1995, pp. 261–268.

[20] M. Čeliković, I. Luković, S. Aleksić, and V. Ivančević, "A MOF based Meta-model and a Concrete DSL Syntax of IIS*Case PIM Concepts," *ComSIS*, vol. 9, no. 3, 2012, pp. 1075–1104.

[21] Eclipse Modeling Framework Project (EMF), http://www.eclipse.org/modeling/emf/

[22] OMG's MetaObject Facility, http://www.omg.org/mof/

[23] I. Luković, M. J. V. Pereira, N. Oliveira, D. d. Cruz, and P. R. Henriques, "A DSL for PIM Specifications: Design and Attribute Grammar based Implementation," *ComSIS*, vol. 8, no. 2, 2011, pp. 379–403.

[24] N. Oliveira, M. J. Varanda Pereira, P. R. Henriques, D. Cruz, and B. Cramer, "VisualLISA: A Visual Environment to Develop Attribute Grammars," *ComSIS*, vol. 7,no. 2, 2010, pp. 265–289.

[25] Meta Object Facility (MOF) 2.0 Query/View/Transformation (QVT), http://www.omg.org/spec/QVT/

[26] A. Ottensooser, A. Fekete, H. A. Reijers, J. Mendling, and C. Menictas, "Making Sense of Business Process Descriptions: An Experimental Comparison of Graphical and Textual Notations," *Journal of Systems and Software*, vol. 85, 2012, pp. 596–606.

[27] C. A. Petri, "Kommunikation mit Automaten," Unpublished doctoral thesis, Fakultät für Mathematik und Physik, Technische Hochschule Darmstadt, Germany, 1962.

[28] R. Milner, *Communicating and Mobile Systems: The π-calculus*, Cambridge University Press, Cambridge, UK, 1999.

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

[30] W. van der Aalst, "Process-aware Information Systems: Lessons to be Learned from Process Mining," *Transactions on Petri Nets and Other Models of Concurrency II* (2009), pp. 1–26.

[31] H. Garcia, "Evaluation of Data-centric Process Modeling Approaches," Master thesis, Eindhoven University of Technology, the Netherlands, 2011.

[32] R. Krishnan, L. Munaga, and K. Karlapalem, "XDoC-WFMS: A framework for document centric workflow management system," in *in Conceptual Modeling for New Information Systems Technologies*, Springer Berlin Heidelberg, 2002, pp. 348–362.

[33] A. Marchetti, M. Tesconi, and S. Minutoli, "Xflow: An xml-based document-centric workflow," in *Web Information Systems Engineering — WISE 2005*, Springer Berlin Heidelberg, 2005, pp. 290–303.

[34] J. Wang and A. Kumar, "A framework for document-driven workflow systems," in *Business Process Management*, Springer Berlin Heidelberg, 2005, pp. 285–301.

[35] B. Weber, M. Reichert, and S. Rinderle-Ma, "Change Patterns and Change Support Features — Enhancing Flexibility in Process-aware Information Systems," *Data & knowledge engineering*, vol. 66, no. 3, 2008, pp. 438–466.

Author Index